Authors, please upload your paper here.

NATIONAL BUREAU OF ECONOMIC RESEARCH, INC.

SUMMER INSTITUTE 2011

Industrial Organization Workshop

Meghan Busse and Kate Ho, Organizers

August 1 - 2, 2011

Royal Sonesta Hotel
40 Edwin H. Land Boulevard
Cambridge, Massachusetts

PROGRAM

Monday, August 1:

8:45 am Coffee and Pastries

9:30 am Silke J. Forbes, University of California at San Diego
Mara Lederman, University of Toronto
Trevor Tombe, University of Toronto
Do Firms Game Quality Ratings? Evidence from Mandatory Disclosure of Airline On-Time Performance

Discussant: Ginger Jin, University of Maryland and NBER

10:45 am Break

11:00 am James W. Roberts, Duke University
Andrew Sweeting, Duke University and NBER
When Should Sellers Use Auctions?

Discussant: Jakub Kastl, Stanford University and NBER

12:15 pm Lunch

1:30 pm Christopher T. Conlon, Columbia University
Julie Holland Mortimer, Harvard University and NBER
Effects of Product Availability: Experimental Evidence

Discussant: Marc Rysman, Boston University

2:45 pm Break

3:00 pm	Meredith Fowlie, University of California at Berkeley and NBER Mar Reguant Rido, Massachusetts Institute of Technology Stephen P. Ryan, Massachusetts Institute of Technology and NBER *Pollution Permits and the Evolution of Market Structure* Discussant: Ryan Kellogg, University of Michigan and NBER
4:15 pm Â	Break
4:30 pm	Steven Tadelis, University of California at Berkeley and NBER Florian Zettelmeyer, Northwestern University and NBER *Information Disclosure as a Matching Mechanism: Theory and Evidence from a Field Experiment* Discussant: Robert Porter, Northwestern University and NBER
5:45 pm	Adjourn
6:00 pm	Clambake: Harvard Faculty Club, 20 Quincy Street, Cambridge, MA

Tuesday, August 2:

8:30 am	Coffee and Pastries
9:15 am	Song Yao, Northwestern University Carl F. Mela, Duke University Jeongwen Chiang, China Europe International Business School Yuxin Chen, Northwestern University *Determining Consumersâ™ Discount Rates With Field Studies* Discussant: Wes Hartmann, Stanford University
10:30 am Â	Break
10:45 am	Connan Snider, University of California at Los Angeles Jonathan W. Williams, University of Georgia *Barriers to Entry in the Airline Industry: An Analysis of the Wendel H. Ford Aviation Act* Discussant: Severin Borenstein, University of California at Berkeley and NBER
12.00 n	Lunch and Adjourn

Do Firms Game Quality Ratings? Evidence from Mandatory Disclosure of Airline On-Time Performance

Silke J. Forbes
University of California, San Diego

Mara Lederman
University of Toronto, Rotman School of Management

Trevor Tombe
University of Toronto

July 2011

Abstract

Many quality disclosure programs provide consumers with information that is based on whether a product meets a particular threshold. This creates the potential for "gaming" as firms have incentives to improve the quality of specifically those products that can easily be brought above the threshold. We investigate this type of behavior in the context of government-mandated disclosure of airline on-time performance. While this program collects data on the actual minutes of delay incurred, it ranks airlines based only on the fraction of their flights that arrive 15 or more minutes late. This creates the incentive for airlines to selectively reduce delays on flights they expect to arrive with about 15 minutes of delay. We estimate the extent to which airlines engage in this type of gaming and, in particular, whether the occurrence of such gaming depends on whether employees are explicitly incentivized based on the airline's performance in the program. We find little evidence of gaming by airlines that have no incentive programs in place or by airlines that have implemented incentive programs with targets that are unrealistically hard to achieve. On the other hand, we find strong evidence of "gaming" by airlines that have incentive programs with a target level of performance that can realistically be achieved. Specifically, for these airlines, we find that their flights that are predicted to arrive with between 15 and 16 minutes delay have significantly shorter taxi-in times than other flights and are significantly more likely to arrive exactly one minute sooner than predicted. Counterfactual exercises that simulate an airline's distribution of delays in the absence of taxi-time distortions indicate that even small improvements in taxi times can – if applied to the "right" set of flights – result in changes in an airline's ranking.

We thank Severin Borenstein, Bob Gibbons, Matt Mitchell, Steve Puller and seminar participants at Boston University, New York University (Stern), UC Berkeley (Haas), UC San Diego, the University of Maryland, the University of Toronto (Rotman), the U.S. Department of Justice, the AEA Meetings (2011), the Berkeley-Stanford IO Fest (2010) and the USC Conference on Game Theory in Law, Business and Political Economy for helpful comments.

I. Introduction

Disclosure programs exist in many industries in which consumers are imperfectly informed about product quality.[1] While the growing empirical literature on these programs has generally found that they result in improvements in product quality, firms also appear to engage in various types of behavior that attempt to "game" the disclosure scheme. For example, when only some dimensions of quality are reported, firms may substitute effort from unreported to reported margins (see, Jacob 2005 and Lu 2009); when quality depends on consumer characteristics, firms may choose to serve only a select set of consumers (see Dranove *et al*. 2003 and Werner and Asch 2005); and when the disclosure program is based on a particular quality threshold, firms may focus on improving the quality of those products that can most easily be brought over the threshold (see Neal and Schanzenbach 2010). The growing body of evidence on gaming implies that, in addition to considering the cost, precision and usefulness of the information being provided, the design of an optimal disclosure program must also consider the potential for firms to game the program. However, anticipating the potential for gaming - as well as the type of gaming - depends not only on the design of the program but also on the characteristics of the product and the internal organization and incentive schemes of the firm. For example, gaming may depend on how product quality can be manipulated and whether those in a position to manipulate it have incentives to influence the information reported.

In this paper, we explore the relationship between gaming of a disclosure program, the design of the program and the incentive schemes in place at the firms covered by the program. Our setting is the U.S. airline industry. Since 1987, airlines have been required to report to the Department of Transportation (DOT) the scheduled and actual arrival times of their domestic flights. Although the DOT collects detailed data about the actual minutes of delay incurred on

[1] See Dranove and Jin (2010) for a review of the literature on disclosure programs.

each flight, it only counts a flight as being “late” if it arrives 15 minutes or more behind schedule. The DOT issues monthly reports that rank airlines based on the percentage of their flights that are late under this definition and excerpts from these rankings are frequently reported in media outlets.[2] The design of this program clearly creates the potential for gaming as airlines have an incentive to reduce delays on specifically those flights that would otherwise arrive just over 15 minutes late. Small reductions in delay on these flights - which can likely be made at low cost - can improve an airline’s performance in the DOT rankings even though they may not necessarily improve its overall on-time performance.

Two features of this setting make it a particularly interesting one in which to investigate how gaming behavior is affected by characteristics of the product as well as organizational features of the firm. First, airlines cannot predict in advance which flights will be candidates for gaming. While airlines may be able to anticipate which routes or flights will, on average, have longer delays, they are unlikely to be able to anticipate which flights will arrive with exactly 14 versus 16 minutes of delay. Thus, to the extent that gaming occurs, it occurs in real time. As a result, the effort to game must come from front-line airline employees rather than executives or managers. This makes a consideration of employee-level incentives particularly relevant.

Second, between 1995 and 2009, five different airlines implemented employee bonus programs based explicitly on the airline’s ranking in the government’s ranking of on-time performance. Under these programs, each airline employee would receive a payment of between $65 and $100 in any month in which the airline as a whole placed at or near the top of the DOT ranking. While all of the programs created a free-rider problem by rewarding individuals based on firm-level performance, the programs differed significantly in how easy it was to achieve the target ranking. Thus, this empirical setting – combined with the richness of the flight-level data

[2] These rankings are published in the DOT’s “Air Travel Consumer Report”, which also contains separate rankings of airlines based on baggage handling, oversales, and customer complaints.

available - allows us to investigate not only the existence of gaming but also explore where and when it occurs and whether it is affected by the incentives provided to the employees most likely to engage in the gaming behavior.

We develop an empirical approach that allows us to estimate whether airlines systematically try to reduce delays on flights which would otherwise arrive slightly above the 15 minute threshold. Much of our empirical analysis focuses on differences in flights' taxi-in times. We focus on taxi-in times because this represents the final stage of a flight and thus the final point at which delays may be incurred or reduced. By the time a flight has touched down at the arrival airport of its route, an airline has a fairly precise estimate of the expected delay that the flight will have and can decide whether or not to try to reduce that delay below 15 minutes. We expect that taxi-in times can be reduced in several ways – for example, by preferential allocation of scarce resources such as ground crew, by employees exerting more effort and, in some cases, by simply lying about a plane's actual arrival time. While we cannot observe what actions airline employees take to reduce delays, we devise an empirical strategy to try to distinguish between lying and actual speeding up of planes that exploits the fact that, during our sample, some airlines reported their on-time performance manually while other reported automatically.

Our empirical analysis uses the very data that is collected by the DOT under the mandatory disclosure program. We construct a dataset that includes a random sample of domestic flights operated by the seven largest carriers between 1995 and 2010. We take advantage of the fact that, starting in 1995, the DOT began collecting information about each flight's wheels-off and wheels-on times (i.e.: the times at which it leaves the runway and touches down on the runway). This additional information allow us to construct a measure of every flight's *predicted delay* at the time that it touches down at the arrival airport. Our main set of regressions relate a flight's taxi-in time to its predicted delay and look for evidence of a nonmonotonicity right around the 15 minute threshold. We also estimate whether flights that are predicted to be 15 (16) minutes late are systematically more likely than any other flights to arrive exactly one (two) minute(s) earlier than predicted. We estimate these relationships for airlines without incentive programs in place and, separately, for each airline that introduced an incentive program. We focus the analysis, for now, on the seven large network carriers who were initially covered by the reporting requirements.

Our empirical analysis does not provide evidence of gaming by airlines without employee bonus programs in place. However, we find strong evidence of gaming by the first two of the five airlines that introduced these types of incentive programs - Continental Airlines (in 1995) and TWA (in 1996). During the first three years of its bonus program, Continental's taxi-in times for flights predicted to be between 15 and 16 minutes late were about 14 percent shorter than its taxi-in times for flights with predicted delays of less than 10 minutes. We see effects of a very similar magnitude when we look at TWA who also introduced a bonus program during this period. Moreover, the estimates for Continental and TWA reveal a discontinuous relationship between taxi-in times and predicted delay right around the 15 minute threshold. While one might have thought that airlines have the greatest incentive to reduce very long delays (because the costs of delays may be convex), we find that taxi-in times for the flights with predicted delays in the critical 15 minute range are significantly shorter than taxi-in times for flights with longer predicted delays. We also find – for both of these carriers – that their flights that we predict to be exactly 15 (16) minutes late are much more likely than any other flights to arrive exactly one (two) minutes sooner than predicted. When we investigate whether this gaming appears to reflect lying or actual reduction in taxi-in times, we find evidence for both. When we carry out the same series of analysis for the three airlines that introduced bonus programs after 2000, we find no evidence of gaming. We suspect that this is due to the much weaker incentives provided by these programs.[3] While the two early programs rewarded their employees if the airline was among the top five of the 10 airlines that were ranked at the time, the three later programs only rewarded employees if the airline achieved first or, in some cases, second place out of a much larger number of airlines that were, by that time, included in the rankings. Some of these airlines – for example, Hawaiian Airlines – consistently had substantially better on-time performance than any of the large network carriers.

In addition to the literature on gaming of quality disclosure programs, this paper is also related to research on gaming of employee incentive programs, such as Oyer (1998), Courty and Marschke (2004) and Larkin (2007). Finally, this work is related to Knez and Simester (2001) which has studied the effect of one of the airline employee bonus programs on the airline's overall delays. Knez and Simester show that overall departure delays decreased after the introduction of the bonus program, but they do not investigate the gaming of the disclosure program which is the focus of our paper.

The rest of the paper is organized as follows. Section II provides institutional background on the government disclosure program and on the airline bonus programs. Section III describes our data and sample. We outline our empirical approach in Section IV and present our results in Section V. A final section concludes.

II. Institutional Background

II.A. Disclosure of Airline On-Time Performance

All airlines that account for at least one percent of U.S. domestic scheduled passenger revenues have been required to submit information on their on-time performance to the

[3] Some of the differences across firms may also be due to differences in communication. For example, Continental's bonus program was introduced during a time period in which on-time performance was explicitly communicated as an important goal for the organization. We are in the process of investigating the communication strategies of the other carriers.

Department of Transportation under Title 14, Part 234 of the Code of Federal Regulations since September 1987. The reporting requirements have increased over time. Originally, airlines were only required to submit information on their scheduled and actual departure and arrival times and on flight cancellations and diversions. The original reporting requirement also did not to include flights that were delayed or cancelled because of mechanical problems. The reporting rule was amended in January 1995 to cover flights with mechanical problems. The 1995 amendment also required that additional data be reported, including taxi times and airborne times, as well as the aircraft's tail number. Additional amendments to the reporting rule required airlines to include delay causes for their flights beginning in November 2002 and to report tarmac delays for flights that are subsequently cancelled, diverted or returned to their gate beginning in October 2008.

These reporting requirements cover all of an airline's flights that depart from or arrive at one of 29 reportable airports. The airlines have the option of reporting these data for all of their other flights as well and all airlines have chosen to do so. They have an incentive to report the additional data because their on-time performance on the voluntarily reported flights is generally better than it is on the flights that are subject to the reporting requirement (because the 29 reportable airports include the some of the most congested airports in the U.S.) and the voluntarily reported flights are included in the main ranking that the DOT publishes.[4]

Airlines can record delays either manually or automatically through technology that is installed in the aircraft. While the automated devices are presumably reliable in recording the actual arrival times, there has been speculation that airlines which record delays manually may not record their arrival times accurately. Indeed, the distribution of arrival delays for manual reporters shows considerable rounding. Our empirical analysis below focuses on the seven largest network carriers. While we believe that most of these airlines reported their delays

[4] The DOT's report also contains a separate ranking based only on the reportable airports, but this ranking is not as highly publicized as the main ranking.

automatically during our sample period, several – including the two which implement the early employee bonus programs – likely used a combination of manual and automatic reporting.[5] This raises the possibility that airline employees who record flight delays manually may report delays of 14 minutes for flights whose actual delays are 15 minutes. While this would appear in our data as shorter taxi-in times for these flights, this would not reflect extra effort or preferential allocation of resources but rather would reflect employees lying about arrival times. Since this represents a different type of gaming, we have developed an approach (described below) for trying to identify the manual aircraft in the data and look separately at gaming behavior on manually and automatically reported flights.

II.B. Airline Bonus Programs

In February 1995, Continental Airlines was the first airline to implement a firm-wide employee bonus program which was based on the DOT's ranking. Under the program Continental would pay $65 to each full-time employee in every month that the airline was among the top five in the DOT's on-time performance ranking. In 1996, the program rules were changed to pay each employee $65 in every month that the airline ranked second or third and to pay $100 in months that the airline ranked first. The bonus program was part of a larger turnaround effort called the "Go Forward Plan" which sought to address poor performance and profitability at the airline.[6] The two other parts of the "Go Forward Plan" which were also related to improving on-time performance were changes in the flight schedule that increased aircraft turnaround time (i.e.: the time between flights) and the replacement or rotation of the senior manager at every airport. Thus, it is important to keep in mind that changes in on-time

[5] Starting in 1998, we know how each carrier reports in each month (automatic, manual or combination). Since our analysis covers the period between 1995 and 1998, we cannot be certain that the manner in which carriers reported in 1998 is the same as how they reported in the earlier years. However, anecdotal evidence and descriptive analysis of their delay distributions suggest they likely did.

[6] In 1994, Continental had the worst average on-time performance ranking among the ten reporting airlines.

performance after the introduction of the bonus program may be the result of a combination of all three changes. While we have no reason to believe that the increased turnaround time would specifically reduce delays on flights near the 15 minute threshold, increased emphasis within the organization on meeting the DOT's on-time target could enhance the effect of the explicit incentives provided by the bonus program.

In June 1996, TWA implemented an employee bonus program which closely resembled Continental's. TWA would pay $65 to each employee in every month in which the airline ranked top five in the following three rankings published by the DOT: on-time performance, baggage handling and customer complaints.[7] The airline would pay a total of $100 to each employee if the airline also ranked first in at least one of those categories. The program was later amended to reward employees if high rankings were sustained for an entire quarter (instead of a single month) and, in 1999, was changed to reward absolute measures of on-time performance (85% or better during the summer months, 80% or better during the winter months) rather than relative rankings. Like Continental's program, TWA's program was introduced after a period of very poor performance. TWA ranked worst in average on-time performance in 1995 and in 1996 and its baggage handling and customer complaints had been ranking among the worst since the beginning of the DOT's disclosure program in 1987.

Three other airlines introduced similarly structured bonus programs in subsequent years. These were American Airlines in April 2003, US Airways in May 2005, and United Airlines in January 2009. Table 1 summarizes the details of these bonus programs and the airlines' on-time performance one year before and after the introduction of their programs. The table also reports the number of months during the first year after the introduction of the bonus program in which the employees in fact earned bonuses. There are a number of things that are interesting to note

[7] The fourth ranked category, oversales, is a function of the airline's reservation system and not directly related to employee effort.

about this table. First, all of these airlines except American improved their ranking in the first year after the introduction of their bonus program, relative to the year prior. However, the accompanying improvement in on-time performance (i.e.: the percentage of flights delayed less than fifteen minutes) varies quite a bit, from less than one percentage point for TWA to almost ten percentage points for United.

Second, while the two earlier bonus programs by Continental and TWA, made it relatively easy for employees to earn bonuses by rewarding any placement in the top 5 spots of the ranking, the three later programs only rewarded first and (for American and United) second places which made it substantially harder for employees to earn bonuses. In fact, American's and US Airways' employees did not earn a single bonus in the first year after the introduction of their programs, and United's employees had only a single month in which they earned a bonus. In contrast, Continental's and TWA's employees earned bonuses in ten and four months, respectively, during the first year after the introduction of the program.

Another factor which appears to substantially affect the chance that an airline's employees might earn a bonus is the number and type of airlines included in the rankings. Until 2002, there were ten airlines that accounted for more than one percent of domestic passenger revenues and which were therefore included in the DOT's ranking. After 2001, the combination of growth by low-cost and regional carriers and reductions in capacity by the large network carriers led to an increase in the number of carriers that met the DOT reporting requirements. By January 2003, there were 17 airlines included in the ranking. Moreover, Hawaiian Airlines was added as an 18th carrier in November 2003. Since then - in every single month that we have looked at (including all of 2005 and all of 2009) - Hawaiian has always occupied the top spot in the ranking, typically with a substantial lead over the second-ranked airline. We suspect that this may have to do with the fact that Hawaiian operates at relatively uncongested airports with few weather disruptions. This combination of factors means that the three later bonus programs may have had much more negligible effects on the incentives of employees than the two earlier programs because employees of American, US Airways and United may have been aware that their likelihood of achieving a bonus was quite small.

III. Data

III. A. Data and Sample

Our empirical analysis uses the flight-level data on on-time performance collected by the U.S. Bureau of Transportation Statistics under the DOT's mandatory reporting program. We have collected these data for all reporting carriers for every year between 1988 and 2008, inclusive. Our empirical work below focuses on the years 1995 to 1998, 2003 to 2006, and 2008 to 2010 since these are the years during which the airlines introduced their employee bonus programs.[8] Because of the volume of data, we cannot investigate all five bonus programs in a single sample that includes data from the 15 years over which these programs are introduced. As a result, we construct separate samples which include several years of data around the introduction of the programs.

Comment [S1]: Edit this if we end up including later years.

Our regression sample includes domestic flights operated by the following seven airlines: American Airlines, Continental Airlines, Delta Air Lines, Northwest Airlines, TWA, United Airlines, and US Airways. Because this dataset is very large, we only include their flights between the 29 airports for which the airlines are required to report their on-time performance. To further reduce the size of the dataset, we take a random sample of flights by restricting to every fifth day of the year. In addition, we drop flights that meet any of the following conditions: depart more than 15 minutes early (since we suspect this may represent a rescheduled

[8] 1995 is also the year in which the DOT began collecting data on wheels-up and wheels-down times and we require this particular data for our empirical analysis.

flight), arrive more than 90 minutes early, depart on what appears to be the following calendar day, have a taxi-out or taxi-in time of more than 60 minutes, have missing values for their scheduled arrival or departure times, have a distance of less than 25 miles, or operate fewer than 20 times during the quarter. Our final sample includes 3, 067,533 flights.

Table 2 presents summary statistics for the main variables in the data.[9] The average arrival delay in the sample is about seven minutes. About 21% of flights arrive 15 or more minutes late and thus are considered "late" under the program's definition. The average air time is 109 minutes, the average taxi-out time is about 15 minutes and the average taxi-in time is 6 minutes. Note that taxi-out time includes the time between when an aircraft leaves the gate and when it leaves the ground. Similarly, taxi-in time includes the time between when an aircraft touches the ground and arrives at the gate. Delays incurred waiting for a runway or waiting for an arrival gate will therefore be included in taxi-out and taxi-in times, respectively.

III. B. Histograms of Arrival Delays

Figure 1 shows the distribution of arrival delays for the seven network carriers in our regression sample as well as the three other carriers that met the DOT's reporting requirements during our initial sample period. These three additional carriers are Southwest Airlines, America West and Alaska Airlines. We truncate the histogram at -20 on the left and at 60 on the right. The histogram reveals a distribution of delays that peaks at 0. The histogram is fairly smooth but shows discrete increases at certain values. As the next set of histograms will show, these discrete increases appear to reflect rounding by carriers who report their delay data manually. It is interesting to note that the spikes generally occur at five minute intervals (e.g. at -5, 0, 5, 10, etc…); however, instead of there being a spike at 15 minutes, the histogram shows a spike at 14

[9] These are based on the 1995 to 1998 sample.

minutes.[10] This could either reflect rounding (or lying) by carriers who report manually or effort by airlines to systematically reduce delays on flights that would otherwise have delays just above the threshold.

In Figures 2A through 2C, we compare the distribution of arrival delays for carriers who report their delays in different ways. Since we only know an airline's reporting type with certainty beginning in March 1998, these histograms only show delays for flights between March and December 1998. Figure 2A shows the distribution of arrival delays for American Airlines, Northwest Airlines, United Airlines and US Airways – all of which reported fully automatically during this period. Their histogram is smooth with a peak around -5 and no apparent spike at 14 minutes. Figure 2B shows the distribution of arrival delays for Southwest Airlines, Alaska Airlines and American West – all of which reported their on-time data manually during this period. This histogram is much less smooth, has a large spike at zero (with almost 10% of flights arriving with exactly zero minutes delay) and suggests that these airlines are rounding their delays at the five minute intervals (i.e.: 0, 5, 10, etc…). However, rather than a spike at 15 minutes – which would be consistent with the pattern – the histogram shows a spike at 14 minutes. Finally, Figure 2C shows arrival delays for Continental, Delta and TWA – the three airlines that used a combination of manual and automatic reporting. This histogram is quite smooth and looks much more like the histogram of the automatic reporters than the histogram of the manual reports – suggesting these airlines were likely reporting most of their data automatically. The histogram for these carriers - which includes the first two airlines to introduce an employee bonus program based on the DOT ranking - shows a distinct spike at 14 minutes.

[10] Much of this pattern is driven by Southwest Airlines, which schedules its flights to arrive on "the 5s" and appears to report many of its delays in five minute intervals.

In Figures 3A and 3B through Figures 7A and 7B, we compare the distribution of an airline's arrival delays before and after it introduces an employee bonus program. We do this for each of the five airlines that we observe introducing such a program. Figures 3A and 3B show arrival delays for Continental in the two years before and two and a half years after the introduction of its employee bonus program.[11] These histograms suggest a marked increase in the number of flights that arrive exactly 14 minutes late and a decrease in the number of flights that arrive 15 or 16 minutes late after the introduction of the bonus program. Figures 4A and 4B plot analogous histograms for TWA. These figures show a very similar pattern. After the introduction of TWA's program, there is an obvious discontinuity in its distribution right around the relevant threshold, with 14 minute delays being more than twice as likely as 15 minute delays. For both Continental and TWA, the difference in the percentage of flights delayed 14 minutes compared to 15 minutes is much larger after the introduction of the bonus program than before and also much larger than any other difference observed anywhere else in their distributions.

Figures 5A and5B plot the arrival delay distribution for American Airlines one year before and one year after the introduction of its bonus program. The figures show a very small discontinuity around the 15 minute mark which is much less pronounced that the discontinuity in the first two sets of histograms. The analogous figures for US Airways and United Airlines before and after the introduction of their programs show no apparent in the relative heights of the bars at 14 and 15 minutes.

IV. Empirical Approach

IV.A. Overview of Empirical Approach

[11] We add data from 1993 for this histogram so that we can have two years of pre-bonus program data.

We define gaming as a systematic effort by an airline to reduce delays on specifically those flights that it expects to arrive with a delay of just over 15 minutes.[12] To empirically identify gaming, we need to be able to do two things. First, we need to be able to identify flights that an airline expects to be close to the 15 minute threshold. These flights are the most likely candidates for gaming since they are the flights that can presumably be brought below the threshold at the lowest cost. Second, we need to be able to measure whether the airline actually reduces delays on these flights below what they would otherwise have been. This requires a counterfactual measure of what a flight's delay would have been absent any incentive for gaming.

We believe that both of these requirements are met particularly well in our setting. Because our data allow us to observe the various stages of each flight – departure from the gate, take-off from the departure runway, landing on the arrival runway, and arrival at the gate – we can construct a flight's expected delay at each stage and, at any given stage, we can identify those flights whose expected delay is close to 15 minutes. We can then investigate whether – in *subsequent* stages of the flight - airlines attempt to reduce delays on specifically those flights that were expected to be around 15 minutes late. Furthermore, we have several ways of controlling for the counterfactual delay that these flights would have had in the subsequent stages absent the airline's incentive to game. First, we can look at flights just outside the critical threshold. That is, at a given stage of a flight, we can assume that – absent incentives to game – subsequent delays on flights that had expected delays of 15 minutes should be similar to subsequent delays on flights with expected delays of, say, 12 or 18 minutes. Second, we can compare flights with expected delays in the critical range to flights with very long expected delays (which we define

[12] The manipulation we focus on here is on effort spent in real-time (i.e.: once a flight is in progress) to reduce delays. This is distinct from manipulation that may occur in advance through what has been termed "schedule padding" – increasing schedule times for the purpose of appearing to be on-time.

to be delays over 25 minutes). If the costs of delays are convex, then the airline should have the greatest incentive to reduce delays on those flights. If we find that airlines make more effort to reduce delays on flights that they expect to arrive close to the 15 minute threshold than on flights that they expect to arrive with very long delays, this would strongly suggest that there is gaming. It is also worth pointing out that, in our setting, the flights that are candidates for gaming – i.e.: whose predicted delay is right around the critical 15 minute mark – will be identified in real-time and will vary from day to day. This means that airlines cannot engage in ex ante behavior that aims to reduce delays on those flights that it expects to arrive right around 15 minutes late since this is simply not known by the airline in advance. This eliminates selection concerns when comparing flights that are candidates for gaming to their "control groups" of flights outside the threshold.

IV.B. Taxi Time Regressions

Before describing our regression analysis in detail, it is useful to consider at what stages of a flight gaming may take place. Delays can be occurred at any of the stages of a given flight. In theory, an airline that is trying to systematically improve the on-time performance of a flight that it expects to arrive just above the 15 minute threshold could try to reduce delays during any of the phases. However, we expect that airlines will be more likely to try to reduce delays during the later stages of a flight. This is because, as the flight progresses, the airline knows the delay that has been incurred so far and therefore can more precisely predict the total delay the flight will have. For example, when a flight is airborne, the airline knows how delayed the plane was leaving the ground but must predict both how delayed it will be in the air and how delayed it will be while taxiing in. However, once a flight has touched down at the arrival airport, the airline knows how delayed the plane was leaving the ground and while in the air and must only predict how delayed it will be while taxiing in. For any given predicted level of delay, reducing the amount of noise associated with that prediction increases the likelihood that the airline's effort at reducing a flight's delay will actually result in the flight having a shorter delay. Based on this logic, our empirical analysis focuses on estimating an airline's effort to reduce delays during the final phase of the flight – i.e.: when it is taxiing in to its arrival gate – as a function of its *predicted delay* at the time that it touches down at the arrival airport.[13]

It is the richness of the DOT data and, in particular, the fact that in 1995 it began collecting information on wheels up and wheels down times which allow us to construct a fairly precise predicted delay measure. Specifically, to construct each flight's predicted delay at the time that its wheels touchdown, we take the flight's wheels-down time and add to it the median taxi-in time for that flight in the quarter.[14] This gives us a predicted arrival time for the flight. The difference between the predicted arrival time and the scheduled arrival time is the flight's predicted delay. For example, consider a flight by Delta Air Lines between Boston and Atlanta in March of 1997. Suppose that is has a scheduled arrival time of 4:30 pm. If its wheels-down time is 4:36 pm and Delta's median taxi-in time for this flight in this quarter is 4 minutes, then the flight's predicted arrival time is 4:40 pm and its predicted delay is 10 minutes.

We then construct a series of dummy variables for each level of predicted delay, in one minute increments. For example, we construct a dummy variable that equals one if a flight's predicted delay is greater than or equal to 10 minutes and less than 11 minutes. We construct another dummy variable that is equal to one if a flight's predicted delay is greater than or equal to 11 minutes and less than 12 minutes. And so on. Flights with predicted delays of greater than

[13] In addition, focusing on taxi-in times has the advantage that it minimizes the number of stages of a flight's progression that we need to predict thus eliminating noise from our measure of predicted delay. For example, were we to calculate a flight's predicted delay at the time that it departs from the ground, we would need to estimate both its airborne time as well as its taxi-in time.

[14] We identify a flight as a unique combination of airline, flight number, departure airport and arrival airport.

25 minutes are grouped together in the top category while flights with predicted delays of less than 10 minutes are used as the excluded group. Thus, we define 16 different predicted delay "bins". To investigate whether the employee bonus programs enhance the incentives to game that are inherent in the government program, we construct the predicted delay bins separately for airlines without bonus programs in place and for each airline with a program in place and, where possible, distinguish between the years before and years after its program was in place. Thus, for the 1995-1998 sample which covers the first two bonus programs, we construct predicted delay bins for four mutually exclusive sets of flights: (1) flights by the five carriers in our data that do not have a bonus program in place during the time period; (2) flights by Continental after the introduction of its bonus program (which is introduced in the second month for which we have taxi-time data); (3) flights by TWA before the introduction of its bonus program; and (4) flights by TWA after the introduction of its bonus program. This means that we have a total of 64 mutually exclusive dummy variables in these models.

We estimate a flight level equation that regresses a flight's taxi-in time, in logs, on these 64 dummy variables, carrier-airport-day fixed effects and a set of control variables which includes a dummy for the departure airport being a hub, controls for two distance categories (500-1500 miles and greater than 1500 miles), and dummies for each (actual) arrival hour. One can think of the model as estimating four vectors of 16 parameters, one for each of the four groups of flights defined above. Within these vectors, each coefficient represents the change in the log(taxi-in time) for flights in a given predicted delay bin relative to the taxi-in time for flights with predicted delay of less than 10 minutes. Because we include carrier-airport-day fixed effects, our coefficients are estimated using variation in predicted delays across an airline's flights that arrive at a given airport on a given day. This variation results from differences in the delays that flights incur *prior* to arrival which will largely be driven by factors at the flights' respective departure airports and in the air. Our primary interest is in testing whether those flights with predicted delay right around the critical 15 minute threshold have systematically shorter taxi times than flights that are above or below the threshold and whether this relationship is affected by the introduction of an employee bonus program. The key identifying assumption of the model is that there are no observable factors that are correlated with a flight having a predicted delay in the threshold range and that affect the flight's taxi-in time. Because evidence of gaming would come from a non-monotonic relationship between predicted delay and taxi time, we can rule out most other possible sources of correlation between predicted delay and taxi time since these are not likely to result in the same non-monotonic pattern.

V. Results

V.A. Taxi-Time Regressions

Our main set of taxi-time results are presented in Tables 3A and 3B. Table 3A shows the results for the two early bonus programs while Table 3B shows the results for the three later programs. Each column of the table represents the coefficients on the 16 predicted delay bins for a particular set of flights. We begin by describing the results in Table 3A. The first column represents the coefficients for airlines without bonus programs, the second column represents the coefficients for Continental, the third column represents the coefficients for TWA prior to the introduction of its bonus program and the final column represents the coefficients for TWA after the introduction of its bonus program. In order to look for evidence of gaming, we perform three hypothesis tests for each group. Specifically, we (separately) test if the coefficient on the 15-16 minute bin is significantly larger in magnitude than the coefficients for the 12-13, 18-19 and 25 and over bins, respectively.

The results in the first column show no evidence of gaming by airlines that do not have bonus programs in place. Flights that are predicted to arrive just above the critical threshold have about 3.5% shorter taxi-in times than flights that are predicted to be less than 10 minutes late; however, flights at every higher level of predicted delay also have taxi-in times that are between 3.5%-5% shorter than those for flights with predicted delays of less than10 minutes. Our hypothesis tests show that the coefficient on the 15-16 minute bin is significantly larger in magnitude than the coefficient on the 12-13 minute bin, but it is significantly smaller in magnitude than the coefficient on the 25 minute and over bin and not significantly different from the coefficient on the 18-19 minute bin.

In contrast, the results for the first two carriers that implemented bonus programs show a different pattern. Looking first at Continental Airlines, its flights with predicted delays of 15 to 16 minutes have taxi-in times that are 14 percent shorter than those of flights with predicted delay of 10 minutes or less. Its flights with predicted delays of 16 to 17 minutes have taxi-in times that are about 14.5 percent shorter. Moreover, the coefficients indicate a non-monotonic relationship between taxi-in times and a flight's predicted delay. Flights with predicted delays above or below the critical range have much smaller coefficients (i.e.: longer taxi-in times) than flights that are within the critical range. All three of our hypothesis tests indicate that the coefficient on the 15-16 minute bin is larger in magnitude than the other coefficients we test it against. Given an average taxi-in time of about 6 minutes, the coefficients we estimate for flights in the critical range translate into average reductions in taxi-in times of about 50 seconds. While this magnitude may appear small, our simulations below reveal that these selective reductions in delay can add up to meaningful changes in on-time performance.

The estimates for TWA after the introduction of its bonus program show a very similar pattern for flights near the 15 minute threshold, with magnitudes that are slightly larger than those estimated for Continental. While we cannot reject equality of the 15-16 minute and the 18-19 minute coefficients, we find that the 15-16 minute coefficient is significantly larger in magnitude than both the 12-13 and the 25 minute and over coefficients. Since TWA's program was introduced in 1996, we are able to separately estimate the relationship for TWA before and after its program is in place. As the third column of the table indicates, we see no systematic evidence of gaming by TWA prior to the introduction of its program. Figures 8A and 8B contain plots of the coefficients for Continental and TWA after their programs are in place. The non-monotonic relationship is very apparent in these plots.

Table 3B shows the results for the airlines that introduced bonus programs in 2003 and later. In the first two columns we show the results for American Airlines and US Airways after they introduced their bonus programs (estimated on the 2002 to 2006 sample). The third column shows the results for United Airlines after it introduced its program (estimated on the 2008-2010) sample. As above, we also include predicted delay dummy variables for these airlines pre-bonus as well as for the other carriers that did not introduce bonus programs during this period. However, because of space constraints, we only present the post-bonus results in the table. None of the columns show any indication that these programs resulted in gaming as we have defined it. The coefficients on predicted delay bins in the threshold range are very similar in magnitude to or smaller than the coefficients on predicted delay bins above the critical range. In the case of United's program, there is no evidence that taxi-in times for flights in the critical range are any different than taxi-in times for flights that are predicted to be less than 10 minutes late. Thus, while we find strong evidence of gaming following the introduction of Continental's and TWA's bonus programs, we do not find similar evidence of gaming following the introduction of American's, US Airways' and United's programs. As described earlier, we suspect that this is due to the fact that the three later provides provided much weaker incentives to employees as the programs were structured in such a way that the likelihood of actually earning the bonus was quite low.

V.B. Does it Work?

The results in Table 3A suggest that airlines are trying to improve the on-time performance of specifically those flights that would otherwise arrive just above the threshold for being on-time. In Tables 4A and 4B, we investigate whether they are successful in doing so.[15] We do this by estimating the probability that flights with predicted delay between 15 and 16 minutes arrive exactly one minute early and compare this to the probability that flights with other levels of predicted delay arrive exactly one minute early. Again, we are looking for a discontinuous relationship right around the relevant threshold. Since our predicted delay measure is not necessarily an integer but the actual delay variable in the data is, we define a flight as arriving exactly one minute earlier than predicted if its actual delay is the integer below its predicted delay (e.g.: a flight that is predicted to have 17.6 minutes of delay would be considered to arrive exactly one minute early if its actual arrival delay was 16 minutes). We regress a dummy variable that equals one if a flight arrives one minute earlier than predicted on the same expected delay dummies and controls as in Table 3A.

The results are presented in Table 4A. As before, each column displays the 16 coefficient estimates for one of the four different groups of flights and we run three separate hypothesis tests for each of these groups to look for evidence of gaming. Consistent with the results presented in Table 3, the estimates in the first column of Table 4A do not suggest gaming by airlines without bonus programs. The results for Continental and TWA in columns 2 and 4, respectively, are again consistent with efforts to systematically reduce delays on flights that

[15] Given that the results Table 3B – as well as the raw data in the histograms presented above - suggest that the later programs did not induce gaming, we restrict our subsequent empirical analyses to Continental and TWA programs.

would otherwise arrive around the threshold for being considered on-time. For Continental and TWA, after the introduction of their bonus programs, their flights with predicted delays between 15 and 16 minutes are 11 percentage points and 9 percentage points, respectively, more likely to arrive exactly one minute earlier than predicted, relative to their flights with less than 10 minutes of predicted delay. For both of these carriers, no other level of predicted delay has a coefficient that is in this range.

In Table 4B, we re-estimate this regression using (as the dependent variable) a dummy variable that equals one if a flight arrives exactly two minutes earlier than expected. The results of this exercise are again consistent with these two airlines attempting to systematically reduce delays on flights that would otherwise arrive just above the threshold for being on-time. For both Continental and TWA, flights that are predicted to be between 16 and 17 minutes late (i.e.: arrive 2 minutes after the cutoff for being considered on-time) are more than 13 percentage points more likely to arrive two minutes sooner than predicted than flights with predicted delay of less than 10 minutes. This effect is again substantially larger than it is for flights with any other level of predicted delay. Note that the results in Tables 4A and 4B are also consistent with what is observed in Continental's and TWA's histograms after they introduce their bonus programs – an increase in the fraction of flights that arrive exactly 14 minutes late.

V.C. Manual vs. Automatic Planes

All of the results presented so far indicate that, after introducing their employee bonus programs, Continental and TWA systematically try to reduce delays on those flights that might otherwise arrive right around the 15 minute threshold. However, as discussed in Section II, we believe that, during our sample period, both of these airlines had some number of aircraft that reported on-time data manually. This raises the possibility that what we are measuring as shorter taxi-in times are simply airline employees lying about the arrival times of flights that would have arrived 15 or 16 minutes late.[16] This would still represent a form of "gaming" of the incentive program; however, it would represent a different type of gaming than actual reductions in taxi-in times. In addition, the welfare implications would be different.

The fact that the histograms for Continental and TWA look much more similar to the histograms for the automatic reporters than the histograms for the manual reporters suggests that most of these two airlines' planes are likely to be reporting automatically. However, we have also developed an approach that tries to identify specifically which aircraft may be reporting manually. We exploit the fact that we can track planes in our data by tail number. We look for evidence that some of the planes of combination reporters appear to be rounded in a way that is similar to how the manual reporters appear to round their delays at zero. Specifically, for each aircraft in each year of our data, we calculate the fraction of its flights in that year that have a reported arrival delay of zero. We then compare the distribution of this plane-year level variable across airlines which report their on-time data in different ways.

Table 5 shows the distribution of this variable for all 10 airlines who report to the DOT in 1996. The 99th percentile of the distribution of this variable for American Airlines – which we expect reported fully automatically in 1996 – is 0.0509 which indicates that only about 1 percent of American's planes were reported to arrive with a delay of zero minutes more than 5% of the time. In contrast, for America West which was a manual reporter during this time, 50% of its planes landed with a reported delay of zero more than 5% of the time. Southwest is clearly an outlier here with the 50th percentile of its distribution being 11.72%, far higher than any other airline's. If we compare Continental and TWA to the carriers that we expect are fully automatic

[16] In our data, taxi-in times are calculated as the difference between arrival times and wheels down times. As a result, given a plane's wheels down time, if its arrival time at the gate is recorded as one minute earlier than it actually was, this would appear in our data as a one minute shorter taxi-in time.

in 1996, we see that TWA's distribution is very close to the automatic reporters while Continental's planes are more likely than the automatic reporters to have reported delays of zero. Based on this table, we categorize planes that have reported delays of zero more than 5% of the time to be manual planes.

Using this approach for identifying manual aircraft, we re-estimate our earlier regressions with separate predicted delay bins for Continental's and TWA's manual and automatic planes. This allows us to investigate whether the patterns we estimated earlier were being driven by planes that we suspect reported manually and where lying may be taking place. Rather than present the results of this exercise in additional tables, we present plots of the coefficients of interest. The coefficients from the taxi-time regression are presented in Figures 9A and 9B while the coefficients from the "arrive 1 minute early" regression are presented in Figures 10A and 10B. The coefficients in these figures show that the non-monotonic relationship between taxi-in times and predicted delays exists for both manual and automatic planes. However, the pattern is more pronounced and the difference in taxi-in times for threshold flights is greater for manual planes. Our hypothesis tests again suggest evidence of gaming for Continental and TWA after the introduction of their bonus programs. Of the six hypotheses that we test, the only hypothesis test we reject is that the 15-16 minutes coefficient for TWA is greater than its 18-19 minute coefficient.

We have tested the robustness of our definition for identifying manual planes by using an alternative definition which is based on rounding of flight delays throughout the distribution, not just at zero. Specifically, we compute the percentage of a plane's flights during a year that have a reported arrival delay that is either equal to 0 or is equal to a number that falls on the five minute intervals, excluding 15. Based on the distribution of this variable for automatic reporters, we define planes as manual if their flights are reported to arrive with a delay of zero or a multiple of five more than 20 percent of the time. This alternative definition has a strong overlap with the definition based zero delay and the results are robust to using this alternative definition.

As a final check of our main definition of manual planes, we have tested it on Continental's planes in the period after Continental had switched to fully automatic reporting of delays. We find that our definition identifies about three percent of Continental's automatic planes as "manual" during that time period which is similar to the fraction of planes that arrive with zero delay more than five percent of the time for the automatic reporters on which the definition was based.

V.D. Analysis of Paired Flights

The results in Table 3A clearly suggest that airline employees are systematically shortening taxi-in times for flights that arrive close to the 15 minute threshold. The identification strategy used in those regressions exploits variation in delays incurred prior to arrival across a carrier's flights arriving at the same airport on the same day. While this identification strategy should be fairly convincing given that it is difficult to think of an unobservable factor that would be correlated with predicted delays and generate the particular relationship between predicted delays and taxi-in times that we find, we nonetheless carry out an additional analysis of taxi-in times that controls even more carefully for possible unobservable factors that may lead to differences in taxi-in times across flights. Specifically, we consider pairs of flights by the same airline that land at the same airport at the precisely the same time.[17] We focus on pairs in which at least one of the flights lands with an expected delay of 25 minutes or more. We construct a variable that equals one if the "late" flight (i.e.: the one that lands with predicted delay of more than 25 minutes) has a shorter taxi-in time than the "early" member of the pair. We relate this

[17] The BTS data rounds arrival times to the nearest minute. Thus, we can only be certain that tied arrivals do not deviate in their true arrival times by more than one minute.

variable to the predicted delay of the early member of the pair by regressing it on the same expected delay bins used in the analysis above. Intuitively, what we are doing is estimating whether the probability that a very late flight has a shorter taxi-in time that an earlier flight that arrives at the exact same time depends on whether the earlier flight is close to the critical threshold. The benefit of this empirical exercise (relative to the earlier regressions) is that if there is some unobservable that is correlated with both the likelihood of a flight having expected delay in the threshold range and that flight's taxi-in time when it arrives, this unobservable should equally affect the threshold flight and the flight with which it is paired because that flight lands at the exact same time.

The results of this exercise are presented in Table 6. Each column again presents the coefficients for one of the four groups of flights that we distinguish. Each coefficient represents the probability that the "late" member of the pair has a shorter taxi time than the "early" member of the pair when the "early" member's expected delay is in the particular bin. The coefficients are relative to the probability that the "late" member has a shorter taxi time when it is paired with a flight with predicted delay less than 10 minutes. The estimates for Continental indicate that, relative to when the late flight lands with a flight that is predicted to be less than 10 minutes late, there is a significant reduction in the probability of the late flight "winning" when it lands at the exact same time as a flight that is predicted to be 14 to 15 or 15 to16 minutes late. While it is reasonable to expect that the probability that the late flight wins falls with the expected delay of its pair, one would expect to observe a monotonic relationship and this is not what the results show. The probability of the late flight having the shorter taxi time is lowest precisely when it is paired with a flight in the critical range. Interestingly, TWA's flights show this pattern prior to the introduction of its bonus program but not after. We are in the process of investigating what may be driving this result for TWA. Perhaps operational changes or changes in scheduling at its hub (where we are most likely to observe more than flight land at the same time) are influencing taxi-in times.

V.E. Externalities (Preliminary Results)

We have also begun an analysis of whether the selective reduction of delays on threshold flights imposes externalities on other flights. Such externalities will occur if – as a result of gaming – scarce resources are reallocated from other flights to threshold flights. They will not occur if airlines simply lie about the delays of threshold flights – as we suspect may happen with the manual planes – or if gaming is achieved through higher levels of effort from slack resources (e.g., ground crew). We should also point out that any externalities that do occur from reallocation of scarce resources will be inherently difficult to detect. This is because resources are scarce during times when the carrier has many flights arriving at the airport, but any threshold flight may only affect a small number of these flights. Since we do not know which of these flights will be affected (and it is likely that it is not always the same flight that is affected) and since we do not know whether arrival or departure delays are affected – we necessarily have to look for average effects which will be hard to detect.

Based on our analysis of this so far, we have not been able to uncover any externalities beyond the effects on paired flights described above. We have run regressions where we relate a flight's arrival delay to the percentage of other flights by the same carrier that arrive at the same airport during the same 15 minute time block that are threshold flights. We do not find that increases in the fraction of threshold flights that land within the 15 minute window that a flight lands has an effect on that flight's arrival delay.

V.F Additional Results and Robustness Checks

We have explored the robustness of our results to two alternative ways of estimating the taxi-in time that is used to calculate a flight's predicted delay. Specifically, instead of computing the median taxi time for a given flight in a given quarter, we have computed the median taxi-in time for a carrier at a given airport in a given month as well as the median taxi-in time for a carrier at a given airport in a given month during arrival time window. The results are robust to these alternative ways of calculating a flight's expected delay.

We have also re-estimated our regressions on a few subsamples of the data in order to explore whether the results differ across these samples. First, we have created separate samples for flights that arrive at a carrier's hub and flights that do not arrive at its hubs. We find evidence of gaming by Continental and TWA in both subsamples. We also find that flights with long expected delays have shorter taxi-in times (relative to flights with expected delays under ten minutes) in the hub sample than in the non-hub sample. This is consistent with the fact that long delays are more costly at hubs, where many more passengers make connections.

Second, we have created subsamples of flights that arrive at times of day where congestion at the arrival airport is above and below the median, respectively. Depending on whether the primary mechanism through which gaming occurs is the reallocation of scarce resources (during congested times) or a higher level of effort from otherwise slack resources, such as ground crew (during uncongested times), gaming may either be more or less prevalent for flights during congested times, compared to flights during uncongested times. We find evidence of gaming in both subsamples for Continental, but only for flights during uncongested times for TWA, suggesting that, for TWA, the primary source of gaming is a higher level of effort from slack resources. Finally, we have explored whether there may be end-of-the-month effects – specifically, whether gaming takes place at the end of months in which the airline is close to achieving the necessary ranking for a bonus payment, but not at the end of months in which the carrier is far away from achieving that target. Similar types of effects have been found in the prior literature on employee bonus programs. Note that, in order for such effects to occur in our setting, employees would have to be informed not only about their own airline's overall on-time performance in the month so far, but also about the on-time performance of all other carriers. The Department of Transportation only releases this information with a two-month lag, so that the information would have to come from other sources. We find no evidence of end-of-the-month effects, which suggests that airline employees may not have the necessary information to distinguish the months in which the airline is close to achieving the bonus target from months in which it is not.

V.G Simulation of Rankings

To investigate whether the distortions in taxi-in times that we find in our regression analysis can actually impact airlines' overall on-time performance and DOT rankings, we perform a counterfactual simulation that estimates what arrival delays and rankings would be absent gaming. To do this, we take the following approach. Our data suggest that taxi-in times are distributed approximately log-normal. We calculate the mean and variance of the log taxi-in time for each carrier-airport-month. Then, for each flight in our data, we replace the actual taxi-in time in the data with a random draw from a log-normal distribution with the mean and variance for the appropriate carrier-airport-month. The idea behind this exercise is to replace a flight's taxi-in time with the taxi-in time it would likely have absent any incentive for the airline to systematically reduce taxi-in times on threshold flights. After doing this exercise for every flight in our data, we can recalculate the fraction of flights that are 15 or more minutes delayed. This leads to counterfactual measures of on-time performance for each airline and these can be used to create counterfactual rankings of airlines. Repeating the simulation a number of times yields standard errors for our simulated on-time performance measures.

We report results from the counterfactual exercises in Tables 7A and 7B. Table 7A shows simulated changes in on-time performance and ranking for Continental in the months after the introduction of its bonus program. Table 7B shows the same thing for TWA. Averaging across months, the difference between actual and simulated on-time performance for Continental is about one full percentage point – that is, the distortion in taxi-in times results in the fraction of flights delayed 15 minutes or more falling by one percentage point. The difference is about 1.3 percentage points for TWA after it introduces its program. These changes in the fraction of delayed flights directly map into changes in rankings. For example, when we simulate Continental's taxi-in time but leave the others carriers' behaviour unchanged, we find that the taxi time distortions result in Continental achieving an improvement in rankings of at least one position in 19 of the 35 months following the introduction of their program. When we simulate Continental as well as all other airlines' taxi-in times, we find that the taxi-time distortions result in Continental achieving an improvement in rankings in 8 of the 36 months. Thus, the results of the simulation exercise indicate that while a 45 to 55 second reduction in delay may be small in absolute value (and in terms of the disutility to consumers), when applied to flights that are close to the relevant threshold, the impact on reported rankings can be significant.

V.G. Are There Any Real Effects of the Bonus Programs?

The results so far suggest that part of the improvements in Continental's and TWA's on-time performance after the introduction of their bonus programs resulted from gaming behaviour. One might question whether these programs - and the other operational and/or managerial changes that accompanied them - resulted in any actual improvements in on-time performance.

Using a very different empirical strategy and different data than us, Knez and Simester (2001) investigate the impact of Continental's program and find that it resulted in a significant improvement in on-time performance measured by the fraction of flights that depart less than 15 minutes late. Since airlines have no incentive to manipulate departure delays, these results would indicate an actual improvement in on-time performance.

In Appendix A we estimate the relationship between the introduction of the bonus programs and several different measures of on-time performance. Using our sample of flights from 1994 to 1998, we estimate a flight-level regression that includes airline and arrival-airport date fixed effects. To estimate whether on-time performance differed after the introduction of the bonus program, we interact the Continental dummy with a variable that equals one in months in which its bonus program is in effect. We do the same with the TWA dummy. Time trends are captured in a very flexible way by the arrival-airport date fixed effects. The results indicate that, in the months after the introduction of its bonus program, Continental's mean arrival was lower by about 2.4 minutes, its likelihood of arriving 15 or more minutes late fell by about 4.8 percentage points, its taxi-in times were on average 0.6 minutes shorter, its departure delays were 1.8 minutes shorter and its taxi-out times were not changed. The results for TWA are roughly similar.

There are a couple of interesting things to note from this table. First, there is evidence of at least some real improvement in on-time performance. Continental's flights are, on average, departing 1.8 minutes less delayed after the introduction of the program. TWA's departure delays are about one minute shorter. Second, the estimates in also suggest the presence of gaming. Specifically, one can take the estimated change in arrival delays – 2.4 minutes – and apply it equally to all of Continental's flights in 1994 (i.e.: reduce each flight's delay by 2.4 minutes). Based on this, one would predict that the fraction of flights with delays of 15 minutes or more would fall by about 3 percentage points, which is less than the 4.8 percentage points estimated in the second column of the table. The same is true for the estimates on TWA's program. Thus, the findings in these fairly descriptive regressions are consistent with the findings from the more nuanced analysis above.

VI. Conclusion

Prior research has shown that while disclosure programs may induce firms to improve product quality, there is also considerable effort by firms to game the schemes under which they are rated. As a result, those designing disclosure programs must try to anticipate the potential for a given scheme to be gamed. However, the potential for gaming of a disclosure program will depend not only the structure of the program but also on the characteristics of the product being rated and the incentives in place at the firm. In this paper, we have begun to explore these issues in the context of airline reporting of on-time performance. While the structure of this program creates obvious incentives for airline to game by selectively reducing delays on flights that would otherwise arrive with 15 minutes of delay, those flights cannot be identified in advance and so gaming must take plane in real-time. Whether such gaming will take place depends on whether those individuals who are in the position to reduce delays on select flights have incentives to do so.

Our empirical analysis finds no evidence of gaming by airlines without explicit employee bonus programs in place and no evidence of gaming by airlines with bonus programs that set targets that cannot realistically be achieved. On the other hand, our empirical analysis finds very strong evidence of gaming by the two airlines who introduced bonus programs with targets that could plausibly be achieved. We find that those airlines have systematically shorter taxi-in times for their flights that are predicted to arrive with 15 or 16 minutes of delay. These flights are also much more likely to end up arriving with exactly 14 minutes of delay. Our analysis suggests that some of this represents lying about planes' arrival times while some represents actual reductions in taxi-in times. While the effects we estimate translate into about 50 second shorter taxi-in times, our simulations show that applying this reduction in taxi-in times to the "right" set of flights can result in meaningful changes in the rankings which is what the bonus programs are based on. This paper contributes the growing empirical literature on gaming of disclosure programs by explicitly considering how gaming is affected by the incentives provided to employees who are in a position to carry out the gaming.

References

Courty, P. and G. Marschke (2004), "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics* 22: 23-56.

Dranove, D. and G. Jin (2010), "Quality Disclosure and Certification: Theory and Practice", *Journal of Economic Literature*.

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite (2003) "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers." *Journal of Political Economy* 111: 555-88.

Jacob, B. (2005), "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago," *Journal of Public Economics*, 89(5-6): 761-796.

Jacob, B. (2007), "Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments", *NBER Working Paper* 12817.

Kenz, M. and D. Simester (2001), "Firm Wide Incentives and Mutual Monitoring at Continental Airlines," Journal of Labor Economics, 19(4): 743-772.

Larkin, I. (2007), "The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software Sales." *Unpublished manuscript*, Harvard Business School.

Lu, Susan F. (2009), "Multitasking, Information Disclosure and Product Quality: Evidence from Nursing Homes", *Working Paper*, University of Rochester (Simon School of Business).

Neal, D. and D. W. Schanzenbach (2010), "Left Behind by Design: Proficiency Counts and Test-Based Accountability", *Review of Economics and Statistics* 92(2), 263-283.

Oyer, P. (1998), "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality." *Quarterly Journal of Economics* 113:149-85.

Werner, R. and D. Asch (2005), "The Unintended Consequences of Publicly Reporting Quality Information," *Journal of the American Medical Association*, 293(10):1239-44.

Figure 1
Distribution of Arrival Delays
Ten Largest U.S. Carriers, 1994-1998

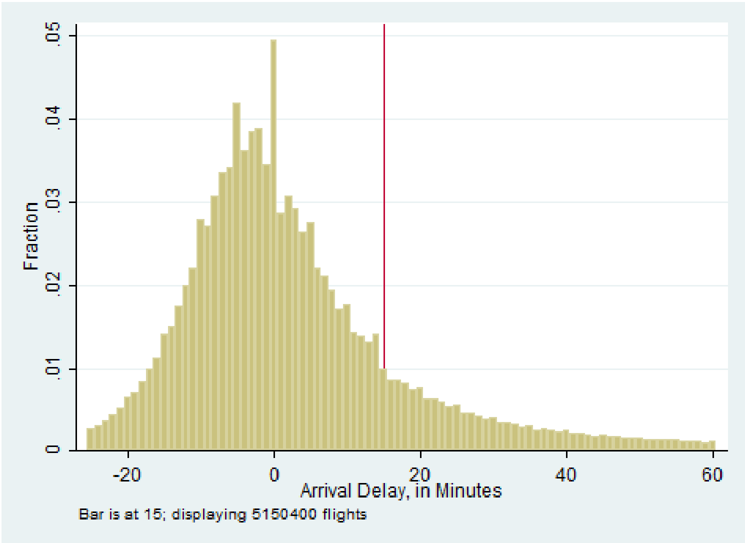


Figure 2A
Distribution of Arrival Delays
Fully Automatic Reporters, March – December 1998

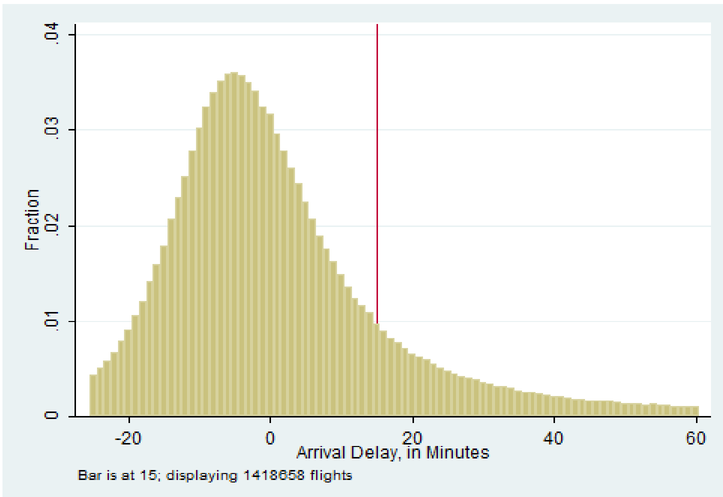


Figure 2B
Distribution of Arrival Delays
Manual Reporters, March – December 1998

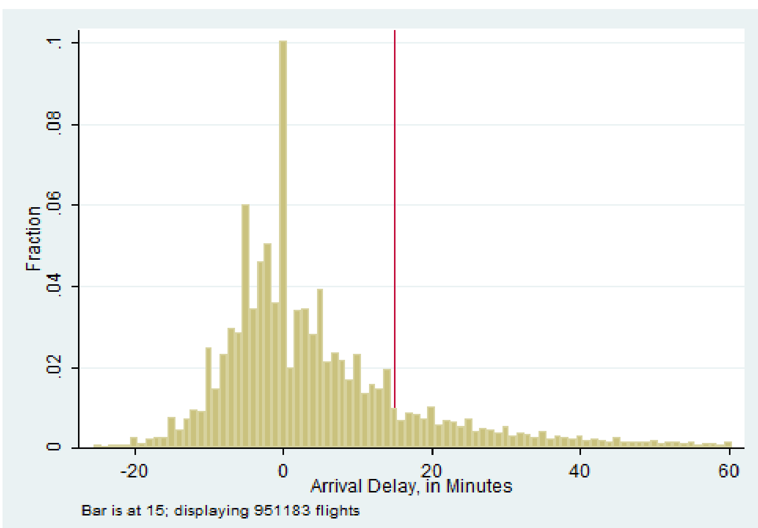


Figure 2C
Distribution of Arrival Delays
Combination Reporters, March – December 1998

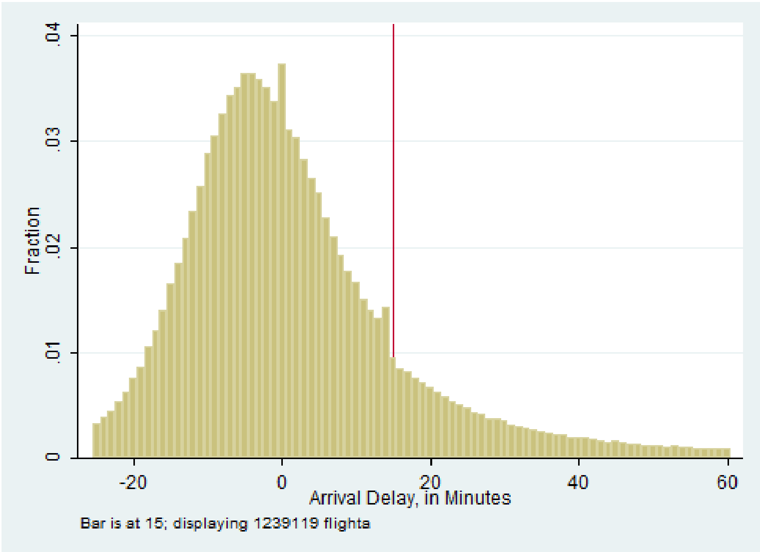


Figure 3A
Distribution of Arrival Delays
Continental Airlines, 1993-1994

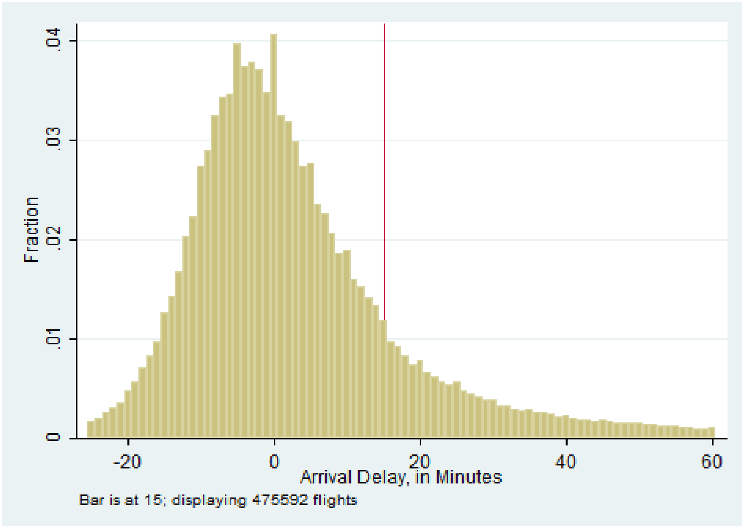


Figure 3B
Distribution of Arrival Delays
Continental Airlines, February 1995-1997

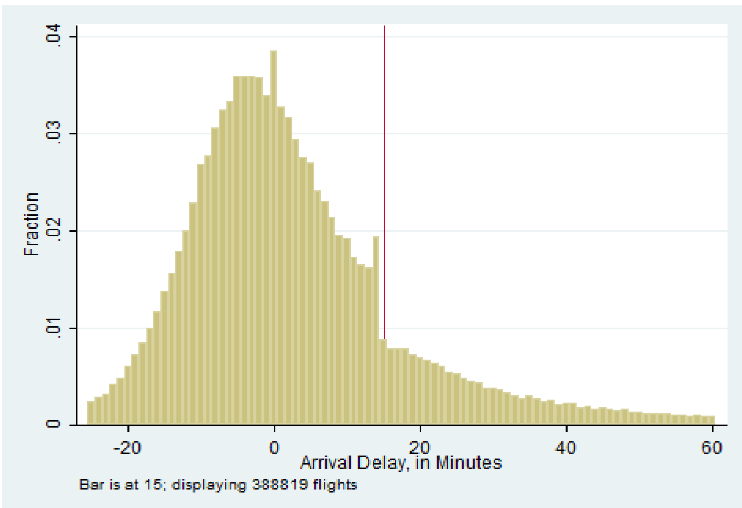


Figure 4A
Distribution of Arrival Delays
TWA, 1994-1995

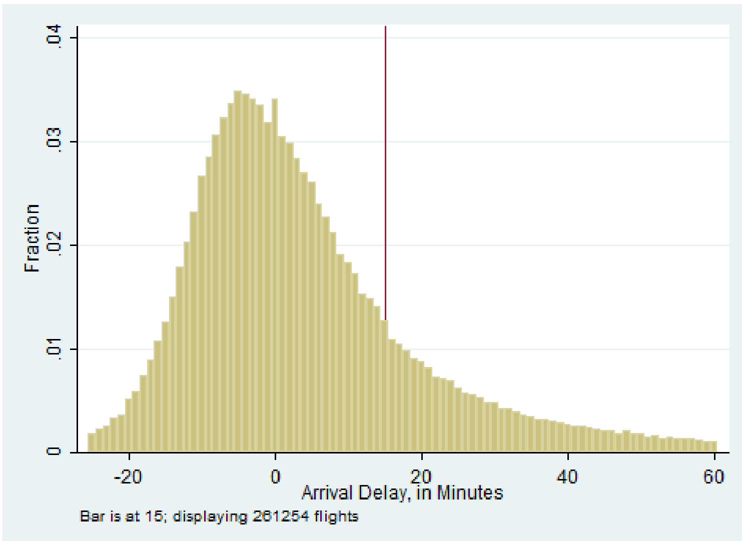


Figure 4B
Distribution of Arrival Delays
TWA, June 1996-1998

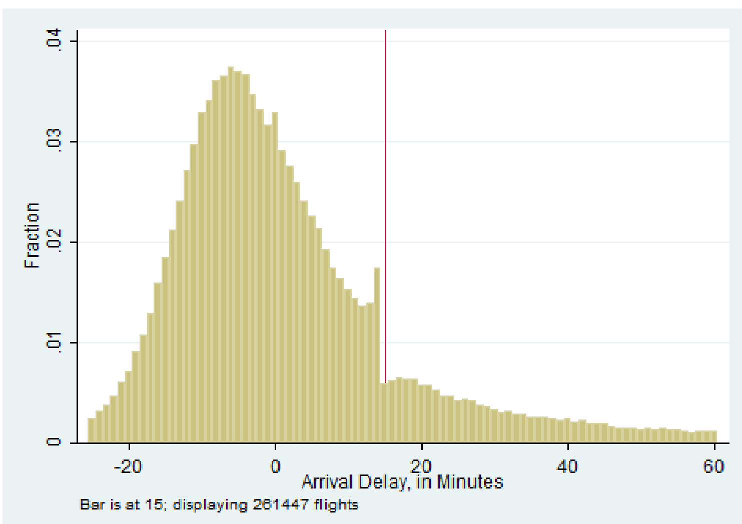


Figure 5A
Distribution of Arrival Delays
American Airlines, 2002

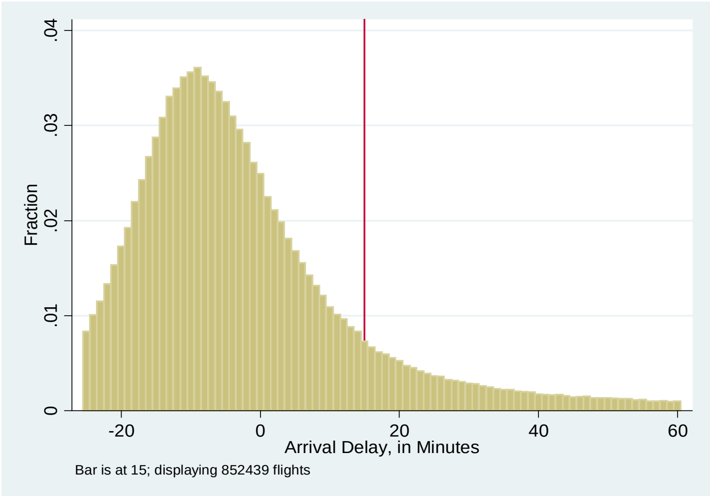


Figure 5B
Distribution of Arrival Delays
American Airlines, 2003

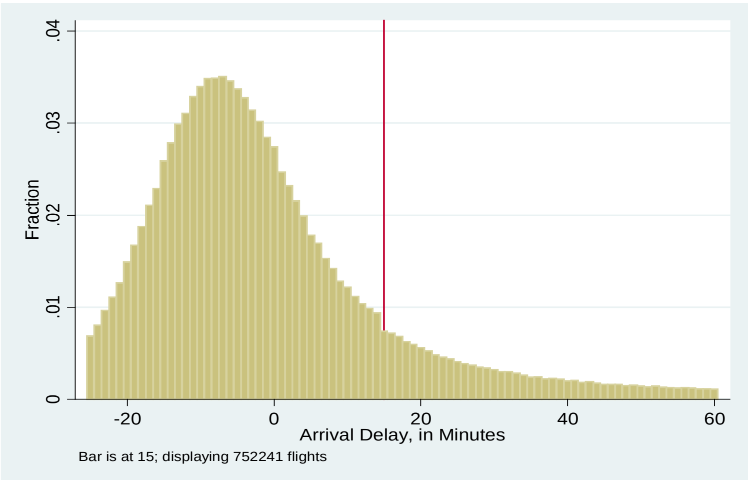


Figure 6A
Distribution of Arrival Delays
US Airways, 2004

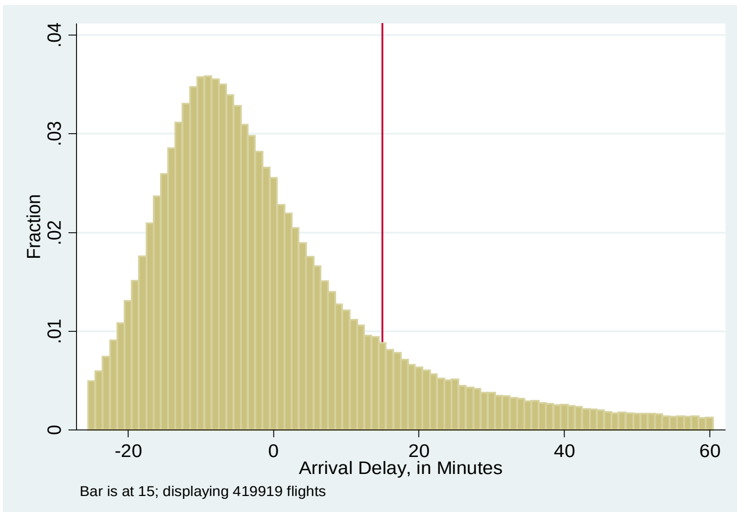


Figure 6B
Distribution of Arrival Delays
US Airways, 2004

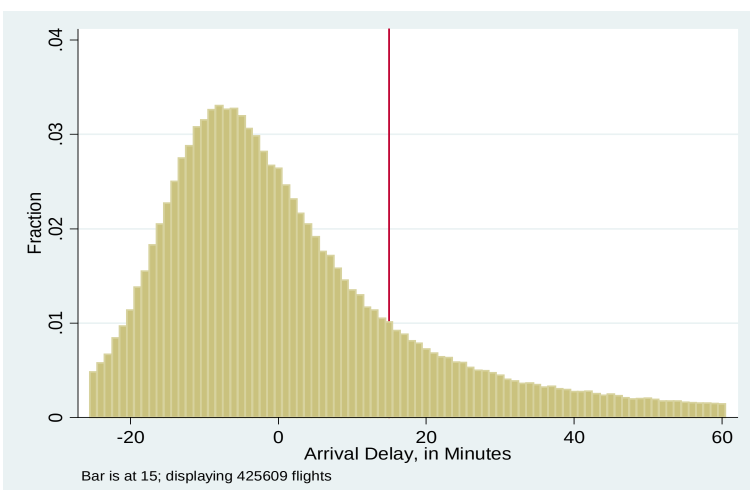


Figure 7A
Distribution of Arrival Delays,
United Airlines, 2008

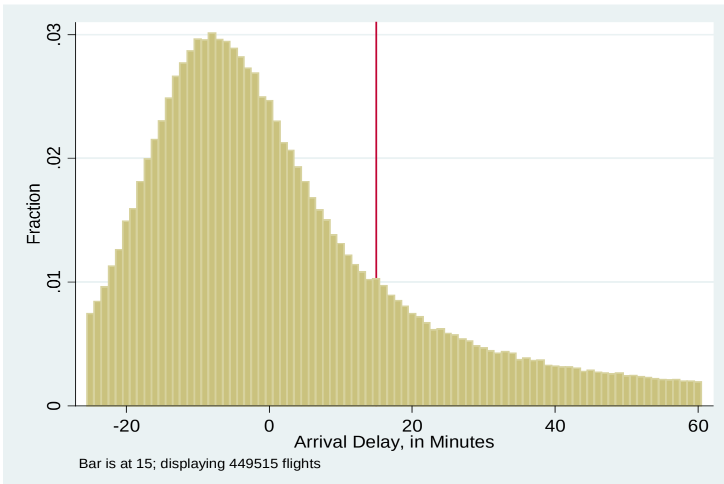


Figure 7B
Distribution of Arrival Delays
United Airlines, 2009

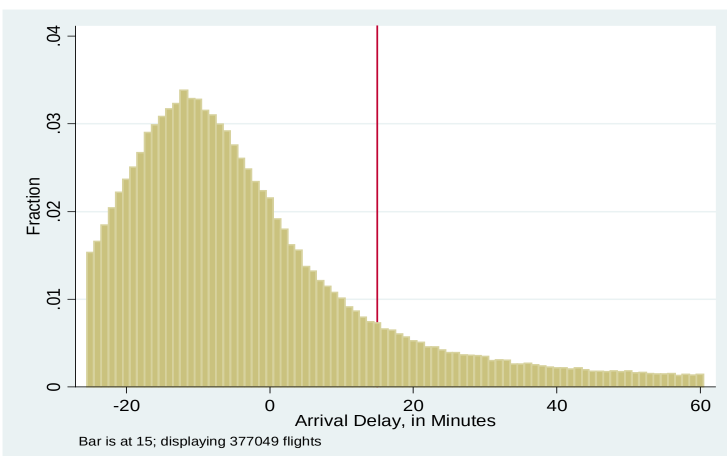


Figure 8A
Coefficients on Continental's Predicted Delay Bins (post-bonus)
(From Table 3)

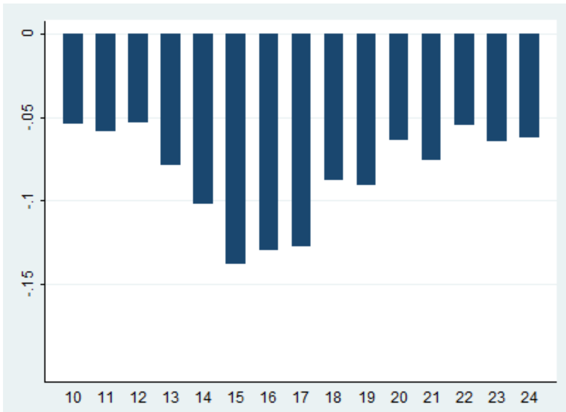


Figure 8B
Coefficients on TWA's Predicted Delay Bins (post-bonus)
(From Table 3)

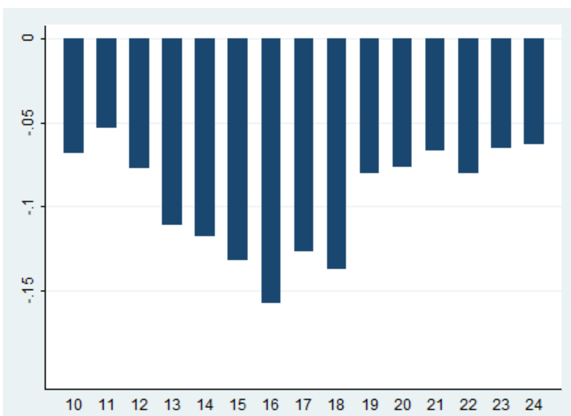


Figure 9A
Coefficients from Taxi-Time Regression
Continental's Predicted Delay Bins – Manual vs. Automatic Planes

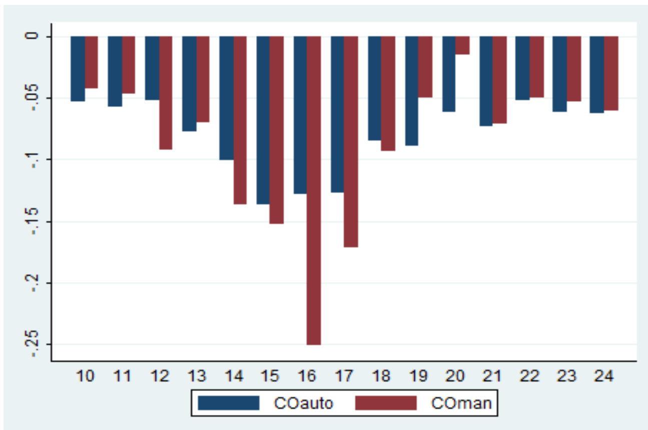


Figure 9B
Coefficients from Taxi-Time Regression
TWA's Predicted Delay Bins – Manual vs. Automatic Planes

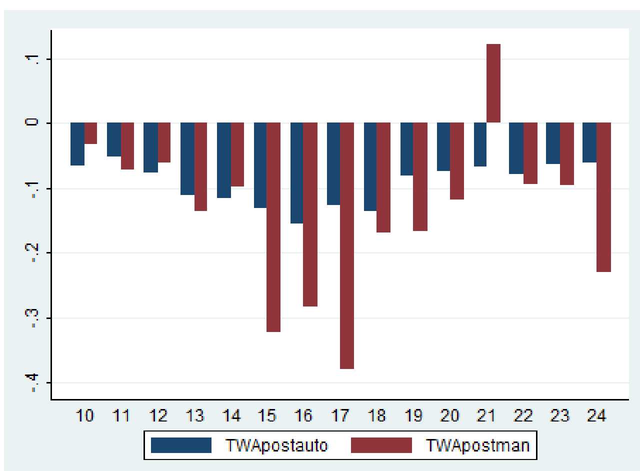


Figure 10A
Coefficients from 1 Minute Early Regression
Continental's Predicted Delay Bins – Manual vs. Automatic Planes

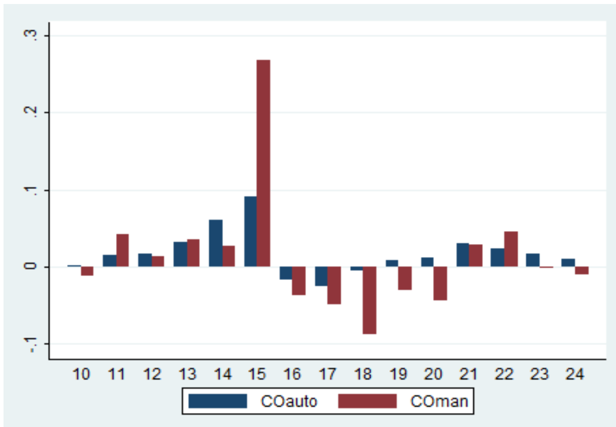


Figure 10B
Coefficients from 1 Minute Early Regression
TWA's Predicted Delay Bins – Manual vs. Automatic Planes

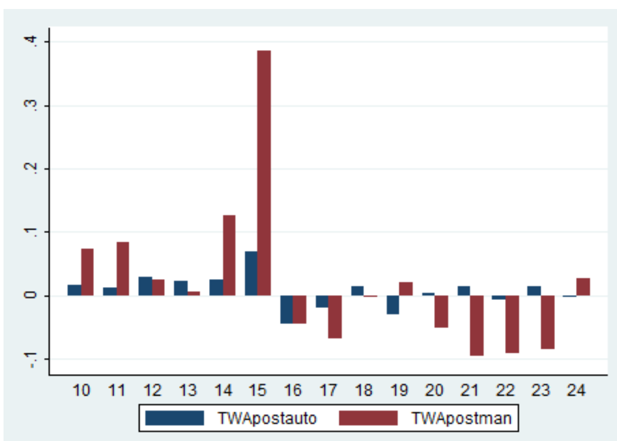


Table 1
Overview of Bonus Programs

Airline	Payment Structure	1 year prior to start of bonus program		1 year after start of bonus program		
		Average Rank	Average On-Time %	Average Rank	Average On-Time %	# Months Bonus Achieved
Continental (Start: Feb 1995)	Initially: $65 per employee in each month that the airline ranked among top 5. Since 1996: $65 for rank 2 and 3; $100 for rank 1.	7.1	80.2	3.4	81.4	10
TWA (Start: Jun 1996)	Initially: $65 per employee in each month that the airline ranked top 5 in on-time, baggage and complaints. $100 if it also ranked 1st in one of the categories. In 1999: $100 if on-time performance exceeds fixed threshold of 80%. In 2000: Seasonal targets: 85% summer, 80% winter.	8.1	74.2	5.7	74.6	4
American (Start: Apr 2003)	Initially: $100 per employee in each month that the airline ranked 1st. $50 in months that the airline ranked 2nd. Since 2009: Bonus based on internal metric that excludes delays that are not under the employees' control.	3.1	81.4	12	79.2	0
US Airways (Start: May 2005)	$75 per employee in each month in which the airline ranks 1st.	9.8	76.1	8.2	79.2	0
United (Start: Jan 2009)	$100 per employee in each month that the airline ranked 1st. $65 in months that the airline ranked 2nd.	14.7	71.6	6.8	81.0	1

Table 2
Summary Statistics for Regression Sample
February1995 - December 1998

	N	Mean	Standard Deviation	Min	Max
Arrival Delay (min)	3,067,533	7.22	27.99	-88	1182
Dummy for Arrive 15 Minutes Late or More	3,067,533	0.21	0.41	0	1
Taxi In Time (min)	3,067,533	6.10	3.92	1	60
Departure Delay (min)	3,067,533	8.43	25.43	-15	1185
Taxi Out Time (min)	3,067,533	14.91	7.44	1	60
Flight Time	3,067,533	108.7	66.50	20	632

Notes**:** Includes flights by American, Continental, Delta, Northwest, TWA, United, and US Airways.

Table 3A
Taxi Time as a Function of *Predicted* Delay, 1995-1998

Dependent Variable	***Log(Taxi In)***			
	Coefficient Estimates for:			
	All Other Carriers	**CO post-Bonus**	**TWA pre-Bonus**	**TWA post-Bonus**
Predicted Delay				
[10,11) min	-0.0218***	-0.0522***	-0.0587***	-0.0656***
	(0.00199)	(0.00553)	(0.0123)	(0.0108)
[11,12) min	-0.0201***	-0.0562***	-0.0373**	-0.0530***
	(0.00204)	(0.00566)	(0.0132)	(0.0106)
[12,13) min	-0.0235***	-0.0563***	-0.00858	-0.0757***
	(0.00212)	(0.00587)	(0.0142)	(0.0109)
[13,14) min	-0.0324***	-0.0772***	-0.0502***	-0.115***
	(0.00230)	(0.00621)	(0.0141)	(0.0119)
[14,15) min	-0.0310***	-0.105***	-0.0726***	-0.116***
	(0.00241)	(0.00660)	(0.0158)	(0.0133)
[15,16) min	-0.0346***	-0.140***	-0.0516**	-0.145***
	(0.00244)	(0.00707)	(0.0163)	(0.0133)
[16,17) min	-0.0390***	-0.144***	-0.0160	-0.165***
	(0.00254)	(0.00781)	(0.0162)	(0.0161)
[17,18) min	-0.0413***	-0.132***	-0.0648***	-0.140***
	(0.00265)	(0.00935)	(0.0178)	(0.0167)
[18,19) min	-0.0392***	-0.0874***	-0.0564**	-0.139***
	(0.00283)	(0.00929)	(0.0175)	(0.0179)
[19,20) min	-0.0405***	-0.0857***	-0.0764***	-0.0835***
	(0.00291)	(0.00880)	(0.0178)	(0.0174)
[20,21) min	-0.0467***	-0.0590***	-0.0609**	-0.0789***
	(0.00293)	(0.00862)	(0.0194)	(0.0171)
[21,22) min	-0.0363***	-0.0728***	-0.0721***	-0.0620***
	(0.00306)	(0.00877)	(0.0175)	(0.0157)
[22,23) min	-0.0411***	-0.0556***	-0.0645**	-0.0811***
	(0.00316)	(0.00892)	(0.0204)	(0.0180)
[23,24) min	-0.0436***	-0.0607***	-0.0938***	-0.0665***
	(0.00331)	(0.00930)	(0.0187)	(0.0183)
[24,25) min	-0.0425***	-0.0615***	-0.0886***	-0.0716***
	(0.00338)	(0.00982)	(0.0207)	(0.0172)
>25 min	-0.0489***	-0.0489***	-0.0841***	-0.0883***
	(0.00145)	(0.00366)	(0.00978)	(0.00846)

Notes: Standard errors are in parentheses and are clustered at the level of the arrival airport-day. Columns display coefficients from a single regression of taxi time on four sets of predicted delay "bins" that are defined to be mutually exclusive. Specification includes carrier-arrival airport-day fixed effects and arrival hour and hub controls. Coefficients represent the change in log(taxi time) relative to flights with predicted delay of less than 10 minutes. Calculation of predicted delay is described in the text on page 13. The regression contains 3,067,533 observations.

Table 3B

Taxi Time as a Function of *Predicted* Delay, 2002-2004 and 2008-2010 samples

Dependent Variable	***Log(Taxi In)***		
	Coefficient Estimates for:		
	American Airlines post-Bonus	**US Airways post-Bonus**	**United Airlines post-Bonus**
Predicted Delay			
[10,11) min	-0.0291***	-0.0206*	-0.0124
	(0.00665)	(0.0105)	(0.0143)
[11,12) min	-0.0351***	-0.0275**	-0.0343*
	(0.00654)	(0.0104)	(0.0139)
[12,13) min	-0.0486***	-0.0260*	0.000440
	(0.00699)	(0.0116)	(0.0147)
[13,14) min	-0.0467***	-0.0211	-0.0288
	(0.00735)	(0.0118)	(0.0170)
[14,15) min	-0.0507***	-0.0273*	-0.00304
	(0.00766)	(0.0115)	(0.0169)
[15,16) min	-0.0685***	-0.0363**	-0.00278
	(0.00781)	(0.0124)	(0.0170)
[16,17) min	-0.0521***	-0.0258*	-0.00686
	(0.00839)	(0.0130)	(0.0183)
[17,18) min	-0.0586***	-0.0306*	0.00393
	(0.00858)	(0.0138)	(0.0161)
[18,19) min	-0.0465***	-0.0403**	-0.0340
	(0.00843)	(0.0131)	(0.0188)
[19,20) min	-0.0762***	-0.0255	-0.0429*
	(0.00914)	(0.0133)	(0.0184)
[20,21) min	-0.0545***	-0.0376*	-0.0276
	(0.00994)	(0.0148)	(0.0174)
[21,22) min	-0.0564***	-0.0599***	-0.0428*
	(0.00970)	(0.0144)	(0.0215)
[22,23) min	-0.0601***	-0.0349*	-0.0304
	(0.0103)	(0.0149)	(0.0202)
[23,24) min	-0.0499***	-0.0644***	-0.0352
	(0.0103)	(0.0145)	(0.0201)
[24,25) min	-0.0755***	-0.0618***	-0.0302
	(0.0104)	(0.0158)	(0.0233)
>25 min	-0.0579***	-0.0617***	-0.0470***
	(0.00360)	(0.00512)	(0.00567)

Notes: Standard errors are in parentheses and are clustered at the level of the arrival airport-day. Columns display coefficients from regression of taxi time on mutually exclusive sets of predicted delay "bins" for individual carriers. This table only shows a selected set of coefficients: for carriers with bonus programs, after the introduction of the program. Columns 1 and 2 are based on data from 2003-2006. Column 3 is based on data from 2008-2010. Specifications includes carrier-arrival airport-day fixed effects and arrival hour and hub controls. Coefficients represent the change in log(taxi time) relative to flights with predicted delay of less than 10 minutes.

Table 4A
Probability of Arriving Exactly *One* Minute Earlier than Predicted, 1995-1998

Dependent Variable	*Arrives One Minute Earlier than Predicted*			
	Coefficient Estimates for:			
	All Other Carriers	**CO post-Bonus**	**TWA pre-Bonus**	**TWA post-Bonus**
Predicted Delay				
[10,11) min	0.00520*	0.000474	-0.0204	0.0185
	(0.00209)	(0.00624)	(0.0121)	(0.0101)
[11,12) min	0.00522*	0.0177*	0.00500	0.0160
	(0.00213)	(0.00686)	(0.0124)	(0.00987)
[12,13) min	0.00290	0.0158*	-0.00768	0.0279**
	(0.00224)	(0.00689)	(0.0132)	(0.0108)
[13,14) min	0.00673**	0.0312***	0.00412	0.0228
	(0.00235)	(0.00736)	(0.0144)	(0.0121)
[14,15) min	0.00997***	0.0560***	-0.0145	0.0318**
	(0.00247)	(0.00803)	(0.0148)	(0.0120)
[15,16) min	0.0101***	0.111***	0.0106	0.0888***
	(0.00257)	(0.00852)	(0.0157)	(0.0132)
[16,17) min	0.00769**	-0.0196**	0.00146	-0.0435***
	(0.00261)	(0.00760)	(0.0151)	(0.0118)
[17,18) min	0.00957***	-0.0274***	-0.0125	-0.0223
	(0.00272)	(0.00779)	(0.0155)	(0.0125)
[18,19) min	0.0128***	-0.0131	0.00905	0.0127
	(0.00285)	(0.00870)	(0.0174)	(0.0134)
[19,20) min	0.00896**	0.00288	-0.000275	-0.0292*
	(0.00295)	(0.00924)	(0.0180)	(0.0122)
[20,21) min	0.0127***	0.00856	0.0258	0.000948
	(0.00306)	(0.00998)	(0.0194)	(0.0147)
[21,22) min	0.00504	0.0302**	-0.00486	0.0109
	(0.00323)	(0.0102)	(0.0188)	(0.0153)
[22,23) min	0.0131***	0.0244*	-0.0230	-0.0119
	(0.00325)	(0.0102)	(0.0185)	(0.0150)
[23,24) min	0.00931**	0.0135	-0.0133	0.00964
	(0.00344)	(0.0105)	(0.0183)	(0.0161)
[24,25) min	0.00837*	0.00808	0.0411	-0.00246
	(0.00346)	(0.0108)	(0.0233)	(0.0170)
>25 min	0.00799***	0.00993***	-0.000805	0.00813
	(0.000916)	(0.00264)	(0.00555)	(0.00441)

Notes: Standard errors are in parentheses and are clustered at the level of the arrival airport-day. Columns display coefficients from a single regression on four sets of predicted delay "bins" that are defined to be mutually exclusive. Specification includes carrier-arrival airport-day fixed effects and arrival hour and hub controls. Coefficients represent the change in the probability of a flight arriving exactly one minute earlier than predicted relative to the probability of arriving exactly one minute earlier than predicted for flights with predicted delay of less than 10 minutes. Calculation of predicted delay is described in the text on page 13. The regression contains 3,067,533 observations.

Table 4B
Probability of Arriving Exactly *Two* Minutes Earlier than Predicted, 1995-1998

Dependent Variable	***Arrives Two Minutes Earlier than Predicted***			
	Coefficient Estimates for:			
	All Other Carriers	**CO post-Bonus**	**TWA pre-Bonus**	**TWA post-Bonus**
Predicted Delay				
[10,11) min	0.00876***	0.0249***	0.00725	0.00968
	(0.00151)	(0.00499)	(0.00949)	(0.00760)
[11,12) min	0.00746***	0.0173***	0.00177	0.0171*
	(0.00155)	(0.00479)	(0.00967)	(0.00780)
[12,13) min	0.0107***	0.0193***	-0.00231	-0.00902
	(0.00163)	(0.00521)	(0.00958)	(0.00772)
[13,14) min	0.00969***	0.0267***	-0.00571	0.0289**
	(0.00167)	(0.00544)	(0.0110)	(0.00914)
[14,15) min	0.0147***	0.0291***	0.0140	0.0252**
	(0.00175)	(0.00577)	(0.0114)	(0.00911)
[15,16) min	0.0165***	0.0638***	0.0164	0.0439***
	(0.00186)	(0.00679)	(0.0119)	(0.00962)
[16,17) min	0.0208***	0.139***	0.0110	0.132***
	(0.00201)	(0.00807)	(0.0114)	(0.0131)
[17,18) min	0.0140***	0.0287***	0.0149	-0.0171
	(0.00198)	(0.00659)	(0.0141)	(0.00900)
[18,19) min	0.0118***	0.0212**	-0.0108	0.00496
	(0.00203)	(0.00667)	(0.0123)	(0.0103)
[19,20) min	0.0137***	0.0305***	0.0223	0.0195
	(0.00214)	(0.00748)	(0.0135)	(0.0106)
[20,21) min	0.0147***	0.0287***	0.000792	0.0113
	(0.00227)	(0.00784)	(0.0130)	(0.0110)
[21,22) min	0.0182***	0.0315***	0.0240	0.0389**
	(0.00239)	(0.00738)	(0.0143)	(0.0124)
[22,23) min	0.0155***	0.0120	0.0100	0.0245
	(0.00238)	(0.00743)	(0.0151)	(0.0127)
[23,24) min	0.0170***	0.0187*	0.0276	0.00868
	(0.00258)	(0.00779)	(0.0152)	(0.0122)
[24,25) min	0.0199***	0.0249**	-0.0178	0.0412**
	(0.00265)	(0.00835)	(0.0145)	(0.0142)
>25 min	0.0188***	0.0209***	0.0209***	0.0234***
	(0.000689)	(0.00199)	(0.00427)	(0.00352)

Notes: Standard errors are in parentheses and are clustered at the level of the arrival airport-day. Columns display coefficients from a single regression on four sets of predicted delay "bins" that are defined to be mutually exclusive. Specification includes carrier-arrival airport-day fixed effects and arrival hour and hub controls. Coefficients represent the change in the probability of a flight arriving exactly two minutes earlier than predicted relative to the probability of arriving exactly two minutes earlier than predicted for flights with predicted delay of less than 10 minutes. Calculation of predicted delay is described in the text on page 13. The regression contains 3,067,533 observations.

Table 5
Identification of "Manual" Planes, 1995-1998
Likelihood of a Plane Landing with Exactly Zero Delay, by Reporting Status

	50^{th} percentile	**75^{th} Percentile**	**90^{th} Percentile**	**95^{th} Percentile**	**99^{th} Percentile**	**Reporting Status in 1998**
Alaska	0.0577	0.0621	0.0652	0.0671	0.0709	Manual
America West	0.05	0.0552	0.0591	0.0604	0.0653	Manual
American	0.0333	0.0384	0.0429	0.0455	0.0509	Auto
Continental	0.0418	0.0459	0.0521	0.0577	0.0689	Combo
Delta	0.0393	0.0464	0.0537	0.0569	0.0620	Combo
Northwest	0.0356	0.0400	0.0433	0.0455	0.0502	Auto
Southwest	0.1172	0.1230	0.1277	0.1299	0.1335	Manual
TWA	0.0327	0.0360	0.0432	0.0559	0.0613	Combo
United	0.0380	0.0421	0.0466	0.0491	0.0553	Auto
US Airways	0.0385	0.0432	0.0464	0.0483	0.0546	Auto

*Notes***:** Table shows the distribution of a plane-year level variable that equals the probability that the plane is reported to have landed with zero minutes of delay. For example, the fourth entry in the row for American Airlines (third row of table) indicates that only 5% of American's planes in 1996 reportedly landed with zero delay more than 4.5% of the time. The entries in the row for Southwest Airlines (final row of table) indicate that 50% of Southwest's planes in 1996 reportedly landed with zero delay more than 11% of the time. The three shaded rows represent the three carriers that we think were combination reporters in 1996. Their entries show that their planes are slightly more likely than the automatic reporters to land with a reported delay of zero but not nearly as likely as the manual reporters.

Table 6
Probability that "Late" Flight Has Shorter Taxi Time as Function of "Early" Flight's *Predicted* Delay (Flights that Land at the Exact Same Time) , 1995-1998

Dependent Variable	***"Late" Member of Pair Has Shorter Taxi Time***			
	Coefficient estimates for:			
	All Other Carriers	**CO post-Bonus**	**TWA pre-Bonus**	**TWA post-Bonus**
Predicted Delay of "Early" Member of Pair				
[10,11) min	-0.0518***	-0.0702	-0.0159	-0.0274
	[0.0125]	[0.0487]	[0.0504]	[0.0621]
[11,12) min	-0.0523***	-0.0727	0.101***	-0.00537
	[0.0136]	[0.0434]	[0.0197]	[0.0512]
[12,13) min	-0.0297	-0.0679	-0.156***	-0.0907*
	[0.0165]	[0.0585]	[0.0264]	[0.0360]
[13,14) min	-0.0447***	-0.0887**	0.0105	-0.176***
	[0.0111]	[0.0335]	[0.0257]	[0.0396]
[14,15) min	-0.0612***	-0.199***	-0.217***	0.00157
	[0.0136]	[0.0308]	[0.0622]	[0.0343]
[15,16) min	-0.0584***	-0.117*	-0.203***	-0.0188
	[0.0105]	[0.0490]	[0.00323]	[0.104]
[16,17) min	-0.0651***	-0.193***	-0.0581*	0.00669
	[0.0144]	[0.0332]	[0.0231]	[0.0900]
[17,18) min	-0.0597***	-0.133*	-0.0732	-0.165
	[0.0112]	[0.0625]	[0.0647]	[0.112]
[18,19) min	-0.0325**	-0.159*	-0.0341	-0.201***
	[0.0119]	[0.0650]	[0.0360]	[0.0175]
[19,20) min	-0.0529***	-0.0621	0.0115	-0.173*
	[0.0104]	[0.0386]	[0.0292]	[0.0814]
[20,21) min	-0.00644	-0.0737	-0.0953***	-0.0673
	[0.0170]	[0.0500]	[0.0261]	[0.0668]
[21,22) min	-0.0235*	-0.0513	0.0899**	-0.183***
	[0.0110]	[0.0394]	[0.0286]	[0.0304]
[22,23) min	-0.0533***	-0.106	-0.00261	-0.0454
	[0.0118]	[0.0654]	[0.0344]	[0.0281]
[23,24) min	-0.0389*	-0.107*	0.0291	-0.141*
	[0.0151]	[0.0453]	[0.0675]	[0.0708]
[24,25) min	-0.0478***	-0.0772	-0.0243	-0.108**
	[0.0137]	[0.0506]	[0.0446]	[0.0353]
>25 min	-0.0432***	-0.0676***	-0.0533***	-0.0374**
	[0.00474]	[0.0169]	[0.00950]	[0.0142]

Notes: Sample includes carriers' flights that touch-down at the exact same minute. Restricted to two-member pairs. Standard errors are in parentheses. Columns display coefficients from a single regression on four sets of predicted delay "bins" that are defined to be mutually exclusive. Coefficients represent the change in the probability that the "late" member of pair has a shorter taxi time, relative to when "late" member is paired with flight with predicted delay of less than 10 minutes.

Table 7A
Simulated Changes in On-Time Performance and Rankings
Continental, 1995-1997

Year	Month	Actual % On-Time	Simulated % On-Time	Standard Error of Simulated % On-Time	Actual Rank	Simulated Rank
1995	2	0.1704	0.1762	0.0007	4	4
1995	3	0.1507	0.1570	0.0006	1	1
1995	4	0.1451	0.1498	0.0007	2	3
1995	5	0.1963	0.1997	0.0006	9	8
1995	6	0.3313	0.3274	0.0008	10	10
1995	7	0.1691	0.1772	0.0008	2	5
1995	8	0.1286	0.1353	0.0005	1	2
1995	9	0.1037	0.1094	0.0006	2	2
1995	10	0.1324	0.1403	0.0006	3	4
1995	11	0.1709	0.1778	0.0007	4	4
1995	12	0.2111	0.2195	0.0007	1	2
1996	1	0.2370	0.2469	0.0008	2	2
1996	2	0.1901	0.2015	0.0008	2	2
1996	3	0.2011	0.2138	0.0007	5	6
1996	4	0.1800	0.1908	0.0008	4	4
1996	5	0.1334	0.1453	0.0009	2	2
1996	6	0.2441	0.2611	0.0011	6	6
1996	7	0.2170	0.2323	0.0005	5	6
1996	8	0.2358	0.2515	0.0006	5	6
1996	9	0.1960	0.2090	0.0009	4	6
1996	10	0.1797	0.1933	0.0005	3	3
1996	11	0.1653	0.1774	0.0005	1	3
1996	12	0.2421	0.2570	0.0007	1	1
1997	1	0.2434	0.2584	0.0007	2	4
1997	2	0.1869	0.2018	0.0007	2	4
1997	3	0.1941	0.2107	0.0008	5	8
1997	4	0.1785	0.1919	0.0006	6	7
1997	5	0.1698	0.1827	0.0008	8	9
1997	6	0.2131	0.2267	0.0007	8	8
1997	7	0.1723	0.1871	0.0009	4	5
1997	8	0.1720	0.1856	0.0008	4	5
1997	9	0.1367	0.1488	0.0005	5	8
1997	10	0.1728	0.1867	0.0008	7	8
1997	11	0.2050	0.2182	0.0007	6	7
1997	12	0.2270	0.2397	0.0006	3	5

Number of months in which actual rank is **better** than simulated: **19**

Number of months in which actual rank is **same** as simulated: **13**

Number of months in which actual rank is **worse** than simulated (others simulated): **1**

Notes: Based on 20 iterations, standard errors average 300 times smaller than the reported on-time.

Table 7B
Simulated Changes in On-Time Performance and Rankings
TWA, 1996-1998

Year	Month	Actual % On-Time	Simulated % On-Time	Standard Error of Simulated % On-Time	Actual Rank	Simulated Rank
1996	6	0.2845	0.2927	0.0008	9	9
1996	7	0.2995	0.3046	0.0010	8	8
1996	8	0.2836	0.2931	0.0009	8	8
1996	9	0.2106	0.2135	0.0008	6	6
1996	10	0.2146	0.2221	0.0010	5	6
1996	11	0.1861	0.1929	0.0010	5	6
1996	12	0.3302	0.3377	0.0010	6	7
1997	1	0.2833	0.2923	0.0009	6	6
1997	2	0.2081	0.2154	0.0008	5	5
1997	3	0.2041	0.2128	0.0010	8	8
1997	4	0.1402	0.1456	0.0006	1	2
1997	5	0.1040	0.1121	0.0007	1	1
1997	6	0.1372	0.1489	0.0008	1	1
1997	7	0.1275	0.1445	0.0009	1	2
1997	8	0.1515	0.1696	0.0007	2	3
1997	9	0.0848	0.0977	0.0006	1	2
1997	10	0.1175	0.1317	0.0005	1	2
1997	11	0.1872	0.2032	0.0009	3	5
1997	12	0.2756	0.2977	0.0008	8	9
1998	1	0.2259	0.2421	0.0007	5	5
1998	2	0.1906	0.2107	0.0012	4	4
1998	3	0.2571	0.2781	0.0009	9	9
1998	4	0.1891	0.2092	0.0012	6	7
1998	5	0.2093	0.2302	0.0011	6	6
1998	6	0.2985	0.3179	0.0010	7	9
1998	7	0.1836	0.2001	0.0007	6	6
1998	8	0.1392	0.1522	0.0007	1	2
1998	9	0.1081	0.1186	0.0007	1	3
1998	10	0.1046	0.1172	0.0008	1	1
1998	11	0.1075	0.1217	0.0007	1	1
1998	12	0.2080	0.2275	0.0013	4	5

Number of months in which actual rank is **better** than simulated (others simulated): **15**

Number of months in which actual rank is **same** as simulated (others simulated): **16**

Number of months in which actual rank is **worse** than simulated (others simulated): **0**

Based on 20 iterations, standard errors average 300 times smaller than the reported on-time.

Appendix A
Changes in On-Time Performance after Introduction of Employee Bonus Programs, 1995-1998

Dependent Variable	*Arrival Delay*	*Arrival Delay≥15 min*	*Taxi In Time*	*Departure Delay*	*Taxi Out Time*
	(1)	(2)	(3)	(4)	(6)
CO*Bonus Period	-2.370***	-0.0476***	-0.585***	-1.797***	-0.227
	(0.177)	(0.00237)	(0.0836)	(0.150)	(0.127)
TW*Bonus Period	-2.609***	-0.0484***	-0.0807**	-0.947***	-0.209***
	(0.207)	(0.00269)	(0.0260)	(0.214)	(0.0482)
Airline Dummies					
CO	2.034***	0.0328***	-0.114	2.482***	0.536***
	(0.171)	(0.00229)	(0.0878)	(0.153)	(0.129)
DL	1.964***	0.0273***	-0.498***	1.170***	0.0290
	(0.0974)	(0.00144)	(0.0352)	(0.112)	(0.0279)
NW	0.289*	0.00992***	-0.0867**	0.372**	0.00792
	(0.113)	(0.00154)	(0.0326)	(0.113)	(0.0311)
TW	1.889***	0.0309***	-0.757***	1.806***	0.276***
	(0.170)	(0.00223)	(0.0352)	(0.179)	(0.0459)
UA	1.742***	0.00991***	-1.266***	3.176***	-1.081***
	(0.107)	(0.00143)	(0.0349)	(0.105)	(0.0291)
US	0.876***	0.0194***	-0.736***	2.193***	-1.873***
	(0.107)	(0.00151)	(0.0314)	(0.106)	(0.0293)
WN	1.040***	0.000669	-2.157***	2.597***	-3.988***
	(0.108)	(0.00159)	(0.0313)	(0.111)	(0.0338)
HP	4.953***	0.0527***	-0.696***	2.940***	-1.150***
	(0.139)	(0.00202)	(0.0325)	(0.144)	(0.0365)
AS	2.818***	0.0284***	-1.535***	0.427*	-1.000***
	(0.209)	(0.00349)	(0.0345)	(0.171)	(0.0426)
N	4,966,448	4,966,448	3,983,280	4,966,448	3,983,280
R-squared					

Notes: Standard errors are in parentheses and are clustered at the arrival airport-day level. All specifications include arrival airport-day fixed effects. All specifications also include departure and arrival hour controls as well as controls airline and airport level controls. Appendix B presents the coefficient estimates on the control variables. Data on taxi time is not available prior to 1995. As a result, columns (3) and (6) have fewer observations.

When Should Sellers Use Auctions?

James W. Roberts* Andrew Sweeting†

June 29, 2011

Abstract

A bidding process can be organized so that offers are submitted simultaneously or sequentially. In the latter case, potential buyers can condition their behavior on previous entrants' decisions. The relative performance of these mechanisms is investigated when entry is costly and selective, meaning that potential buyers with higher values are more likely to participate. A simple sequential mechanism can give both buyers and sellers significantly higher payoffs than the commonly used simultaneous bid auction. The findings are illustrated with parameters estimated from simultaneous entry USFS timber auctions where our estimates predict that the sequential mechanism would increase revenue and efficiency.

JEL CODES: D44, L20

Keywords: Auctions, Entry, Selection, Mechanism Design, Mergers and Acquisitions

*Department of Economics, Duke University. Contact: j.roberts@duke.edu.

†Department of Economics, Duke University and NBER. Contact: atsweet@duke.edu.

We are extremely grateful to Susan Athey, Jonathan Levin and Enrique Seira for sharing the data with us and to Coleman Doggett, Kent Lou and Doug MacDonald for their insight on the timber industry. We would also like to thank John Asker, Brendan Daley, Ken Hendricks, Jakub Kastl, Greg Lewis, David McAdams, Brian McManus, Rob Porter, Steve Tadelis and numerous seminar participants for useful feedback. Excellent research assistance was provided by Vivek Bhattacharya. Both authors gratefully acknowledge financial support from Duke University. Any errors are our own.

1 Introduction

The simultaneous bid auction is a standard method for sellers to solicit offers from buyers. A simple alternative is for a seller to ask buyers to make offers sequentially. If it is costly for buyers to participate, the sequential mechanism will tend to be more efficient than the simultaneous auction because later potential buyers can condition their participation decisions on earlier bids. However, the sequential mechanism's greater efficiency may not produce higher revenues because while the possibility of deterring later potential entrants can lead early bidders to bid aggressively, the fact that later firms might be deterred will tend to reduce revenues. The relative revenue performance of the mechanisms will therefore depend on whether the threat of potential future competition, which can raise bids in the sequential mechanism, is more valuable to the seller than actual competition, which will tend to be greater in the simultaneous auction.

The relative revenue performance of these alternative mechanisms has direct implications for how assets should be sold. In the case of how to structure the sale of corporations, this question has attracted attention from practitioners and other commentators since the Delaware Supreme Court's 1986 *Revlon* decision charged a board overseeing the sale of a company with the duty of "getting the best price for the stockholders" (Revlon v McAndrews & Forbes Holdings (1986)). In practice, corporate sales occur through a mixture of simultaneous and sequential mechanisms, with sequential mechanisms sometimes taking the form of "go-shop" arrangements where a seller may reach an agreement with one firm while retaining the right to solicit other offers, to which the first firm may be able to respond.[1,2]

Surprisingly, the only attempt to date to directly address this relative performance question is Bulow and Klemperer (2009) (BK hereafter). They compare the revenue and efficiency performances of the commonly-used simultaneous bid second-price auction with a similarly simple, sequential mechanism. In this second mechanism, buyers are approached in turn, and upon observing the history of offers, each chooses whether to enter and attempt to outbid the current high bidder. If the incumbent is outbid, the new entrant can make a jump bid that may potentially deter later firms from participating. The incumbent at the end of the game pays the standing price. As BK note (see also Subramanian (2008), Wasserstein (2000))[3],

[1] A "go-shop" clause allows a seller to come to an agreement on an initial price with a buyer and retain the right to solicit bids from other buyers for the next 30-60 days. If a new, higher offer is received, then according to the "match right", which is often included in the agreement with the initial buyer, the seller must negotiate with the first buyer (for 3-5 days, for example) to see if it can match the terms of the new, higher offer.

[2] There are numerous theory papers, some related directly to the field of corporate finance, that consider sequential mechanisms similar to the one considered here. Examples include Fishman (1988), Daniel and Hirshleifer (1998) and Horner and Sahuguet (2007).

[3] Boone and Mulherin (2007) describe various sale methods for corporate takeovers and establish that the

these simple mechanisms can be thought of as spanning the range of sale processes that are actually used. In the comparison, BK assume that potential bidders only know the distribution from which values are drawn prior to entering, and have no additional information about their own value. After entry they find out their values for sure. These assumptions are common in the auction literature as they provide greater analytic tractability. Under this informational assumption, together with the assumption that bidders are symmetric and the seller cannot set a reserve price, BK show that "sellers will generally prefer auctions and buyers will generally prefer sequential mechanisms" (p. 1547).

This result holds in BK's model because, in the equilibrium they consider, early bidders with high enough values submit bids that deter all future potential entry (all future potential entrants have the same beliefs about their values prior to entry), and there is too much deterrence from the seller's perspective. Thus, he would prefer the greater actual competition provided by the auction. In particular, deterrence means that later potential entrants with high values will not enter, which decreases both the expected value of the winning bidder and the value that an incumbent has to pay. In contrast, buyers prefer the sequential mechanism as expenditures on entry costs are lower. This effect also tends to increase social efficiency.

In light of their result, BK interpret the use of sequential mechanisms as evidence that buyer's preferences can determine the choice of mechanism, consistent with the fact that some influential buyers, such as Warren Buffett, have explicit policies that they will not "waste time" by participating in auctions.

In this paper, we consider a similar comparison, except that we extend BK's model to allow potential buyers to receive a noisy signal about their valuation prior to deciding whether to enter either mechanism. After entry, they find out their values for sure, as in BK's model. This structure results in a "selective entry" model, where firms enter if they receive high enough signals, and firms with higher values are more likely to enter.[4,5] We believe that this is a natural model to describe settings where firms are likely to have some imperfect information about their value for an asset based on publicly available information, but must conduct costly research to discover additional information that will affect their value.[6] We

set of potential bidders in these contests is larger than the set submitting public bids.

[4]The precision of the signal determines how selective the entry process is. In its limits, the model can approach the polar cases of (a) perfect selection, which we term the S model after Samuelson (1985), whereby a firm knows its value exactly when taking its entry decision, and (b) no selection, which we term the LS model after Levin and Smith (1994), whereby a firm knows nothing of its value when taking its entry decision.

[5]Selective entry contrasts with standard assumptions in the empirical entry literature (e.g., Berry (1992)) where entrants may differ from non-entering potential entrants in their fixed costs or entry costs, but not along dimensions such as marginal costs or product quality that affect competitiveness or the profits of other firms once they enter.

[6]Examples include oil and gas leases, timber sales, government procurement contracts and firm takeovers. Recently there has been some work allowing for selection in empirical auction research. The dominant way this is done is by assuming that bidders know their value precisely prior to entry, i.e. by assuming perfect

also allow for potential buyers to be asymmetric, which is another important feature of many real-world settings.

Using numerical analysis, which becomes necessary once either asymmetries or selective entry are added to the model, we show that the sequential mechanism can give the seller *higher* expected revenues than the simultaneous auction even when buyers' signals about their values are quite noisy. When the entry process is quite selective and/or entry costs are large, the difference in revenues can be substantial, and, as a comparison, the increase in revenues from using the sequential mechanism is much larger than the returns to using an optimal reserve price in the simultaneous auction. As in BK's analysis, the sequential mechanism is more efficient, and the sequential mechanism generally gives higher expected payoffs to both buyers and sellers. This result would obviously lead to a different interpretation of why sequential mechanisms are sometimes used, and because the sequential mechanism increases the payoffs of buyers, it is still consistent with comments like those of Warren Buffett. Our findings are also consistent with observed differences in target shareholder returns in private equity transactions documented by Subramanian (2008). He compares returns when companies are sold using a go-shop process and a process where many firms are simultaneously asked to submit bids before a winner is selected. He finds that target shareholder returns are 5% higher for go-shops and he argues that, even though go-shop agreements introduce asymmetries between bidders into the sale process, they are preferable for both buyers and sellers.

We illustrate our findings using parameters estimated from a sample of (simultaneous) open outcry US Forest Service (USFS) timber auctions. This setting provides a close match to the information structure assumed in our model as a potential bidder can form a rough-estimate of its value based on tract information published by the USFS and knowledge of its own sales contracts and capabilities, and it is also standard for interested bidders to conduct their own tract surveys ("cruises") prior to bidding. It is also a setting where various auction design tools, such as reserve price policies, have been studied by both academics and practitioners in order to try to raise revenues which have often been regarded as too low.[7]

selection. For example, Li and Zheng (2009) compare estimates from both the LS and S models using data on highway lawn mowing contracts from Texas to understand how potential competition may affect procurement costs, and Li and Zheng (2010) test the LS and S models using timber auctions in Michigan. Marmer, Shneyerov, and Xu (2010) extend this literature by testing whether the Li and Zheng (2009) data is best explained by the LS, S or a more general affiliated signal model. They find support for the S and signal models, and they also estimate a very simple version of their signal model. Finally, Gentry and Li (2010) show how partial identification techniques can be used to construct bounds on the primitives of a signal model.

[7]Some examples of studies of timber auction reserve prices include Mead, Schniepp, and Watson (1981), Paarsch (1997), Haile and Tamer (2003), Li and Perrigne (2003) and Aradillas-Lopez, Gandhi, and Quint (2010). All of these papers assume that entry is not endogenous. Academics have also provided expert advice to government agencies about how to set reserve prices (stumpage rates) for timber (e.g. Athey, Cramton, and Ingraham (2003)). In 2006, Governor Tim Pawlenty of Minnesota commissioned a task force

Timber auctions are also characterized by important asymmetries between potential buyers, with sawmills tending to have systematically higher values than loggers.

Our estimates imply that the entry process into timber auctions is moderately selective, while average entry costs are 2.3% of the average winning bid, which is large enough to prevent many loggers from entering auctions. For the (mean) representative auction in our data, our results imply that using a sequential mechanism (with no reserve) would generate a nine times larger increase in revenues than setting the optimal reserve price (the focus of the existing literature) in the simultaneous auction. We also find that the efficiency gains from using the sequential mechanism are large enough that both the USFS revenues and firm profits can increase. Additionally, loggers (the weaker type) win more often. These results suggest that the sequential mechanism may present the USFS and other procurement agencies with an effective, new alternative to commonly used set-aside programs and bid subsidies for ensuring that a certain fraction of projects are won by a targeted set of bidders.[8]

Why does the sequential mechanism tend to produce higher revenues when entry is selective? The key reason is that selective entry changes the nature of the equilibrium in the sequential mechanism in a way that tends to increase both its relative efficiency and the revenues that the seller can extract. With no selection, BK show that the "pre-emptive bidding [which occurs in equilibrium] is crucial: jump-bidding allows buyers to choose partial-pooling deterrence equilibria which over-deter entry relative to the social optimum" (p. 1546). Introducing any degree of selection into the entry process causes the bidding equilibrium (in the unique equilibrium under the D1 refinement which we focus on) to change so that there is full separation, with bids perfectly revealing the value of the incumbent.[9] At the same time, a potential entrant will enter if it receives a high enough signal about its value. These changes increase the efficiency of the outcome in the sequential mechanism as higher value incumbents deter more entry with higher value potential entrants being more likely to enter. Unlike in BK's model, the expected value of the winner can be higher in the sequential mechanism, which increases the rents available to all parties. In addition, the change to a separating equilibrium affects the equilibrium level of jump bids. For some values, this will increase the expected amount that bidders pay, benefiting the seller.

Some comments about the nature of our results are appropriate. First, we do not seek

to investigate the performance of the state's timber sale policies, and its report indicates an openness to considering alternative sales mechanisms as well as different reserve prices (Kilgore, Brown, Coggins, and Pfender (2010)).

[8]The USFS has historically used set-asides and recent work (Athey, Coey, and Levin (2011)) suggests this may come at a substantial revenue and efficiency loss relative to using bid subsidies.

[9]This is correct for values less than the upper limit of the value distribution minus the cost of entry. An upper limit on the value distribution is required for technical reasons but we assume that it is sufficiently high that, for practical purposes, all incumbent values are revealed.

to compare revenues with those from the optimal mechanism. Instead, in the same spirit as BK, we are interested in the relative performance of stylized versions of commonly used sales mechanisms, whereas the seller optimal mechanism, which is not known for a model with imperfectly selective entry (Milgrom (2004)), is likely to involve features, such as side payments or entry fees that are rarely observed in practice, and which might require the seller to have implausibly detailed information.[10] The seller would also need to know this information if he wants to set the optimal reserve price in an auction, and, indeed, an attraction of the simple sequential mechanism that we consider is that the only required information concerns the set of potential entrants who should be approached.[11] We show below that if the seller has enough information to set an optimal reserve in the sequential mechanism, he can do even better.

On the other hand, what is known about the optimal mechanism in models with costly entry and either no selection or perfect selection suggests that the optimal mechanism should be sequential, which helps to rationalize our results. For example, Cremer, Spiegel, and Zheng (2009) consider the case with *no* selection and McAfee and McMillan (1988) consider a model where buyers know their values but it is costly for the seller to engage additional buyers. In both cases, the optimal mechanism involves some type of sequential search procedure, which stops when a buyer with a high enough value is identified.[12]

Second, while we characterize the unique equilibrium of each mechanism under standard refinements, our revenue comparisons are numerical in nature. This is a necessary cost of allowing for either a more general model of entry, or bidder asymmetries. Our results show that these features matter because the relative performance of the mechanisms can change even when selection is quite imperfect. The computational approach also allows us to provide a substantive empirical application of our model as selective entry and bidder asymmetries are clear features of our data.

Third, the sequential mechanism can be characterized as a multi-round extension of a standard two-player signaling game where an incumbent bidder can use a jump bid to signal its value to later potential entrants. We contribute to the literature on extensions of two-

[10]Our approach is therefore similar to analyses of practical mechanisms in other settings, such as Chu, Leslie, and Sorensen (2011) (bundling), Rogerson (2003) (contracts), McAfee (2002) (nonlinear pricing) and Neeman (2003) (auctions).

[11]In this sense the sequential mechanism satisfies what has come to be known as the "Wilson doctrine" (Wilson (1987)), which suggests that, from a practical standpoint, we ought to be concerned with mechanisms that do not rely on the seller possessing unrealistically detailed information about buyers.

[12]In this environment, Ye (2007) considers two-stage bidding structures where the seller must choose how many firms pay the entry cost ahead of the first stage. His paper clearly shows how to determine the optimal number of entrants. However, he does not consider a wider range of mechanisms that might allow, for example, the seller to set a reserve price or to decide how many firms should enter only after the first stage bids are submitted.

player signaling games by characterizing the unique sequential equilibrium under standard refinements and providing a straightforward recursive algorithm for calculating equilibrium strategies. We also note that our findings relate to the classic limit pricing result of Milgrom and Roberts (1982). As they show in a two-period, two-firm setting, an incumbent's incentive to deter a competitor's entry can benefit consumers through lower prices. We find a similar result that the incentive to deter later potential competitors can benefit a seller through higher prices.

Fourth, we note two differences, beside the introduction of selective entry and bidder asymmetries, between our model and the model considered by BK. First, we assume that the number N of potential entrants is fixed and common knowledge to all players, whereas BK's model allows for some probability ($0 \leq \rho_j \leq 1$) of a j^{th} potential entrant if there are $j-1$ potential entrants. As these probabilities may equal 1 for $j < N$, and 0 for $j \geq N$ for any N, our model is a special case of theirs. Our choice reflects the standard practice in the empirical literature, which we want to follow when estimating our model.[13] Second, when modeling the auction mechanism, we focus on the model where potential buyers make simultaneous entry decisions as well as simultaneous bid choices, whereas BK's primary focus is on a model where firms make sequential entry decisions before bidding simultaneously. However, in BK's model "no important result is affected if potential bidders make simultaneous, instead of sequential, entry decisions into the auction" (p.1560). We also give some consideration to a sequential entry, simultaneous bid model, and show that our qualitative results are unchanged. Our choice to focus on simultaneous entry into the auction reflects a desire to reduce the computational burden and, more importantly, the fact that simultaneous entry is the appropriate way to model entry into the auctions in our empirical sample (Athey, Levin, and Seira (forthcoming) also apply a simultaneous entry model (with no selection) to USFS timber auctions).[14]

The paper proceeds as follows. Section 2 introduces the models of each mechanism and characterizes the equilibria that we examine. Section 3 compares expected revenue and efficiency from the two mechanisms for wide ranges of parameters, and provides intuition for when the sequential mechanism outperforms the auction. Section 4 describes the empirical setting of USFS timber auctions and explains how we estimate our model. Section 5 presents

[13]Examples of this assumption in the auction literature include Athey, Levin, and Seira (forthcoming) and Li and Zheng (2009). Examples elsewhere in empirical work on entry games include Berry (1992), Seim (2006) and Ciliberto and Tamer (2009).

[14]The computational burden in the sequential entry, simultaneous bid auction model arises from the fact that later potential entrants' equilibrium entry thresholds are a function of the complete history of the game and thresholds in earlier rounds, so that it is necessary to solve for all of the thresholds simultaneously. In contrast, in the sequential mechanism, a potential entrant's equilibrium threshold only depends on the value of the incumbent which (with any degree of selection) is completely revealed by its jump bid.

the parameter estimates and counterfactual results showing that the USFS could improve its revenues by implementing a sequential mechanism. Section 6 concludes.

2 Model

We now describe the model of firms' values and signals, before describing the mechanisms that we are going to compare.

2.1 A General Entry Model with Selection

Suppose that a seller has one unit of a good to sell and gets a payoff of zero if the good is unsold. There is a set of potential buyers who may be one of $\tau = 1, ..., \overline{\tau}$ types, with N_τ of type τ. In practice $\overline{\tau} = 2$. Buyers have independent private values (IPV), which can lie on $[0, \overline{V}]$, distributed according to $F_\tau^V(V)$. F_τ^V is continuous and differentiable for all types. In this paper we will assume that the density of V is proportional to the log-normal distribution on $[0, \overline{V}]$, and that $\overline{V}$ is high, so that the density of values at $\overline{V}$ is very small.[15] Before participating in any mechanism, a potential buyer must pay an entry cost K_τ. This entry cost can be interpreted as a combination of research costs necessary to learn one's value and participation/bidding costs. Once it pays K_τ, a potential buyer learns its value. We assume that a firm cannot participate without paying K_τ. However, prior to deciding whether to enter, a bidder receives a private information signal about its value. We focus on the case where the signal of potential buyer i of type τ is given by $s_{i\tau} = v_{i\tau} a_{i\tau}$, where $A_\tau = e^{\varepsilon_\tau}, \varepsilon_\tau \sim N(0, \sigma_{\varepsilon\tau}^2)$ and draws of ε are assumed to be i.i.d. across bidders.

Let $F_\tau^S(s)$ be the unconditional distribution of a bidder's signal and $F_\tau^S(s|v)$ be the distribution conditional on a particular value v. In this model, $\sigma_{\varepsilon\tau}^2$ controls how much potential buyers know about their values before deciding whether to enter. As $\sigma_{\varepsilon\tau}^2 \to \infty$, the model will tend towards the informational assumptions of the Levin and Smith (1994) (LS) model in which pre-entry signals contain no information about values. As $\sigma_{\varepsilon\tau}^2 \to 0$, it tends towards the informational assumptions of the Samuelson (1985) (S) model where firms know their values prior to paying an entry cost (which is therefore interpreted as a bid preparation or attendance cost). Intermediate values of $\sigma_{\varepsilon\tau}^2$, implying that buyers have some idea of their values but have to conduct costly research to learn them for sure, seem plausible for most empirical settings. Having received his signal, a potential buyer forms posterior beliefs about his valuation using Bayes Rule.

[15] To be precise, $f^V(v|\theta) = \frac{h(v|\theta)}{\int_0^{\overline{V}} h(x|\theta)dx}$, where $h(v|\theta)$ is the pdf of the log-normal distribution.

2.2 Mechanism 1: Simultaneous Entry Second Price Auction

The first mechanism we consider is a simultaneous entry second price or open outcry auction that we model as a two-stage game. In the first stage all potential buyers simultaneously decide whether to enter the auction (pay K_τ) based on their signal, the number of potential entrants of each type and the auction reserve price. In the second stage, entrants then learn their values and submit bids. We assume that an open outcry auction would give the same outcome as an English button auction, so that the good would be awarded to the firm with the highest value at a price equal to the value of the second highest valued entrant or the reserve price if one is used.[16]

Following the literature (e.g. Athey, Levin, and Seira (forthcoming)), we assume that players use strategies that form type-symmetric Bayesian Nash equilibria, where "type-symmetric" means that every player of the same type will use the same strategy. In the auction's second stage, entrants know their values so it is a dominant strategy for each entrant to bid its value. In the first stage, players take entry decisions based on what they believe about their value given their signal. The (posterior) conditional density $g_\tau(v|s_i)$ that a player of type τ's value is v when its signal is s_i is defined via Bayes Rule.

The weights that a player places on its prior and its signal when updating its beliefs about its true value depend on the relative variances of the distribution of values and ε (signal noise), and this will also control the degree of selection. A natural measure of the relative variances is $\frac{\sigma_\varepsilon^2}{\sigma_V^2+\sigma_\varepsilon^2}$, which we will denote α. If the value distribution were not truncated above, player i's (posterior) conditional value distribution would be lognormal with location parameter $\alpha\mu_\tau + (1-\alpha)ln(s_i)$ and squared scale parameter $\alpha\sigma_{V\tau}^2$.

The optimal entry strategy in a type-symmetric equilibrium is a pure-strategy threshold rule where the firm enters if and only if its signal is above a cutoff, $S_\tau'^*$. $S_\tau'^*$ is implicitly defined by the zero-profit condition that the expected profit from entering the auction of a firm with the threshold signal will be equal to the entry cost:

$$\int_R^{\overline{V}} \left[\int_R^v (v-x) h_\tau(x|S_\tau'^*, S_{-\tau}'^*) dx \right] g_\tau(v|S_\tau'^*) dv - K_\tau = 0 \tag{1}$$

where $g_\tau(v|s)$ is defined above, and $h_\tau(x|S_\tau'^*, S_{-\tau}'^*)$ is the pdf of the highest value of other entering firms (or the reserve price R if no value is higher than the reserve) in the auction, given equilibrium strategies. A pure strategy type-symmetric Bayesian Nash equilibrium exists because optimal entry thresholds for each type are continuous and decreasing in the threshold of the other type.

[16] Our estimation procedure does not require that other losing bidders bid up to their values.

With multiple types, there can be multiple equilibria in the entry game when types are similar (for example, in the means of their values) even when we assume that only type-symmetric equilibria are played. As explained in Roberts and Sweeting (2011), we choose to focus on an equilibrium where the type with higher mean values has a lower entry threshold (lower thresholds make entry more likely). This type of equilibrium is intuitively appealing and when firms' reaction functions are S-shaped (reflecting, for example, normal or log-normal value and signal noise distributions) and types only differ in the location parameters of their value distributions (i.e., the scale parameter, signal noise variance and entry costs are the same) then there is exactly one equilibrium of this form.[17] Therefore, we assume that types only differ in the location parameters of their value distributions from now on. Given our focus on this type of equilibrium, solving the model is straightforward: we find the S'^{*} values that satisfy the zero profit conditions for each type and which satisfy the constraint that $S_1'^{*} < S_2'^{*}$, where a type 1 firm is the high type (larger location parameter). It is important to note that the issue of type-symmetric multiple equilibria affects only the auction, not the sequential mechanism.

2.3 Mechanism 2: Sequential Mechanism

As BK and others note, the standard alternative to buyers submitting bids simultaneously is a sequential bid process. Here we describe a very simple sequential bid process like that in BK. Potential buyers are placed in some order (which does not depend on their signals, but may depend on types), and the seller approaches each potential buyer in turn. We will call what happens between the seller's approach to one potential buyer and its approach to the next potential buyer a "round". In the first round, the first potential buyer observes his signal and then decides whether to enter the mechanism and learn his value by paying K. If he enters he can choose to place a 'jump bid' above the reserve price, which we assume to be zero. Given entry, submitting a bid is costless.

In the second round the potential buyer observes his signal, the entry decision of the first buyer and his jump bid, and then decides whether to enter himself. If the first firm did not enter and the second firm does, then the second firm can place a bid in exactly the same way as the first firm would have been able to do had he entered. If both enter, the firms bid against each other in a knockout button auction until one firm drops out, in which case it can never return to the mechanism. The remaining firm then has an opportunity to submit

[17] We have also estimated the model using a nested pseudo-likelihood procedure which does not require us to use an equilibrium selection rule. The parameter estimates in this case indicate that the difference in mean values between our two types (sawmills and logging companies) are so large that multiple equilibria cannot be supported.

an additional higher jump bid above the bid at which the other firm dropped out. If the second firm does not enter, but the first firm did, then the first firm can either keep its initial bid or submit a higher jump bid.

This procedure is then repeated for each remaining potential buyer, so that in each round there is at most one incumbent bidder and one potential entrant. The complete history of the game (entry decisions and bids, but not signals) is observed by all players. If a firm drops out, or chooses not enter, it is assumed to be unable to re-enter at a later date. The good is allocated to the last remaining bidder at a price equal to the current bid.

A strategy in the sequential model consists of an entry rule and a bidding rule as a function of the round, the potential buyer's signal and value (for bidding) and the observed history. When a potential buyer is bidding against an active opponent in the knockout auction, the dominant strategy is to bid up to its value, so that the firm with the lower value will drop out at a price equal to its value. This does not depend on the selective entry model because values are known at this stage. However, the strategies that firms use to determine their jump bids and entry decisions do depend on selective entry. To place our equilibrium in context, we begin describing what happens when there are no signals and symmetric firms, which are the assumptions made by BK.

Before describing this mechanism's equilibrium, we note that it is straightforward for a seller to implement this mechanism. In particular, the seller needs only to identify potential entrants, specify and commit to a buyer order and establish a program for collecting and distributing information on the entry and bidding behavior of all firms. For sellers that will implement the mechanism many times, such as the USFS, any fixed costs involved in setting up the mechanism should be relatively small.

2.3.1 Equilibrium with No Pre-Entry Signals

Assuming symmetric firms and no pre-entry signals, BK show that any entering firm that learns its value is below some endogenously determined V^S will keep the existing standing bid, while firms with values above V^S will submit a jump bid that deters all future entry, no matter how many rounds are left. This is because all future potential entrants have identical information about their values prior to taking entry decisions. V^S is independent of the round of the game and history to that point and it is determined by the condition that future potential entrants should be indifferent to entering when the incumbent firm's value is above V^S. The deterring bid is determined by the condition that the bidder with a value V^S is indifferent between deterring future entry with this deterring bid and accommodating entry by keeping the standing bid. Thus while in any round all firms with values above V^S submit the same deterring bid, this bid may depend on the round and history of the game.

Equilibrium with no signals is therefore characterized by entry in every round until a firm with a value greater than V^S participates, in which case entry ceases forever. BK show that while this leads to higher expected efficiency than the auction, from the seller's standpoint, too many bidders are deterred from participating and in equilibrium revenues tend to be lower than in the auction.

2.3.2 Equilibrium with Pre-Entry Signals

There are important changes to the nature of the equilibrium when potential buyers receive pre-entry signals. We begin by describing the equilibrium we consider, before explaining the refinements that lead us to focus on it.

A potential entrant in any round n participates if and only if his signal exceeds some threshold $S_n'(v)$, at which the expected profits from entering are zero and is a function of the round, his beliefs about the current incumbent's (if there is one) value (v) and the expected behavior of future potential entrants. Upon entry, an incumbent and a new entrant bid up to their values in the knockout auction. The winner of the knockout auction may then submit a jump bid that may deter future entry. For a bidder with values on $[0, \overline{V} - K]$, its jump bid will perfectly reveal its value and so we assume that a new incumbent jump bids the first time he is able and after placing one jump bid he will not do so again. Therefore, given a bidding function in round n, $\beta(v, \widehat{b}_n, n)$, which depends on the bidder's value (v), the standing bid prior to the jump bid being placed ($\widehat{b}_n$ - this will be zero when the bidder is the first entrant and otherwise it will be the previous incumbent's value since they will have just lost in a knockout auction prior to a new jump bid being placed) and the round, an incumbent with value v must decide which v''s bid he should submit to maximize $\pi(v'|v, \widehat{b}_n, n)$, given:

$$\left[v - \beta(v', \widehat{b}_n, n)\right] \left[\Pi_{k=n+1}^{N} F^S(S_k'(v')) + \overline{F}_{n,v'}(\beta(v', \widehat{b}_n, n))\right] + \int_{\beta(v', \widehat{b}_n, n)}^{v} (v - x)\, \overline{f}_{n,v'}(x) dx \quad (2)$$

where $\overline{F}_{n,v'}(t) = \Pi_{k=n+1}^{N} \int_0^t f^V(x) \left(1 - F^S(S_k'(v')|x)\right) dx$ is the probability that entry occurs and that the maximum value of all future entrants, when the incumbent's value at the end of round n is believed to be v', is less than t, and $\overline{f}_{n,v'}(t) = \frac{\partial \overline{F}_{n,v'}(t)}{\partial v}$.[18] The first part of equation 2 reflects the incumbent's profits when either there is no more entry or, all future entrants have values less than the jump bid, so the incumbent can win at his jump bid. The second part reflects the incumbent's profit if an entrant has a value above the jump bid. Differentiating equation 2 with respect to v', and requiring that the first order condition equals zero when $v = v'$ (so that local incentive compatibility constraints are satisfied), gives

[18] Note that $\overline{F}_{n,v'}(\overline{V}) = 1 - \Pr[\textit{no entry in future}]$, so that we are not double counting in equation 2.

the differential equation that defines the bid function:

$$\frac{d\beta(\cdot)}{dv} = \frac{[v-\beta(\cdot)]\left[\frac{d\left[\Pi_{k=n+1}^{N}F^{S}(S_k'(v))\right]}{dv} + \frac{\partial \overline{F}_{n,v}(\beta(\cdot))}{\partial v}\right] + \int_{\beta(\cdot)}^{v}(v-\widehat{v})\frac{\partial \overline{f}_{n,v}(\widehat{v})}{\partial v}d\widehat{v}}{\Pi_{k=n+1}^{N}F^{S}(S_k'(v)) + \overline{F}_{n,v}(\beta(\cdot))} \tag{3}$$

The lower boundary condition is provided by the condition that incumbents with values less than or equal to the standing bid will submit the standing bid. For values on $[\overline{V}-K,\overline{V}]$ bidders pool and submit $\beta(\overline{V}-K,\widehat{b}_n,n)$.

When an incumbent bids $b \leq \beta(\overline{V}-K,\widehat{b}_n,n)$, the posterior belief of any potential entrant in round $m > n$ about the incumbent's value will place all of the weight on $\beta^{-1}(b,\widehat{b}_n,n)$. Thus, this potential entrant's entry threshold $S_m^{*\prime}(\beta^{-1}(b,\widehat{b},n))$ is implicitly defined by the following zero profit condition:

$$K = \int_{\beta^{-1}(b,\widehat{b}_n,n)}^{\overline{V}} \pi(x|x,\beta^{-1}(b,\widehat{b}_n,n),m)f^{V}(x|S_m^{*\prime}(\beta^{-1}(b,\widehat{b},n))dx \tag{4}$$

Upon observing a bid at $\beta(\overline{V}-K,\widehat{b}_n,n)$, beliefs will be consistent with Bayes Rule and a potential entrant will not participate.

Given the nature of this equilibrium we can solve the game recursively. For the final potential entrant, who believes that he will win if his value is greater than the incumbent's (in which case the final price will be the incumbent's value), we can solve for the equilibrium entry thresholds on a grid of possible values for an incumbent firm. Next we consider the previous potential entrant, and, given these final round thresholds, we can solve for both this entrant's entry thresholds and its equilibrium bid functions for a grid of possible values for an incumbent using equations 3 and 4. We then repeat the procedure for the third-from-last potential entrant, and so on until we reach the first round, where there will be no incumbent.

2.3.3 Equilibrium and Refinement

We now explain why the equilibrium just described exists and is the only equilibrium consistent with the D1 refinement (Banks and Sobel (1987), Cho and Kreps (1987)), which is a commonly used refinement for signaling models (Fudenberg and Tirole (1991)). For clarity we first consider the case with two potential entrants, which matches existing models in the signaling literature closely. We then consider the extension to the case with more firms. As the distribution of values has the same support for all types, adding more types has no effect on our arguments, so we assume there is only one type to reduce notation.

Our arguments will make use of three properties of the game. Let $\pi_v(b_1,S_2')$ be the expected profit of a first round incumbent with value v where b_1 is his bid and S_2' is the entry threshold chosen by the second round potential entrant. Given our assumptions and the dominant strategies in the knockout game, $\pi_v(b_1, S_2')$ will be continuous and differentiable in both arguments. The three properties are:

1. $\frac{\partial \pi_v(b_1,S_2')}{\partial S_2'} > 0$;

2. $\frac{\partial \pi_v(b_1,S_2')}{\partial b_1} / \frac{\partial \pi_v(b_1,S_2')}{\partial S_2'}$ is monotonic in v;

3. S_2' is uniquely defined for any belief about the first potential entrant's value, and the potential entrant's response is more favorable to the incumbent when the potential entrant thinks that the incumbent's value is higher.

The appendix shows that these properties hold in our model.

The results in Mailath (1987) imply that there is an unique separating equilibrium on the $[0, \overline{V} - K]$ interval which can be found using the differential equation given by equation 3 and the boundary condition in a continuous type signaling model when the single crossing condition (property 2 above) holds. His results do not rule out the possibility of pooling equilibria on this interval. However, Ramey (1996) (who extends the results in Cho and Sobel (1990) to the case of an unbounded action space and a continuum of types on an interval) shows that these three properties imply that only a separating equilibrium will satisfy the D1 refinement, so our equilibrium must be the only one satisfying D1. As noted by Mailath (1987), this equilibrium will also be the separating equilibrium which is least costly to the first round potential entrant.[19]

The conditions also imply that, if an incumbent with value $\overline{V} - K$ prefers $\beta(\overline{V} - K)$, which will stop all future entry, to a lower bid then all incumbents with values above $\overline{V} - K$ will prefer $\beta(\overline{V} - K)$ to lower bids. But, firms with values above $\overline{V} - K$ will also strictly prefer to bid $\beta(\overline{V} - K)$ than any higher bid (for any beliefs of the potential entrant following a higher bid), because by bidding $\beta(\overline{V} - K)$ the incumbent can get the asset for sure at a lower price. Therefore, in equilibrium all entrants with values greater than $\overline{V} - K$ pool.

Three (or More) Rounds We now consider a model with three potential entrants (arguments for more rounds would follow directly from this case). We make the natural simplification by restricting ourselves to equilibria where all potential entrants make the same inferences from a bid by an incumbent and incumbents only make jump bids in the first round

[19] Given that we show below that the sequential mechanism tends to generate higher revenues than the auction, the fact that we focus on the *least cost separating* equilibrium implies that our results may be conservative since other equilibria in the sequential mechanism would give even higher revenues to the seller.

that they enter.[20] The two period equilibrium discussed above would define strategies for the final two rounds if the second period entrant enters and defeats any incumbent entrant from the first round (with an adjusted boundary condition to reflect the new standing bid). It therefore only remains to be shown that there is a unique sequential equilibrium bid function, which is fully separating for values $[0, \overline{V} - K]$, for a first round entrant. A first round entrant's jump bid sends a signal to the second round potential entrant, and, if he is still an incumbent in the final round, which must be the case if he is to win, the final potential entrant. Conditional on the incumbent surviving the second round, the third round is just a repeat of another two round game. The first round entrant's expected profit function is now $\pi_v(b_1, S_2', S_3')$, and the following properties hold:

1. $\frac{\partial \pi_v(b_1,S_2',S_3')}{\partial S_2'} > 0$ and $\frac{\partial \pi_v(b_1,S_2',S_3')}{\partial S_3'} > 0$;
2. $\frac{\partial \pi_v(b_1,S_2',S_3')}{\partial b_1} / \frac{\partial \pi_v(b_1,S_2',S_3')}{\partial S_2'}$ and $\frac{\partial \pi_v(b_1,S_2',S_3')}{\partial b_1} / \frac{\partial \pi_v(b_1,S_2',S_3')}{\partial S_3'}$ are both monotonic in v;
3. both S_2' and S_3' are uniquely defined for any belief about the first entrant's value, and both potential entrants' responses are more favorable to the first entrant when they think that its value is higher.

These conditions allow us to apply the D1 refinement to the signaling game between the incumbent making the jump bid and every subsequent potential entrant to identify the unique separating equilibrium.

2.3.4 Illustrative Example of the Sequential Mechanism's Equilibrium

To provide additional clarity about how the mechanism works, given equilibrium strategies, Table 1 presents what happens in a game with four potential entrants and one type of firm with values distributed proportional to $LN(4.5, 0.2)$ on [0,200], $K = 1$ and $\sigma_\varepsilon = 0.2$ ($\alpha = 0.5$).

In the example, the first potential entrant enters if he receives a signal greater than 75.0, which is the case here. The signal thresholds in later rounds depend on the number of rounds remaining and the incumbent's value. So, when the incumbent is the same as in the previous round, the threshold S'^* falls since the expected profits of an entrant who beats the incumbent rise (because he will face less competition in the future). On the other hand, S'^* does not depend on the level of the standing bid given the incumbent's value, because it has no effect on the entrant's profits if he beats the incumbent in a knockout (since the standing bid must

[20] It is possible that future potential entrants could ignore the information that they have on games before the last round. In this case, incumbents would choose to submit jump bids every round to signal information to the next potential entrant. This simplification allows us to consider a model where a firm sends at most one signal to many possible receivers.

	Initial	Potential Entrant				Post-Knockout	Post-Jump Bid
Round	Standing Bid	Value	Signal	S'^*	Entry	Standing Bid	Standing Bid
1	-	80.0	90.1	75.0	Yes	-	69.3
2	69.3	75.4	50.5	69.4	No	69.3	69.3
3	69.3	116.0	114.9	61.7	Yes	80.0	87.1
4	87.1	100.0	114.0	107.0	Yes	100.0	100.0
Seller's Revenue = 100.0, social surplus (winner's value less total entry costs) =113.0							

Table 1: A simple example of how the sequential mechanism works in a game with four potential entrants and one type of firm with values distributed proportional to $LN(4.5, 0.2)$ on [0,200], $K = 1$ and $\sigma_\varepsilon = 0.2$.

be below the incumbent's value). In round 2, the incumbent does not face entry, so there is no change in the standing bid because incumbents do not place additional jump bids.[21] In round 3, the standing bid rises during the knockout, and the *new* incumbent places an additional jump bid. In round 4 the last potential entrant participates, but his value is less than the incumbent's and so revenue is the price at which this last entrant drops out.

We can also use this example to give intuition for how introducing selection affects bid functions and entry probabilities. With selection, the level of bids is determined by the fact that bids must be sufficiently high that firms with lower values will not want to copy them. In particular, if the entry decisions of later potential entrants are likely to be more sensitive to beliefs about the incumbent's value, then the equilibrium bid function must increase more quickly in v. A straightforward way to illustrate this is to focus on the last two rounds of the sequential mechanism when a new incumbent in the penultimate round only needs to worry about one more potential entrant, and the final round potential entrant would face no further entry if he enters and outbids the incumbent. This is illustrated in Figure 1, which compares the equilibrium bid functions in the penultimate round, and equilibrium probabilities of entry in the final round of the sequential mechanism for varying degrees of selection.

Specifically, the left panel displays bid functions for a new incumbent in the penultimate round, when the previous incumbent's value was 80. The right panel gives the probability that the final round potential entrant participates as a function of this new incumbent's value. Successively lower degrees of selection change the bid function so that when $\alpha \to 1$ it approaches the bid function in the LS (no selection) model (the bold line), which is a step function with a jump at a value of 119 (the level of the incumbent's value that deters all future entry). The slope of the bid function is more gradual for lower αs since the probability that the final round potential entrant participates declines more smoothly when α is lower.

[21] There would also have been no change in the standing bid if the entrant had come in, because the entrant's value was below the current bid, so the standing bid would not have risen in the knockout.

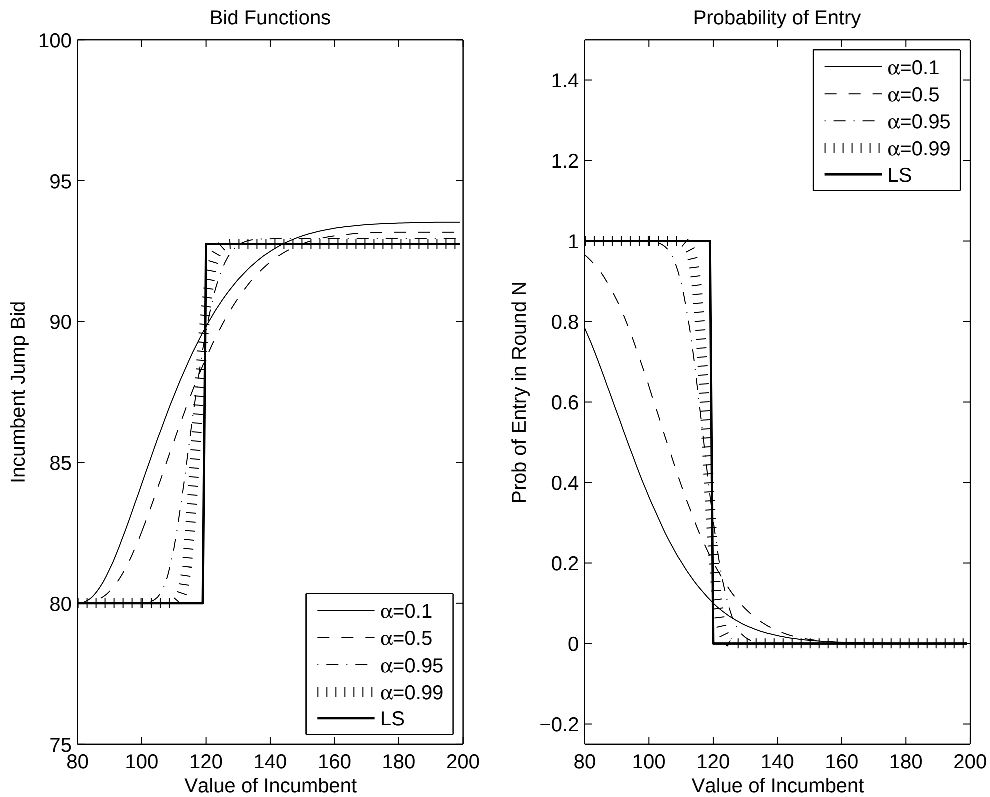


Figure 1: Penultimate round bid function for a new incumbent and probability of entry for the final round potential entrant, with symmetric firms, values LN(4.5,0.2) on [0,200], K=1 and standing bid of 80.

3 Comparison of Expected Revenues and Efficiency

Before introducing specific parameters estimated from data for USFS timber auctions, we present a more general comparison of expected revenues and efficiency between the sequential mechanism and the simultaneous entry auction. We see this general comparison as valuable, because it shows that our results in the empirical application are not going to be particularly sensitive to the parameters that we estimate, and they provide guidance about when auctions should perform well in other settings. Additionally, we allow a reserve price to be used in the simultaneous auction but restrict attention, for now, to a sequential mechanism with no reserve. In this way the results are biased *against* a seller preferring the sequential mechanism.

We focus on how the performance of the mechanisms depends on the level of entry costs (K) and the precision of the signal. We measure the precision of the signal by the parameter $\alpha = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_V^2}$, where a higher value of α indicates that signals are less precise. As a base case, we consider 8 symmetric firms whose values are distributed $LN(4.5, 0.2)$ so that the value distribution has a mean of 91.6 and a standard deviation of 18.6.

Figure 2 shows the results of comparing expected revenues from the sequential mechanism (with no reserve) and a simultaneous entry auction with an optimal reserve in (K, α) space.[22] Filled squares represent outcomes where the expected revenues from the sequential mechanism are higher by more than 4% (of auction revenues), while hollow squares are outcomes where they are higher but only by between 0.1% and 4%. Diamonds represent cases where the simultaneous auction gives higher revenues. Crosses on the grid mark locations where the difference in revenues is less than 0.1%. Due to small numerical errors in solving differential equations and simulation error in calculating expected revenue, we take the conservative approach of not signing revenue differences in these cases.

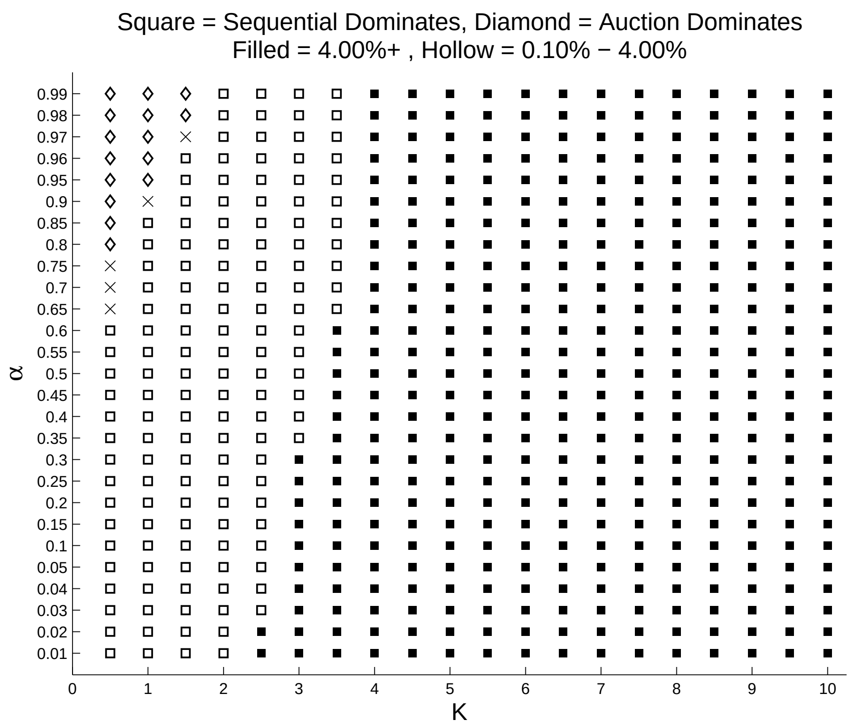


Figure 2: Expected revenue comparison for 8 symmetric firms, values LN(4.5,0.2), optimal reserve price in auction, no reserve price in sequential mechanism.

The results indicate that the sequential mechanism generally produces higher expected revenues than the auction, even when no reserve price is used in the sequential mechanism, whereas the optimal reserve price is used in the auction. The exception is for very low values of K and high α, but the revenue advantage of the auction is always small (the maximum difference is 1.1%). These points are consistent with BK's theoretical results as their model assumes no signals and requires that at least two firms will enter the auction, which implies that K must be low.

[22]Sequential (auction) mechanism's expected revenues are calculated using 200,000 (5,000,000) simulations.

Economists and mechanism designers may be concerned with efficiency as well as revenues, and the sequential mechanism outperforms the auction along both dimensions. With no selection, low entry costs and symmetric bidders, BK show that the sequential mechanism is always more efficient, where efficiency is measured by the expected value of the winner less total entry costs paid. Increasing selection reduces the overall amount of entry and further wasteful entry costs, which serves to raise efficiency. It can also be the case that the expected value of winner in the sequential mechanism is higher than that in the auction, despite the lower number of entrants. For example, taking the parameters from Figure 2, when $K = 1$, the expected value of the winner is greater in the sequential mechanism for $\alpha \leq 0.25$. For higher K the expected value of the winner is always greater in the sequential mechanism. This is contrary to results from a no selection model, where "the expected value of the top bidder in the auction must be higher than in the sequential mechanism" (BK p. 1546). However, even when the expected value of the winner in the auction exceeds that in the sequential mechanism, the elimination of wasteful entry costs tends to sufficiently compensate so as to raise efficiency in the sequential mechanism. For example, over the grid given in Figure 2, there is no case in which the auction is more efficient than the sequential mechanism.

Regarding asymmetries, in a simultaneous entry auction weaker bidders need to consider the odds of competing against stronger bidders. While this is still true in the sequential mechanism, if the weaker bidders are approached last, given the separating equilibrium they know the values of the higher types that have entered. This permits more efficient entry of the weaker bidders and achieves a more efficient allocation of the good relative to the auction. For example suppose that $K = 5$ and $\alpha = 0.4$, $N = 4$ and the first two bidders approached have values proportional to $LN(4.5, 0.2)$, while the last two bidders approached have values proportional to $LN(4.4, 0.2)$.[23] The probability that each of the weaker firms enters the simultaneous auction is 0.20 and the probability that one of them wins is only 0.17. On the other hand, in the sequential mechanism the entry probabilities are 0.143 for the first one and 0.139 for the second and the probability that one of them wins is 0.28. This is much closer to the probability that one of the weaker firms will have the highest value (0.33). In this case, the sequential mechanism's expected revenues of 83.34 exceed those of the auction, which are 78.40.

When bidders are asymmetric, sellers may prefer a first price auction with type-specific reserve prices to a second price auction with a uniform reserve. However, continuing with this example, even a first price auction with type-specific optimal reserve prices only generates expected revenues of 80.41, and so it is outperformed by the sequential mechanism with no

[23] Our simulations show that approaching all of the high value firms first, followed by all of the low values firms is better than doing the opposite.

reserve price.

The separating equilibrium also has the effect of improving sellers' abilities to appropriate the larger rents in the sequential mechanism. This is important to note since, in BK's partial pooling equilibrium, high value participants' ability to completely forestall entry permits them to capture enough of the sequential mechanism's greater overall rents that sellers expect lower revenues than in the auction. Revenues in the auction are obviously determined by the second highest value of the bidders. In the sequential mechanism, revenues are determined by the maximum of the second highest value of participants and the deterring bid of the eventual winner. Thus, the larger rents can be appropriated either through stronger actual competition (the value of the second highest bidder is greater in the sequential mechanism) or by forcing the eventual winner to bid more aggressively to deter future potential competitors from participating. There are forces working towards and against encouraging stronger actual competition in the sequential mechanism. On the one hand, it can do better at selecting high value potential entrants into the mechanism. This is evidenced by the expected value of the winner in the sequential process sometimes exceeding that in the auction despite reduced entry. On the other hand, fewer bidders tend to participate, which lowers the expected value of the second highest value bidder. However, the threat of future potential entry by strong bidders forces the eventual winner to bid more aggressively and this tends to raise the seller's revenues. For example, the expected value of the second highest entrant is greater in the auction than in the sequential mechanism for each grid point in Figure 2. However, the winner's deterring bid is high enough to earn the seller higher revenues for almost all cases. In this sense, we find that the threat of future potential competition in the sequential mechanism leads to higher prices than does the greater actual competition in a simultaneous bid auction.

By way of example, Figure 1 illustrates how selection's effect on bid functions and equilibrium probabilities of entry serve to increase revenues in the sequential mechanism. Compared to the LS model, the bid functions with selection are higher for values less than 119, which is where this incumbent's value distribution is most dense. For example, when $\alpha = 0.1$, the mean and 90th percentile of a new incumbent who would find himself in the position of submitting a jump bid in the penultimate round are 101 and 121, respectively. While for a portion of the value distribution the bid functions are lower than when $\alpha = 1$, they again (slightly) exceed this bid function for very high values. This is because there is always a chance that the final potential entrant receives an optimistic signal and enters (unless it is inferred that the new incumbent's value is greater than $\overline{V} - K$, here 199), which cannot happen when the incumbent has a high value in BK's model. This is clearly illustrated in the right panel, which gives the probability that the final round potential entrant participates as a function of the new incumbent's value. This probability of entry is also a step function in the LS model. Once there is some selection, there is always a chance that the final round potential entrant participates, even if the incumbent is thought to have a high value (again, assuming its value is less than 199).

3.1 Sequential Mechanism with Reserve Prices

The numerical examples above indicate that a simple, stylized version of real-world sequential mechanisms tends to outperform the commonly used auction, even when the optimal reserve price is set in the auction. The sequential mechanism's advantage over the auction could be increased through additional design elements, an obvious option being a reserve price. Figure 3 computes expected revenues when an optimal reserve price is added to each mechanism when there are five or eight symmetric bidders using the same value distribution parameters as before and assuming $K = 5$. For the sequential mechanism, only one reserve price is used, which is constant across all rounds in the mechanism. Generally, the seller could do better with a round-specific reserve price, but we view a constant reserve price as approximately imposing the same informational demands on the seller as does setting the optimal reserve price in the simultaneous auction.

Figure 3 shows that adding a constant reserve price to the sequential mechanism may substantially improve revenues. The reserve price affects sequential mechanism revenues in two ways. First, in the event that no firm has entered through the first $N - 1$ rounds, a reserve price guards the seller against giving the good away for free to the last potential entrant. Second, a reserve price raises the first entrant's deterring bid function.

The effect of a reserve price varies across mechanisms and for different values of N and α. There are two main reasons for this. First, when entry is endogenous, a reserve price has a smaller impact when the level of entry is greater, as is generally the case (i) in the auction or (ii) when N is greater, as is clearly shown in Figure 3. Second, a reserve price excludes some bidders and if these were valuable to the seller, this reduces the value of a reserve price. This effect can be seen by noticing that the impact of a reserve price in the sequential mechanism falls for higher values of α: less selection implies that marginal and inframarginal entrants are more similar, which makes excluded bidders more valuable to the seller (it is also true that the level of entry increases in α, which also limits a reserve price's impact).

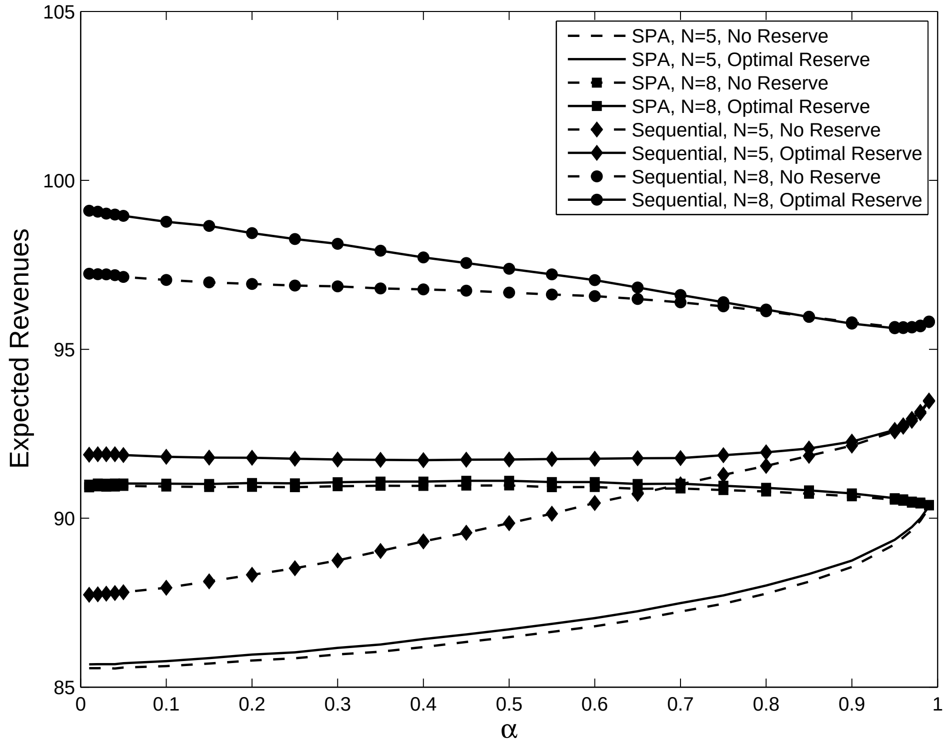


Figure 3: Expected revenue comparison for varying N, with and without reserves. Firms are symmetric with values distributed LN(4.5,0.2) and $K = 5$.

4 Empirical Application

We now turn to our empirical application that focuses on USFS timber auctions, which we view as a sensible environment for evaluating the effects of switching to a sequential mechanism. First, we find that these auctions are characterized by a costly and moderately selective entry process, features that we view as holding more generally across a wide variety of auction environments. Second, unlike other environments, such as the M&A market, we are able to observe the sale of many similar objects which facilitates estimation of bidder values. Third, while a great deal of work has concentrated on auction design tools, such as reserve prices, as means to increasing revenues in timber auctions, we show that a shift to a sequential sales process has a much larger impact. We are brief in our discussions of some reduced form evidence of selection and estimation method since Roberts and Sweeting (2011) provide a more detailed discussion of these topics. Additionally, Gentry and Li (2010) explore conditions under which an imperfectly selective entry model is non-parametrically identified for first price auctions.

4.1 Data

We analyze federal auctions of timberland in California. In these auctions the USFS sells logging contracts to individual bidders who may or may not have manufacturing capabilities (mills and loggers, respectively). When the sale is announced, the USFS provides its own "cruise" estimate of the volume and value of timber for each species on the tract as well as estimated costs of removing and processing the timber. It also announces a reserve price and bidders must indicate a willingness to pay at least this amount to qualify for the auction. After the sale is announced, interested potential bidders perform their own private cruises in order to assess the tract's value. These cruises are informative about the tract's volume, species make-up and timber quality.

We assume that bidders have independent private values. This assumption is also made in other work with similar timber auction data (see for example Baldwin, Marshall, and Richard (1997), Haile (2001) or Athey, Levin, and Seira (forthcoming)). A bidder's private value is primarily related to its own contracts to sell the harvest, inventories and private costs of harvesting. In addition, we focus on the period 1982-1989 when resale, which can introduce a common value element, was limited (see Haile (2001) for an analysis of timber auctions with resale).

We also assume non-collusive bidder behavior. While there has been some evidence of bidder collusion in open outcry timber auctions, Athey, Levin, and Seira (forthcoming) find strong evidence of competitive bidding in these California auctions.

Our model assumes that bidders receive an imperfect signal of their value and they must pay a participation cost to enter the auction.[24] We interpret the USFS's publicly available tract appraisal and a firm's own knowledge of its sales contracts and capabilities as generating its pre-entry signal. Participation in these auctions is costly for numerous reasons. In addition to the cost of attending the auction, a large fraction of a bidder's entry cost is its private cruise. People in the industry tell us that firms do not bid without doing their own cruise, which can provide information that bidders find useful, such as trunk diameters, but is not provided in USFS appraisals.

We use data on 887 ascending auctions.[25] Table 2 shows summary statistics for our sample. Bids are given in $ per thousand board feet (mbf) in 1983 dollars. The average mill bid is 20.3% higher than the average logger bid. As suggested in Athey, Levin, and Seira (forthcoming), mills may be willing to bid more than loggers due to cost differences or the imperfect competition loggers face when selling felled timber to mills.

[24] We note that we are not the first to model a costly entry decision into these auctions (e.g. Athey, Levin, and Seira (forthcoming)).

[25] Roberts and Sweeting (2011) include a detailed description of the sample selection process.

Variable	Mean	Std. Dev.	25^{th}-tile	50^{th}-tile	75^{th}-tile	N
WINNING BID ($/mbf)	86.01	62.12	38.74	69.36	119.11	847
BID ($/mbf)	74.96	57.68	30.46	58.46	105.01	3426
LOGGER	65.16	52.65	26.49	49.93	90.93	876
MILL	78.36	58.94	32.84	61.67	110.91	2550
LOGGER WINS	0.15	0.36	0	0	0	887
FAIL	0.05	0.21	0	0	0	887
ENTRANTS	3.86	2.35	2	4	5	887
LOGGERS	0.99	1.17	0	1	1	887
MILLS	2.87	1.85	1	3	4	887
POTENTIAL ENTRANTS	8.93	5.13	5	8	13	887
LOGGER	4.60	3.72	2	4	7	887
MILL	4.34	2.57	2	4	6	887
SPECIES HHI	0.54	0.22	0.35	0.50	0.71	887
DENSITY (hundred mbf/acre)	0.21	0.21	0.07	0.15	0.27	887
VOLUME (hundred mbf)	76.26	43.97	43.60	70.01	103.40	887
RESERVE ($/mbf)	37.47	29.51	16.81	27.77	48.98	887
SELL VALUE ($/mbf)	295.52	47.86	260.67	292.87	325.40	887
LOG COSTS ($/mbf)	118.57	29.19	99.57	113.84	133.77	887
MFCT COSTS ($/mbf)	136.88	14.02	127.33	136.14	145.73	887

Table 2: Summary statistics for sample of California ascending auctions from 1982-1989. All monetary figures in 1983 dollars. SPECIES HHI is the Herfindahl index for wood species concentration. SELL VALUE, LOG COSTS and MFCT COSTS are USFS estimates of the value of the tract and the logging and manufacturing costs of the tract, respectively.

We define potential entrants as the auction's bidders plus those firms who bid within 50 km of an auction over the next month. One way of assessing the appropriateness of this definition is that 98% of the bidders in any auction also bid in another auction within 50 km of this auction over the next month and so we are unlikely to be missing many actual potential entrants. The median number of potential bidders is eight (mean of 8.93) and this is evenly divided between mills and loggers.

In Table 2, entrants are defined as the set of bidders we observe at the auction, even if they did not submit a bid above the reserve price.[26] The median number of mill and logger entrants are three and one, respectively. Among the set of potential logger entrants, on average 21.5% enter, whereas on average 66.1% of potential mill entrants enter. The differences in bids and entry decisions are consistent with mills having significantly higher values than loggers.[27]

[26]However, in our preferred empirical specification below, we interpret the data more cautiously and allow bidders that do not submit bids to have entered (paid K), but learned that their value was less than the reserve price.

[27]Roberts and Sweeting (2011) present evidence that differences in values, and not entry costs, explain

4.2 Evidence of Selection

Roberts and Sweeting (2011) present reduced form evidence that the data are best explained by a model allows for selection. There are two main pieces of evidence. First, Athey, Levin, and Seira (forthcoming) show that in the type-symmetric mixed strategy equilibrium of a model with endogenous, but non-selective, entry and asymmetric bidder types, whenever the weaker type enters with positive probability, the stronger type enters with probability one. Thus, for any auction with some logger entry, a model with no selection would imply that all potential mill entrants enter. In 54.5% of auctions in which loggers participate, and there are some potential mill entrants, some, but not all, mills participate. Likewise, they show that whenever the stronger type enters with probability less than one, a model with no selection implies that weaker types enter with probability zero. However, in the data we find that in 61.1% of auctions in which only some mill potential entrants participate and potential logger entrants exist, some loggers enter. A model with selective entry can rationalize partial entry of both bidder types into the same auction.

Second, a model without selection implies that bidders are a random sample of potential entrants. Roberts and Sweeting (2011) test this by estimating a Heckman selection model with the exclusion restriction that potential competition affects a bidder's decision to enter an auction, but has no direct effect on values. The second stage regression of all bids on auction covariates and the estimated inverse Mills ratio from a first stage probit of the decision to participate shows a positive and significant coefficient on the inverse Mills ratio. This is consistent with bidders being a selected sample of potential entrants.

The evidence presented in this section strongly suggests that the entry process is selective. However, it does not pin down the degree of selection. Therefore, we now describe how we estimate our model to measure the degree of selection.

4.3 Estimation Using Importance Sampling

To take the model to data, we need to specify how the parameters of the model may vary across auctions, as a function of observed auction characteristics and unobserved heterogeneity. Both types of heterogeneity are likely to be important as the tracts we use differ greatly in observed characteristics, such as sale value, size and wood type, and they also come from different forests over several years so they may differ in other characteristics as well. Both observed and unobserved (to the econometrician) heterogeneity may affect entry costs and the degree of selection, as well as mean values.[28]

---

why mills are more likely than loggers to enter an auction.

[28] In the specification below, α is not a function of observables as when we allowed for this the estimated effects of observables on the degree of selection were small and imprecise.

Our estimation approach is based on Ackerberg (2009)'s method of simulated maximum likelihood with importance sampling. We fully describe our estimation method in the appendix and in Roberts and Sweeting (2011). However, here we note several features of our specification.

We assume that the parameters are distributed across auctions according to the following distributions, where X_a is a vector of observed auction characteristics and $TRN(\mu, \sigma^2, a, b)$ is a truncated normal distribution with parameters μ and σ^2, and upper and lower truncation points a and b.

$$\text{Location Parameter of Logger Value Distribution: } \mu_{a,\text{logger}} \sim N(X_a\beta_1, \omega^2_{\mu,\text{logger}})$$
$$\text{Difference in Mill/Logger Location Parameters: } \mu_{a,\text{mill}} - \mu_{a,\text{logger}} \sim TRN(X_a\beta_3, \omega^2_{\mu,\text{diff}}, 0, \infty)$$
$$\text{Scale Parameter of Mill and Logger Value Distributions: } \sigma_{Va} \sim TRN(X_a\beta_2, \omega^2_{\sigma_V}, 0.01, \infty)$$
$$\alpha\text{: } \alpha_a \sim TRN(\beta_4, \omega^2_{\alpha}, 0, 1)$$
$$\text{Entry Costs: } K_a \sim TRN(X_a\beta_5, \omega^2_K, 0, \infty)$$

These specifications reflect our assumptions that σ_V, α and K are the same for mills and loggers within any particular auction, even though they may differ across auctions.

To apply the estimator, we also need to define the likelihood function based on the open outcry auction data. Two problems arise when interpreting these data. First, a bidder's highest announced bid in an open outcry auction may be below its value, and it is not obvious which mechanism leads to the bids that are announced (Haile and Tamer (2003)). Second, if a firm does not know its value when taking the entry decision, it may learn (after paying the entry cost) that its value is less than the reserve price and so not submit a bid. We take a conservative approach (the details of which are provided in the appendix) when interpreting the data by assuming that the winning bidder has a value greater than the second highest bid, the second highest observed bid is equal to the value of the second-highest bidder, all other bidders had values less than the highest observed bid and that potential entrants that we do not see bid may or may not have paid the entry cost.

5 Empirical Results

In this section we present estimates of our structural model and counterfactual results measuring the benefits to the USFS of switching from the current simultaneous entry and simultaneous bid auction to our simple sequential process.

5.1 Parameter Estimates

Table 3 presents the parameter estimates for our structural model.[29] We allow the USFS estimate of sale value and its estimate of logging costs to affect mill and logger values and entry costs since these are consistently the most significant variables in regressions of reserve prices or winning bids on observables, including controls for potential entry. We also control for species concentration since our discussions with industry experts lead us to believe that can matter to firms. We allow for auction-level unobserved heterogeneity (to the econometrician) in all parameters. The righthand columns show the mean and median values of the structural parameters when we take 10 simulated draws of the parameters for each auction. For the rest of the paper, we refer to these as the "mean" and "median" values of the parameters. All standard errors are based on a non-parametric bootstrap, where both auctions and draws are re-sampled, with 100 repetitions.

The coefficients show that tracts with greater sale values and lower costs are more valuable, as one would expect. There is unobserved heterogeneity in values across auctions (the standard deviation of μ_{logger}) and some unobserved heterogeneity in the difference between mill and logger mean values across auctions (the standard deviation of $\mu_{\text{mill}} - \mu_{\text{logger}}$).

Based on the mean value of the parameters, the mean values of mills and loggers in the population are, in 1983 dollars, \$61.95/mbf and \$42.45/mbf, respectively, a 46% difference. We estimate a mean entry cost of \$2.05/mbf, also in 1983 dollars. One forester we spoke with estimated modern day cruising costs of approximately \$6.50/mbf, or \$2.97/mbf in 1983 dollars. It is sensible that our estimate is less than the forester's estimate if firms in our data are able to use any information they learn when deciding whether to enter other auctions.

Our estimates of the αs across auctions indicate a moderate amount of selection in the data. This is illustrated by the difference in expected values for marginal and inframarginal bidders in a representative auction where the reserve price and the number of potential mill and logger entrants are set to their respective medians of \$27.77/mbf, four and four. Based on the mean parameter values, the expected values of a marginal and inframarginal mill entrant are \$45.22/mbf and \$68.13/mbf, respectively (the former is lower than the population average because most mills enter). The comparable numbers for loggers are \$48.13/mbf and \$59.80/mbf, respectively.

Our estimation approach assumes that, if there are multiple equilibria, the firms will play the equilibrium where mills have the lower S'^*. We can check whether our parameter estimates can support multiple equilibria by plotting type-symmetric "equilibrium best response functions" for mills and loggers for each auction. For each auction, our parameter estimates

[29] Roberts and Sweeting (2011) discuss alternative estimation methods that were attempted, such as Nested Pseudo-Likelihood, and model fit.

		β parameters			ω parameter		
Parameter	Constant	log SELL VALUE	log LOG COSTS	SPECIES HHI		Mean	Median
$\mu_{a,\text{logger}}$	-9.6936	3.3925	-1.2904	0.2675	0.3107	3.5824	3.5375
$\sim N(X_a\beta_1, \omega^2_{\mu,\text{logger}})$	(1.3690)	(0.1911)	(0.1332)	(0.1386)	(0.0213)	(0.0423)	(0.0456)
$\mu_{a,\text{mill}} - \mu_{a,\text{logger}}$	3.6637	-0.4998	-0.0745	-0.1827	0.1255	0.3783	0.3755
$\sim TRN(X_a\beta_3, \omega^2_{\mu,\text{diff}}, 0, \infty)$	(0.8890)	(0.1339)	(0.0919)	(0.1007)	(0.0163)	(0.0242)	(0.0249)
σ_{Va}	4.0546	-0.7379	0.1393	0.0895	0.0796	0.5763	0.5770
$\sim TRN(X_a\beta_2, \omega^2_{\sigma_V}, 0.01, \infty)$	(0.7872)	(0.0994)	(0.1025)	(0.0813)	(0.0188)	(0.0273)	(0.0302)
α_a	0.7127	-	-	-	0.1837	0.6890	0.6992
$\sim TRN(\beta_4, \omega^2_\alpha, 0, 1)$	(0.0509)				(0.0446)	(0.0362)	(0.0381)
K_a	1.9622	-3.3006	3.5172	-1.1876	2.8354	2.0543	1.6750
$\sim TRN(X_a\beta_5, \omega^2_K, 0, \infty)$	(13.2526)	(2.7167)	(2.4808)	(1.5721)	(0.6865)	(0.2817)	(0.3277)

Table 3: Simulated maximum likelihood with importance sampling estimates allowing for non-entrants to have paid the entry cost. The rightmost columns show the mean and median values of the structural parameters when we take 10 simulated draws of the parameter for each auction. Standard errors based non-parametric bootstrap with 100 repetitions. $TRN(\mu, \sigma^2, a, b)$ is a truncated normal distribution with parameters μ and σ^2, and upper and lower truncation points a and b. Based on 887 auctions.

support only a single equilibrium. This is because our estimates imply a large difference in the mean values of loggers and mills, relatively low entry costs and a moderate amount of selection, all of which promote a unique equilibrium.

5.2 Counterfactual Results

Table 4 compares expected revenues and efficiency from the sequential mechanism and the simultaneous entry auction for a range of parameters and different numbers of firms. The simulations assume mills are approached first (in a random order) followed by loggers, although we have found some cases where a different order can strengthen the results below.

The first line in Table 4 gives the results for the representative auction (four mills and four loggers) based on the mean parameter estimates used in the constructing the figures above. Relative to setting no reserve price in the simultaneous entry, simultaneous bid auction, the sequential mechanism with no reserve price improves the USFS's revenues by 1.81%. For a tract of average size (7,626 mbf) the expected revenue difference would be $9,834.

The increase in revenues in this representative case of switching from the simultaneous bid auction with no reserve price to the sequential mechanism with no reserve price is 9.05 times as large as the improvement from using an optimal reserve in the simultaneous bid auction, which is just 0.2%. The finding that the revenue increase from using the sequential mechanism is much larger than the returns to using a reserve price in the current auction format is important since understanding optimal reserve price policies for timber auctions has been the subject of significant interest (examples include Mead, Schniepp, and Watson (1981), Paarsch (1997), Haile and Tamer (2003), Li and Perrigne (2003) and Aradillas-Lopez, Gandhi, and Quint (2010)). Additionally, the sequential mechanism provides an easily implementable mechanism that does not require the USFS to possess detailed information on all of the model's primitives. Such information would be required to set an optimal reserve price. However, were the USFS to possess such information, a reserve price could also be set in the sequential mechanism. If a reserve price is used in the sequential mechanism, the increase in revenues becomes 10.43 times as large as the gain to setting an optimal reserve price in the auction. This advantage would increase if we considered round-specific, optimal reserve prices in the sequential mechanism.

Not only does the sequential mechanism have a much larger impact on revenues than does setting an optimal reserve price in the standard auction format, it also increases efficiency, as shown in the penultimate column in Table 4. In the representative case given in the first row of the table, the USFS captures the majority of the increase in surplus, but expected firm profits still increase in the sequential mechanism. As mentioned in Section 3, the sequential mechanism tends to promote more efficient entry of weaker bidders and this increases their

Case	Parameters							Expected Revenues ($/mbf)			Expected Efficiency SPA → SEQ, $R=0$	
	N_{mill}	N_{logger}	μ_{logger}	μ_{diff}	σ_V	α	K	SPA $R=0$	SEQ $R=0$	SPA $R=R^*$	Total Δ ($/mbf)	% of Δ → to Firms
1	4	4	3.582	0.378	0.576	0.689	2.05	71.25	72.54	71.39	1.88	31.38
2	*1*	4	3.582	0.378	0.576	0.689	2.05	50.96	52.32	51.63	1.45	6.21
3	*7*	4	3.582	0.378	0.576	0.689	2.05	83.91	85.21	83.98	2.16	39.81
4	4	*0*	3.582	0.378	0.576	0.689	2.05	64.42	64.46	64.69	0.91	95.60
5	4	*8*	3.582	0.378	0.576	0.689	2.05	75.87	77.60	75.97	2.31	25.11
6	4	4	*2.921*	0.378	0.576	0.689	2.05	34.10	35.50	34.34	1.58	11.39
7	4	4	*4.243*	0.378	0.576	0.689	2.05	142.75	143.75	142.81	2.25	55.56
8	4	4	3.582	*0.169*	0.576	0.689	2.05	61.90	62.93	62.02	1.88	45.21
9	4	4	3.582	*0.587*	0.576	0.689	2.05	83.91	85.32	84.07	1.66	15.06
10	4	4	3.582	0.378	*0.349*	0.689	2.05	57.13	58.79	57.47	1.51	-9.93
11	4	4	3.582	0.378	*0.804*	0.689	2.05	91.08	91.64	91.16	1.98	71.72
12	4	4	3.582	0.378	0.576	*0.505*	2.05	70.64	72.36	71.02	1.95	11.79
13	4	4	3.582	0.378	0.576	*0.872*	2.05	71.89	73.08	71.91	2.03	41.38
14	4	4	3.582	0.378	0.576	0.689	*0.39*	75.74	75.71	75.75	0.54	105.56
15	4	4	3.582	0.378	0.576	0.689	*3.72*	66.69	69.32	67.12	2.93	10.24

Table 4: Comparing the impact of the sequential mechanism on expected revenues and efficiency. The first line shows the results for the representative auction with four mills and four loggers and the mean parameter estimates from Table 3. This is what we refer to as the "representative case". Italics indicate changes from this representative auction and these changes are ±1 standard deviation changes based on our estimates of the distributions of the structural parameters. SEQ refers to the sequential mechanism, SPA refers to the auction, $R = 0$ and $R = R^*$ indicates that either no reserve price or the optimal reserve price is used, respectively. The first first three columns following the parameters give the expected revenues in the three mechanisms. The final two columns compare the expected efficiency of a sequential mechanism and an auction with no reserve. The first of these is the total change in efficiency. The second gives the percent of the change which accrues to firms. Auction results based on 5,000,000 simulations and sequential results based on 200,000 simulations.

expected profits. In the USFS auctions, switching from the current auction format to the sequential mechanism tends to increase expected logger profits without substantially harming those of mills. For example, in the representative case, expected logger profits increase 21% when the sequential mechanism is used, while mill profits only fall by 0.60%.

The other rows in the table compare outcomes when we increase or decrease the number of potential entrants or structural parameters by one standard deviation (the changing parameter is in *italics*), reflecting the fact that our estimates imply that the coefficients will differ across sales. The cases we consider indicate that using the sequential mechanism generally raises expected revenues. In case 14 the entry cost is very low and in either mechanism almost all firms participate so that revenues are essentially the same. In all cases, once a constant reserve price is used in the sequential mechanism, it earns higher revenues than the current auction format even with an optimal reserve price. We can see that setting a reserve price in the standard auction format is particularly ineffective when there are many potential entrants or when entry is less selective (α is high). In all cases the sequential mechanism increases efficiency and in only one example does total bidder surplus fall (case 10). The finding from the first row that loggers benefit from switching to the sequential mechanism holds in all rows. Additionally, when expected mill profit falls, it tends to be by a small amount, and in some cases it rises. As an example, in case 8, when μ_{diff} is low (0.169), loggers' expected profit increases by 10.18% and mills' increases by 1.40%.

The USFS also uses first price, sealed bid auctions to sell timber. We can also compare the performance of the sequential mechanism to this alternative. Across all of the cases in Table 4, with the exception of cases 6 and 14, a sequential mechanism with no reserve price earns the USFS higher revenues than a first price auction with an optimal reserve price. Introducing a reserve price to the sequential mechanism increases its advantage over the first price auction by even more so that it now dominates in all cases.

While we believe that the simultaneous entry auction is the natural way to think about how USFS auctions currently operate, we have also computed expected auction revenues if, instead, firms enter sequentially (in the same order as the sequential mechanism) before simultaneously submitting bids. For the representative auction, expected revenues in this case are \$71.84/mbf which is still less than the revenues from the sequential bidding mechanism. This pattern holds more generally in the other rows in Table 4 where we were able to solve a sequential entry auction game: in only two cases (6 and 14) did the sequential entry auction give higher revenues than the sequential mechanism with no reserve price and in both cases the differences were small (\$0.40/mbf and \$0.06/mbf, respectively).[30]

[30] As explained in footnote 14, there is a high computational burden to solving the sequential entry auction because it is necessary to solve for a threshold as a function of all possible histories of the game simultaneously. For cases 3, 5, 13 and 15 we could not do so satisfactorily.

Government sales and procurement programs often have distributional requirements that a certain portion of contracts be awarded to targeted firms. The US federal government seeks to award 23% of the $400 billion worth of annual contracts to small businesses (see Athey, Coey, and Levin (2011) for additional discussion). The primary ways of favoring smaller businesses are through set-asides, where only targeted firms can participate, and bid subsidies for preferred firms. The USFS has historically used set-asides and recent work (Athey, Coey, and Levin (2011)) suggests this may come at a substantial revenue and efficiency loss relative to using bid subsidies. In this light, our findings that the USFS can increase revenues, efficiency and the profits of loggers with only small decreases in mill profits (and sometimes increases) by switching from the current auction format to a sequential mechanism may be particularly useful. Although a full comparison of bid subsidies, set-asides and the sequential mechanism is beyond the scope of this paper, we see our findings as suggesting that the sequential mechanism may present procurement agencies with an effective alternative method for allocating projects to targeted bidders. Additionally, the sequential mechanism requires only that the agency be able to identify targeted firms, which is also required in the use of set-asides and subsidies, and does not require determining optimal subsidy amounts or even setting reserve prices. For timber auctions, we have been told by USFS officials that they believe that they can accurately identify the set of potential entrants for any given sale. Even if at times they are unsure, it would be straightforward to allow potential participants to costlessly identify themselves before the full details of a sale are announced.

Our discussion so far has largely ignored potential practical impediments to implementing the sequential mechanism. First, were the USFS to use the sequential mechanism, there may be concern that approaching firms in an order places some of them at an advantage over others and may lead firms to try to affect the order in which they are approached. However, for all of the examples that we have considered, expected firm profits are fairly constant across the order of moves within bidder type, and there is no systematic pattern suggesting that a particular spot in the order is best. Intuitively, while the first potential entrant will be more likely to participate, he also must pay more to win. For example, in the representative auction, where the four mills are approached first followed by the four loggers, the expected profits (in \$/mbf) by order are $\{6.07, 6.09, 6.14, 6.18, 1.08, 1.05, 1.09, 1.04\}$. The maximum amount by which expected mill (logger) profits differ in this case is 0.016 (0.042). Second, there may be some concern about whether the USFS can commit to an order. However, repeated use of the mechanism likely would incentivize the USFS to maintain its credibility through consistent commitment to stated orders. Additionally, the lack of variation of profits across spots in the order could mean that firm lobbying efforts, which might dissuade a seller from sticking to a stated order, are likely to be small. Third, collusion may be a concern given the existing evidence from other USFS regions consistent with noncompetitive bidding (Athey, Levin, and Seira (forthcoming)). However, as Bulow and Klemperer (2009) note (their footnote 40), the "simple auction is perhaps more easily undermined, than a sequential process, by collusion."

There may also be some concern that switching to the sequential mechanism would greatly increase the time required to sell any stand of timber. While the length of the bidding process would necessarily increase, we note that there is already a sizable gap (over a month) between when a sale is announced and when it is completed.[31] Since cruising takes between a day and seven days, depending on the size of the sale, even in the extreme (assuming a large sale in which 8 potential bidders all decide to participate), a sequential mechanism could be run in under two months. Often the sequential process could happen much faster, but even an extra month may be a small price to pay to realize the sequential mechanism's advantages.

We have also ignored the USFS's cost of switching to the sequential mechanism. Although the cost of selling a stand of timber is likely to be similar across mechanisms, there may be a fixed cost associated with switching from the currently used format to a sequential process, which would have to be measured against the potential gains from doing so. Based on the 15 cases in Table 4 alone, USFS revenues would increase by approximately $315,000 (in 2011 dollars) compared to using the current auction format with no reserve price.[32] Given that these 15 sales represent less than 0.3% of the tracts sold by the USFS in CA between 1982 and 1989, this one-time, sunk cost is likely to be small relative to the associated increase in revenues.

6 Conclusion

This paper compares the performance of a sequential and a simultaneous bidding mechanism in an environment where it is costly for potential buyers to participate and they receive imperfectly-informative signals about their values prior to deciding whether to enter, so that the entry process is selective. In contrast to results when there is no selection, a very simple sequential mechanism can generate higher expected revenues for the seller than the commonly used auction, and it also has an efficiency advantage so that buyers may prefer it is as well. The revenue result holds even though there is less entry (actual competition) into the sequential mechanism. Instead, with selection, the sequential mechanism can do a better job of allocating the good to the firm with the highest value and this fact, combined

[31] In fact the gap is usually much longer since the USFS must file documents to comply with the National Environmental Policy Act.

[32] These calculations assume the average tract size in the 15 cases is 7,626 mbf.

with the feature that firms with high values have to bid aggressively in an attempt to deter future entry, provides its revenue advantage.

We view our results as relevant and important for at least three reasons. First, a selective entry process is likely an appropriate description of many real-world settings where a bidding process is used to sell an asset. This is because potential buyers often possess some pre-existing knowledge of their match for the asset, but will need to conduct costly additional research to determine how much they should be willing to pay. In these cases, our results point to conclusions about how bidding should be structured that are different to those in the existing literature. Our model also allows us to explain certain features of the data, such as jump-bids not deterring all future entry in takeover contests, so that multiple jump bids, sometimes by different firms, are observed (e.g., Betton and Eckbo (2000) or Betton, Eckbo, and Thorburn (2008)). These facts cannot be explained by a model with no selection.

Second, the revenue differences that we identify are not trivial. As a comparison, we consider the seller's return to setting an optimal reserve price in a simultaneous auction, which is the type of relative small design change that is the focus of the existing empirical literature. For the representative auction in our data, we estimate that the seller's return to switching to the sequential mechanism would be nine times greater than the return to setting the optimal reserve price. The absolute difference in revenues can also be large when entry costs are higher or entry is more selective than we estimate to be the case in USFS auctions.

Third, our results are directly relevant to an on-going legal debate about how corporate sales should be structured in order to allow boards to fulfill their *Revlon* duties to maximize shareholder value. At the very least, our results suggest that there are circumstances in which a sequential bidding process will achieve this more effectively than a simultaneous one, and they highlight two factors (entry costs and selection) on which the results are likely to depend. One concern that has been raised with sequential processes is that all potential buyers are not treated equally, so that firms that move first may be able to deter later ones and retain a right to match the prices offered by any later competition that emerges. This is true in our model, but it does not necessarily mean that the firms that move first earn higher revenues. In fact, our results suggest that expected payoffs are fairly equal across the order in the presence of selection, because early movers also pay entry costs more often and are less likely to win when they enter.

One might believe that while simultaneous auctions often operate in exactly the way modeled here, the stylized sequential mechanism that we consider is not widely implemented in its exact form, perhaps suggesting that it is impractical or has some hidden disadvantage. We do not believe this to be the case. The only thing that the seller needs to know is the set of potential buyers, and in many cases it would be straightforward for these firms to identify themselves. The seller does need to be able to commit to approaching potential buyers in a particular order, and to develop a system for distributing information about previous bids. For many assets, any costs involved are likely to be small, and for a firm or government agency involved in repeated transactions (e.g., procurement) they would be spread over a large number of contracts. Instead, a more plausible reason for why the exact sequential mechanism considered here is not used is that there are alternative sequential mechanisms that can do even better, consistent with the fact that the seller optimal mechanism is almost certainly some sort of sequential search process and that, within the sequential mechanism, unlike the simultaneous auction, the ability to set a reserve price, and possibly other design elements, can increase seller revenues substantially.

There are, of course, some limitations of the model that we consider here, which may be important in some real-world settings. For example, our IPV assumption will not be satisfied for assets where potential buyers have to form imperfect opinions about some innate future potential. A common value component would change strategies significantly in the sequential mechanism as the incumbent bidder could signal that he believes the common value to be low in order to deter entry. We also assume that firms act competitively, while the structure of the selective mechanism might affect incentives for collusion on either entry decisions or bids. Understanding how these factors would affect the relative performance of sequential and simultaneous mechanisms appear to be profitable directions for future research.

References

Ackerberg, D. (2009): "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," *Quantitative Marketing and Economics*, 7, 343–376.

Aradillas-Lopez, A., A. Gandhi, and D. Quint (2010): "Indentification and Testing in Ascending Auctions with Unobserved Heterogeneity," *Working Paper, University of Wisconsin.*

Athey, S., D. Coey, and J. Levin (2011): "Set-Asides and Subsides in Auctions," *Working Paper.*

Athey, S., P. Cramton, and A. Ingraham (2003): "Upset Pricing in Auction Markets: An Overview," White Paper, Market Design, Inc., on behalf of British Columbia Ministry of Forests.

Athey, S., J. Levin, and E. Seira (forthcoming): "Comparing Open and Sealed Bid Auctions: Theory and Evidence from Timber Auctions," *Quarterly Journal of Economics.*

Bajari, P., H. Hong, and S. Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78(5), 1529–1568.

Baldwin, L., R. Marshall, and J.-F. Richard (1997): "Bidder Collusion at Forest Service Timber Sales," *Journal of Political Economy*, 105(4), 657–699.

Banks, J., and J. Sobel (1987): "Equilibrium Selection in Signalling Games," *Econometrica*, 55(647-662).

Berry, S. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889–917.

Betton, S., and B. Eckbo (2000): "Toeholds, Bid Jumps, and Expected Payoffs in Takeovers," *Review of Financial Studies*, 13, 841–882.

Betton, S., B. Eckbo, and K. Thorburn (2008): "Merger Negotiations and the Toehold Puzzle," *Journal of Financial Economics*, 91, 158–178.

Boone, A. L., and J. H. Mulherin (2007): "How Are Firms Sold?," *Journal of Finance*, 62(2), 847–875.

Bulow, J., and P. Klemperer (2009): "Why Do Sellers (Usually) Prefer Auctions?," *American Economic Review*, 99(4), 1544–1575.

Cho, I., and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2), 179–222.

Cho, I., and J. Sobel (1990): "Strategic Stability and Uniqueness in Signaling Games," *Journal of Economic Theory*, 50(2), 381–413.

Chu, C., P. Leslie, and A. Sorensen (2011): "Bundle-Size Pricing as an Approximation to Mixed Bundling," *American Economic Review*, 101(1), 263–303.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

Cremer, J., Y. Spiegel, and C. Z. Zheng (2009): "Symposium on Auction Theory: Auctions with Costly Information Acquisition.," *Economic Theory*, 38(1), 41–72.

Daniel, K., and D. Hirshleifer (1998): "A Theory of Costly Sequential Bidding," *University of Michigan Business School Working Paper No. 98028.*

Fishman, M. (1988): "A Theory of Preemptive Takeover Bidding," *RAND Journal of Economics*, 19(1), 88–101.

Fudenberg, D., and J. Tirole (1991): *Game Theory.* Cambridge, Mass. and London: MIT.

Gentry, M., and T. Li (2010): "Partial Identification in First-Price Auctions with Selective Entry," *Working Paper.*

Haile, P. (2001): "Auctions with Resale Markets: An Application to U.S. Forest Service Timber Sales," *American Economic Review*, 91(3), 399–427.

Haile, P., and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1), 1–51.

Horner, J., and N. Sahuguet (2007): "Costly Signalling in Auctions," *Review of Economic Studies*, 74, 173–206.

Kilgore, M., R. Brown, J. Coggins, and C. Pfender (2010): "A National Review of State Timber Sale Programs and An Analysis of Factors Influencing Minnesota State Stumpage Prices," Report to the Minnesota Forest Resources Council.

Levin, D., and J. Smith (1994): "Equilibrium in Auctions with Entry," *American Economic Review*, 84, 585–599.

Li, T., and I. Perrigne (2003): "Timber Sale Auctions with Random Reserve Prices," *Review of Economics and Statistics*, 85(1), 189–200.

Li, T., and X. Zheng (2009): "Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions," *Review of Economic Studies*, 76, 1397–1429.

——— (2010): "Information Acquisition and/or Bid Preparation: A Structural Analysis of Entry and Bidding in Timber Sale Auctions," *Working Paper.*

Mailath, G. (1987): "Incentive Compatibility in Signaling Games with a Continuum of Types," *Econometrica*, 55(6), 1349–1365.

Marmer, V., A. Shneyerov, and P. Xu (2010): "What Model for Entry in First-Price Auctions? A Nonparametric Approach," *Working Paper*.

McAfee, R. P. (2002): "Coarse Matching," *Econometrica*, 70(5), 2025–2034.

McAfee, R. P., and J. McMillan (1988): "Search Mechanisms," *Journal of Economic Theory*, 44(1), 99–123.

Mead, W., M. Schniepp, and R. Watson (1981): "The Effectiveness of Competition and Appraisals in the Auction Markets for National Forest Timber in the Pacific Northwest," Washington, DC: U.S. Department of Agriculture, Forest Service.

Milgrom, P. (2004): *Putting Auction Theory to Work*. Cambridge University Press.

Milgrom, P., and J. Roberts (1982): "Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis," *Econometrica*, 50(2), 443–459.

Neeman, Z. (2003): "The Effectiveness of English Auctions," *Games and Economic Behavior*, 43, 214–238.

Paarsch, H. (1997): "Deriving an Estimate of the Optimal Reserve Price: An Application to British Columbia Timber Sales," *Journal of Econometrics*, 78, 333–357.

Ramey, G. (1996): "D1 Signaling Equilibria with Multiple Signals and a Continuum of Types," *Journal of Economic Theory*, 69, 508–531.

Revlon v McAndrews & Forbes Holdings, I. (1986): Del. Supr., 506 A.2d 173.

Roberts, J., and A. Sweeting (2011): "Competition versus Auction Design," *Working Paper*.

Rogerson, W. (2003): "Simple Menus of Contracts in Cost-Based Procurement and Regulation," *American Economic Review*, 93(3), 919–926.

Samuelson, W. (1985): "Competitive Bidding with Entry Costs," *Economic Letters*, 17, 53–57.

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *RAND Journal of Economics*, 37(3), 619–640.

Subramanian, G. (2008): "Go-Shops vs. No-Shops in Private Equity Deals: Evidence and Implications," *The Business Lawyer*, 63(3), 729–760.

Wasserstein, B. (2000): *Big Deal: Mergers and Acquisitions in the Digital Age*. Warner Business Books.

Wilson, R. (1987): *New Palgrave: A Dictionary of Economics* chap. Bidding. MacMillan Press.

Ye, L. (2007): "Indicative Bidding and a Theory of Two Stage Auctions," *Games and Economic Behavior*, 58, 181–207.

A Conditions for Unique Sequential Equilibrium Under the D1 Refinement

We now verify the three conditions for our equilibrium to be the unique sequential equilibrium under the D1 refinement.

(1) $\frac{\partial \pi_v(b_1,S_2')}{\partial S_2'} > 0$. An increase in the signal threshold keeps out more second round potential entrants. The bidding behavior of those entrants who have signals above the threshold is unchanged and so an increase in S_2' must strictly raise the incumbent's probability of winning and lower the expected price paid.

(2) $\frac{\partial \pi_v(b_1,S_2')}{\partial b_1} / \frac{\partial \pi_v(b_1,S_2')}{\partial S_2'}$ *is monotonic in* v. Differentiating $\frac{\partial \pi_v(b_1,S_2')}{\partial b_1} / \frac{\partial \pi_v(b_1,S_2')}{\partial S_2'}$ gives

$$\frac{\partial^2 \pi_v(b_1,S_2')}{\partial b_1 \partial v}\left(\frac{\partial \pi_v(b_1,S_2')}{\partial S_2'}\right)^{-1} - \left(\frac{\partial^2 \pi_v(b_1,S_2')}{\partial S_2' \partial v}\right)\left(\frac{\partial \pi_v(b_1,S_2')}{\partial S_2'}\right)^{-2} \frac{\partial \pi_v(b_1,S_2')}{\partial b_1} \tag{5}$$

Monotonicity requires that this expression is either always positive or always negative. We show that it is always positive by establishing (a)-(d) below.

(a) $\frac{\partial^2 \pi_v(b_1,S_2')}{\partial b_1 \partial v} = 0$. Consider two types of first round bidders v_H and v_L, $v_H > v_L$, with each considering increasing their bid b_1 to $b_1 + \varepsilon$, $\varepsilon > 0$. If the second bidder stays out then the change in profits for each first round type is the same, $-\varepsilon$. We now show that if the second round bidder enters the profit is still the same to each type of first round bidder.

Consider three cases. (i) $v_2 < b_1$. The first round bidder will pay $b_1 + \varepsilon$ whatever his value. (ii) $v_2 > b_1 + \varepsilon$. The final price will equal the value of the lower-valued firm and will not depend on the first round bid.[33] (iii) $b_1 \le v_2 \le b_1 + \varepsilon$. The first round bidder still wins, regardless of type, but now he has to pay more since before he would have won at a price of v_2 but now he wins at a price of $b_1 + \varepsilon$, yielding the same cost of $b_1 + \varepsilon - v_2$ to each type of first round bidder. Therefore, the cost of raising the deterring bid, all else constant, is independent of the first bidder's value.

(b) $\frac{\partial^2 \pi_v(b_1,S_2')}{\partial S_2' \partial v} > 0$. To show that the benefit of increasing the signal entry threshold is

[33] There are three cases within this case. The first is when $v_2 > v_H > v_L$. Regardless of deterring bid, both first round types would lose and so increasing the deterring bid has no effect on their profit. The second is when $v_H > v_2 > v_L$. Here the low type was going to lose regardless, and so it has no effect on his cost. Here the high type was going to win but pay v_2 no matter what and so increasing the bid has no effect on his cost. The third is when $v_H > v_L > v_2$. In either case both types were going to win but have to pay v_2 and so increasing the bid had no effect on either types' costs.

greater the higher is the first bidder's value, we can show that the benefit of excluding any second bidder type v_2 is greater, the higher is the first bidder type, regardless of v_2. Consider the value of excluding a second round bidder whose value is v_2 for any two types of first round bidders v_H and v_L, $v_H > v_L$, both using deterring bid b_1. If $v_2 \leq b_1$ there is no change in benefit from exclusion for either first bidder type. If $b_1 < v_2$ there are three cases. (i) $v_2 \leq v_L < v_H$. In this case the benefit of excluding the second round bidder is $v_2 - b_1$ for each first round bidder type. (ii) $v_L < v_2 \leq v_H$. In this case the benefit of exclusion is $v_L - b_1$ for the low type and $v_2 - b_1$ for the high type. Since by assumption $v_2 > v_L$, the benefit of exclusion is greater for the higher type. (iii) $v_L < v_H < v_2$. In this case the benefit of exclusion is $v_L - b_1$ for the low type and $v_H - b_1$ for the high type and so the benefit is greater for the higher first bidder type. Therefore, the benefit of excluding more second round bidders is greater the higher is the first round bidder's value.

(c) $\frac{\pi_v(b_1,S_2')}{\partial S_2'} > 0$. This was shown above when we verified condition (1).

(d) $\frac{\partial \pi_v(b_1,S_2')}{\partial b_1} < 0$. Increasing the bid is costly when it does not affect the second round potential entrant's decision. In particular, it reduces a firm's payoff when the second round firm does not enter or it enters and has a value less than b_1. If the potential entrant enters with a value above b_1 then changing b_1 has no effect.

Combining (a)-(d), we conclude that, for all v, b_1 and S_2':

$$\underbrace{\frac{\partial^2 \pi_v(b_1, S_2')}{\partial b_1 \partial v}\left(\frac{\partial \pi_v(b_1, S_2')}{\partial S_2'}\right)^{-1}}_{=0} - \underbrace{\left(\frac{\partial^2 \pi_v(b_1, S_2')}{\partial S_2' \partial v}\right)\left(\frac{\partial \pi_v(b_1, S_2')}{\partial S_2'}\right)^{-2}\frac{\partial \pi_v(b_1, S_2')}{\partial b_1}}_{<0} > 0 \qquad (6)$$

and so the monotonicity condition is satisfied.

(3) S_2' *is uniquely defined for any belief about the first potential entrant's value, and the potential entrant's response is more favorable to the incumbent when the potential entrant thinks that the incumbent's value is higher.* This is true since S_2' is a continuous function of the second period potential entrant's belief about the incumbent's value (reflecting the zero profit condition, as in equation 4, and the potential entrant's beliefs about his own value as a function of its signal) and the second potential entrant will increase S_2' if he believes bidder 1's type is higher because his expected profits are decreasing in bidder 1's type for any signal he receives.

B Details of Estimation Method

In this appendix we more fully describe our estimation procedure based on Ackerberg (2009)'s method of simulated maximum likelihood with importance sampling.

This method involves solving a large number of games with different parameters once, calculating the likelihoods of the observed data for each of these games, and then re-weighting these likelihoods during the estimation of the distributions for the structural parameters. This method is attractive when it is believed that the parameters of the model are heterogeneous across auctions and it would be computationally prohibitive to re-solve the model (possibly many times in order to integrate out over the heterogeneity) each time one of the parameters changes.[34]

To apply the method, we assume that the parameters are distributed across auctions according to the specification given in Section 4.3. These specifications reflect our assumptions that σ_V, α and K are the same for mills and loggers within any particular auction, even though they may differ across auctions. The lower bound on σ_{Va} is set slightly above zero simply to avoid computational problems that were sometimes encountered when there was almost no dispersion of values. Our estimated specifications also assume that the various parameters are distributed independently across auctions. This assumption could be relaxed, although introducing a full covariance matrix would significantly increase the number of parameters to be estimated and, when we have tried to estimate these parameters, we have not found these coefficients to be consistently significant across specifications. The set of parameters to be estimated are $\Gamma = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \omega^2_{\mu,\text{logger}}, \omega^2_{\mu,\text{diff}}, \omega^2_{\sigma_V}, \omega^2_{\alpha}, \omega^2_K\}$, and a particular draw of the parameters $\{\mu_{a,\text{logger}}, \mu_{a,\text{mill}}, \sigma_{Va}, \alpha_a, K_a\}$ is denoted θ.

Denoting the outcome for an observed auction by y_a, the log-likelihood function for a sample of A auctions is

$$\sum_{a=1}^{A} \log \left(\int L_a(y_a|\theta)\phi(\theta|X_a, \Gamma)d\theta \right) \tag{7}$$

where $L_a(y_a|\theta)$ is the likelihood of the outcome y in auction a given structural parameters θ, $\phi(\theta|X_a, \Gamma)$ is the pdf of the parameter draw θ given Γ, our distributional assumptions, the unique equilibrium strategies implied by our equilibrium concept and auction characteristics including the number of potential entrants, the reserve price and observed characteristics X_a.

Unfortunately, the integral in (7) is multi-dimensional and cannot be calculated exactly.

[34]Bajari, Hong, and Ryan (2010) use a related method to analyze entry into a complete information entry game with no selection.

We follow Ackerberg by recognizing that

$$\int L_a(y_a|\theta)\phi(\theta|X_a,\Gamma)d\theta = \int L_a(y_a|\theta)\frac{\phi(\theta|X_a,\Gamma)}{g(\theta|X_a)}g(\theta|X_a)d\theta \tag{8}$$

where $g(\theta|X_a)$ is the importance sampling density whose support does not depend on Γ, which is true in our case because the truncation points are not functions of the parameters. This can be simulated using

$$\frac{1}{S}\sum_s L_a(y_a|\theta_s)\frac{\phi(\theta_s|X_a,\Gamma)}{g(\theta_s|X_a)} \tag{9}$$

where θ_s is one of S draws from $g(\theta|X_a)$. Critically, this means that we can calculate $L_a(y_a|\theta_s)$ for a given set of S draws that do not vary during estimation, and simply change the weights $\frac{\phi(\theta_s|X_a,\Gamma)}{g(\theta_s|X_a)}$, which only involves calculating a pdf when we change the value of Γ rather than re-solving the game.

This simulation estimator will only be accurate if a large number of θ_s draws are in the range where $\phi(\theta_s|X_a,\Gamma)$ is relatively high, and, as is well known, simulated maximum likelihood estimators are only consistent when the number of simulations grows fast enough relative to the sample size. We therefore proceed in two stages. First, we estimate Γ using $S = 2,500$ draws, where $g(\cdot)$ is a multivariate uniform distribution over a large range of parameters which includes all of the parameter values that are plausible. Second, we use these estimates $\widehat{\Gamma}$ to repeat the estimation using a new importance sampling density $g(\theta|X_a) = \phi(\theta_s|X_a,\widehat{\Gamma})$ with $S = 500$ per auction. Roberts and Sweeting (2011) provide Monte Carlo evidence that the estimation procedure works well even for smaller values of S.

To apply the estimator, we also need to define the likelihood function $L_a(y_a|\theta)$ based on the data we observe about the auction's outcome, which includes the number of potential entrants of each type, the winning bidder and the highest bids announced during the open outcry auction by the set of firms that indicated that they were willing to meet the reserve price. Two problems arise when interpreting these data. First, a bidder's highest announced bid in an open outcry auction may be below its value, and it is not obvious which mechanism leads to the bids that are announced (Haile and Tamer (2003)). Second, if a firm does not know its value when taking the entry decision, it may learn (after paying the entry cost) that its value is less than the reserve price and so not submit a bid.

We therefore make the following assumptions (Roberts and Sweeting (2011) present estimates based on alternative assumptions about the data generating process that deliver similar results) that are intended to be conservative interpretations of the information that is in the data: (i) the second highest observed bid (assuming one is observed above the reserve price) is equal to the value of the second-highest bidder; [35] (ii) the winning bidder has a value greater than the second highest bid; (iii) both the winner and the second highest bidder entered and paid K_a; (iv) other firms that indicated that they would meet the reserve price or announced bids entered and paid K_a and had values between the reserve price and the second highest bid; and, (v) all other potential entrants may have entered (paid K_a) and found out that they had values less than the reserve, or they did not enter (did not pay K_a). If a firm wins at the reserve price we assume that the winner's value is above the reserve price.

[35] Alternative assumptions could be made. For example, we might assume that the second highest bidder has a value equal to the winning bid, or that the second highest bidder's value is some explicit function of his bid and the winning bid. In practice, 96% of second highest bids are within 1% of the high bid, so that any of these alternative assumptions give similar results. We have computed some estimates using the winning bid as the second highest value and the coefficient estimates are indeed similar.

Effects of Product Availability: Experimental Evidence [*]

Christopher T. Conlon[†]
Julie Holland Mortimer[‡]

July 18, 2011

Abstract

Product availability impacts many markets in which stochastic demand and fixed inventories may lead to stock-outs. In vertically-separated markets, optimal inventory decisions may differ substantially between upstream and downstream firms, making vertical arrangements necessary for coordination. We examine the impact of exogenously removing products from vending machines, and analyze subsequent changes in purchasing patterns and profits using both nonparametric analyses and structural demand estimation. We find substantial switching to alternate products, and evidence of misaligned stocking incentives between upstream and downstream firms. We discuss the trade-offs of both empirical approaches for analyzing product availability effects generally.

[*]We thank Mark Stein, Bill Hannon, and the drivers at Mark Vend Company for implementing the experiments used in this paper, providing data, and generally educating us about the vending industry. We thank Dan Ackerberg, Kate Ho, Greg Lewis, Richard Mortimer, and seminar participants at Harvard University and UCLA for helpful comments. Tom Gole, Adam Kapor and Sharon Traiberman provided exceptional research assistance. Financial support for this research was generously provided through NSF grant SES-0617896. Any remaining errors are our own.

[†]Department of Economics, Columbia University, 420 W. 118th St., New York City, NY 10027. email: cc3264@columbia.edu

[‡]Department of Economics, Harvard University, 1805 Cambridge St., Cambridge, MA, 02138, and NBER. email: mortimer@fas.harvard.edu

1 Introduction

Product availability impacts many markets, particularly those for which storage costs or capacity constraints matter. For example, inventory decisions may lead to stock-outs in retail markets, and capacity decisions may affect transportation, performance events, health care provision, and school choice. Firms in these markets may optimize product availability to influence consumer decisions about where and when to shop. In vertically-separated markets, optimal stocking choices for downstream firms may differ substantially from those of upstream manufacturers. In such settings, manufacturers may use vertical arrangements to try to align the stocking decisions of the downstream firms with their own interests.

Despite the key role that product availability plays in many markets, little empirical evidence exists on the importance of product availability for firms or consumers. We investigate the impact of product availability through a field experiment in which we exogenously remove top-selling products from a set of vending machines, and track subsequent consumer responses and profit impacts. Product availability in the context of vending has been the subject of many recent debates about obesity and the appropriate public policy response to the mix of products offered in vending machines, particularly in school settings.[1]

We find that most consumers purchase another good when a top-selling product is removed. The profit impacts indicate that stocking incentives may be misaligned between downstream and upstream firms in the absence of vertical arrangements. For example, some product removals result in lower revenues for the upstream manufacturer but higher profits for the downstream firm as consumers substitute to products with higher downstream margins. This effect may provide a rationale for the use of tying or forcing contracts by upstream firms that is unrelated to competitive considerations at the upstream level.

The experimental nature of the data allows us to analyze the substitution patterns and profit impacts of each product removal using two alternative methodologies. First, we report nonparametric analyses of the data using techniques common to the treatment-effects literature, and applicable here due to the exogenous variation in product availability introduced by the experiment. Second, we analyze the data using structural demand estimation. The implementation of the two estimation methods allows us to generate insights into the relative benefits and drawbacks of applying these alternative methodological approaches to other data sources or in other settings.

We run our field experiment on a group of 60 vending machines located in five office buildings in downtown Chicago. Over the course of a three-year period, we implement six single-product removals (in which a single product is removed from all sites), and two double-

[1] Forty states now tax junk food or soda products, and cities, school districts, and other local jurisdictions have proposed or implemented restrictions on the set of products that may be offered in vending machines. See Engber (2009) for a recent press article summarizing many policy responses in this area. More recent examples include rules requiring that the mix of beverages in city vending machines favor water in New York City, a ban on sales of sugary drinks in city buildings in San Francisco, and a similar proposed ban in Boston (Smith 2010). The medical literature has also weighed in on the issue of taxing sugary drinks (e.g., see Brownell and Frieden (2009) and Brownell, Farley, Willett, Popkin, Chaloupka, Thompson, and Ludwig (2009)).

product removals (in which two products are removed simultaneously). Each removal lasts for 2-3 weeks. For the nonparametric analyses, we choose control weeks from the three-year panel using nearest-neighbor matching methods and report the substitution patterns implied by a comparison of the treatment and control weeks. Using the same control weeks, we also report the profit impacts of each experiment for the downstream firm and the revenue impacts for all upstream firms.

The nonparametric approach allows us to document changes in purchasing patterns and profit impacts directly, with no need to make any parametric assumptions about the functional form of demand. The changes in purchase frequencies across remaining products are quite sensible; most consumers switch to a similar product when their first choice is removed, and relatively few consumers walk away. In some cases, the downstream firm appears to profit from a product removal, because consumers switch to products with higher downstream profit margins. However, accounting for the vertical contracts between the vending operator and upstream manufacturers reverses this effect, indicating that one rationale for the vertical arrangements in this industry may be to align the downstream firm's incentive to carry and service the products of upstream manufacturers.

The nonparametric approach requires no functional form assumptions, but this flexibility can produce noisy measures of the true effect of each product removal. For example, total vends are occasionally higher during the treatment period (when a product is removed) than during the control period, and vends of substitute goods are sometimes lower during the treatment period than in the control. We would not want to assign a causal interpretation to these types of outcomes, and thus, they represent a limitation to the otherwise very informative results of this approach. More broadly, these sorts of problems may be endemic to any large-scale field experiment in industrial organization (IO). Unlike studies of individual behavior in the lab or the field, randomization may not be able to control for all possible market-level variables in most IO contexts. For example, we can't prevent Mars, Inc. from advertising Snickers in different ways over time or in different geographic markets.

The second way in which we analyze the substitution patterns and profit impacts of changes in product availability is through the use of structural demand estimation. Using the full dataset, we estimate nested logit and random-coefficients logit models of demand. We predict vends during the treatment and control periods using the estimated model parameters, and compare these outcomes. The models perform well in many respects, and capture much of the variation that is observed in the nonparametric results. However, they tend to predict fewer sales to the products identified as 'top substitutes' in the nonparametric analysis, and predict more sales to other products. We speculate on two reasons for this result; namely, properties of the logit error term, and the endogeneity of changes in the retailer's product mix. We also provide estimates of downstream profit and upstream revenue, and again find evidence of mismatched incentives in the vertical supply chain that are addressed by the vertical contracts used between this vending operator and the upstream manufacturers.

Next, we extend the structural analyses by simulating the effect of the product removal directly. This differs from the first exercise, in which we compare estimates of the model during the treatment and control periods, because that exercise allows for other changes to occur in addition to the stock-out of the focal product. For example, if a different brand of pretzels was carried during a control week, the first exercise adjusts for that. In this sense, neither the nonparametric analyses nor the predicted structural estimates of vends during the treatment and control periods provide a clean measure of the impact of the product removal *per se*. Simulating the product removal, on the other hand, holds all other factors fixed, and isolates the "pure" treatment effect when the model is correctly specified. The results of simulating the product removal are qualitatively similar to the 'prediction' exercise. The profit(revenue) impacts for downstream(upstream) firms are quite similar to the impacts we estimate using the nonparametric approach, with the same instances of mismatched incentives in the vertical chain.

In most cases in which one estimates demand, the type of exogenous variation in product availability that we create through our experiments does not exist in the data. Thus, structural demand models are often identified solely from naturally-occurring variation in choice sets (which may also include variation in product characteristics, such as price). One way to examine how successful these sources of variation are for identifying model parameters is to remove the variation in choice sets that arises from our experimental interventions. Thus, we conduct a series of hold-out analyses, in which we estimate the nested logit and random-coefficients logit models using subsets of the data that drop treatment periods.[2] We find that the estimated model parameters are generally robust to the exclusion of data from the experimental interventions, with the exception of the parameters that govern correlation in consumer tastes for candy and sugar. These parameters are sensitive to the exclusion of data from the candy bar stock outs. The finding is intuitive because there are fewer "naturally occurring" changes over time in the product mix of the candy category, and the candy bars that we exogenously remove have much higher market shares than the salty snack and cookie products that we exogenously remove.

There are several advantages to studying product availability in the context of vending machines. One advantage is the ability to successfully implement the field experiments logistically (ie., to exogenously change the set of products that are available to consumers). Even a "simple" retail setting would introduce complications to this change that are absent with vending machines.[3] Second, the scale of revenues that are potentially at stake in a vending machine is relatively small, so the experiments are not prohibitively expensive to run. Third, we observe the wholesale costs of the vending operator, which makes calculation of the upstream and downstream revenue/profit impacts possible.

Some features of vending machines are both advantages and disadvantages for the purpose of our study. Competition between retail outlets is not a feature of our setting. While this simplifies many aspects of the study, it is also a limitation, in the sense that we cannot

[2]One could describe the baseline estimates from our full dataset as an "in-sample" prediction exercise for the structural models. The hold-out analyses provide an "out-of-sample" prediction, in which the model is asked to predict the results of experiments that have not already occurred in the data.

[3]For example, one would need to remove the focal product not only from shelves, but also from back-room storage areas, or alternatively prevent clerks from responding to special consumer requests to retrieve an item from storage when it is not on the shelf.

study how competition among downstream firms responds to changes in product availability. Similarly, price variation is quite limited, especially within a site for a category (e.g., candy) over a period of time.

Finally, we note that the field experiment is not a randomized trial. Rather, each removal exogenously varies the set of products available to all consumers for a period of time, and control weeks are selected from an observational dataset during which no intervention occurred.[4] We did not have the ability to randomize the set of products offered to any given consumer at the time of her purchase. Such an experiment would represent a true randomized trial (absent other market-level effects), but is not feasible technologically in brick-and-morter retail contexts.

Relationship to Literature

This paper connects several different literatures. The first is a growing literature in economics, marketing, and operations research that focuses on firms' stocking decisions and the importance of product availability for vertical arrangements. The "newsvendor" problem dates back to Edgeworth (1888), Spengler (1950), and Arrow, Harris, and Marschak (1951), and describes the potential for mis-aligned inventory incentives between upstream and downstream firms. More recent theoretical work formalizes and extends the solution to this problem (e.g., Kraiselburd, Narayanan, and Raman (2004), Schweitzer and Cachon (2000), and many others). In more recent empirical work, Anupindi, Dada, and Gupta (1998) study product availability, also in the context of vending machines. Several examples in this literature focus on scanner data and availability at supermarkets and convenience stores, such as Bruno and Vilcassim (2008), and Musalem, Olivares, Bradlow, Terwiesch, and Corsten (2010), and Matsa (2010). Aguirregabiria (1999) uses scanner data to examine the strategic implications of dynamic inventory decisions in the context of vertically-separated markets.

The second is a wider literature on field experiments in economics. A rather extensive review of this literature is presented in Levitt and List (2009). Some recent and notable examples include Karlan and List (2007) who study the impact of price on charitable giving, and Bertrand, Karlan, Mullainathan, Shafir, and Zinman (2010) who examine the impact of advertising using a direct-mail experiment involving consumer lending. Much of the field experiments literature focuses on direct-mail, charitable giving, or auction settings (such as Engelbrecht-Wiggans, List, and Reiley (2005) or Ostrovsky and Schwarz (2009)). In a retail setting Cai, Fang, and Yuyu (2009) examine observational learning by randomly marking menu items as "favorites" and analyzing the impact of the designation on customer demand. There is a small related literature that uses experiments to study the effects of stockouts. Fitzsimons (2000) studies psychological effects of stockouts on consumers in the laboratory,

[4]From the point of view of experimental economics, our intervention is more closely related to laboratory experiments that attempt to understand markets as a whole (e.g. Smith (1962) and more recently, Eriv and Roth (1998), among others), than to experiments that examine individual responses (Tversky and Kahneman (1991) summarize many examples). Despite the fact that such interventions fall short of a fully randomized trial, they are in fact exactly the type of experimentation that one might expect retailers to do when trying to learn about demand or set prices in an optimal way. Online retailers may differ from brick-and-morter stores in this respect, but the nature of competition and consumer search may also differ in online settings.

and Anderson, Fitzsimons, and Simester (2006) examine psychological framing effects of how stockouts are presented to consumers in the context of a mail-order company.

The paper also contributes to a recent discussion about the role of different methods in empirical work going back to Leamer (1983), and discussed recently by Heckman (2010), Angrist and Pischke (2010), Leamer (2010), Keane (2010), Sims (2010), Nevo and Whinston (2010), Stock (2010), and Einav and Levin (2010). A central issue in this debate is what role experimental or quasi-experimental methods should play in empirical economic analyses in contrast to structural methods. Several of these recent papers essentially argue that both types of approaches have advantages and drawbacks. Our setting provides the opportunity to examine empirically the trade-offs to which these papers refer. For example, while our experimental estimates are quite informative in many respects, there are cases in which we would not want to infer causality (e.g., when overall sales increase during a stock-out event). The structural demand models use economic theory to rule out such an effect, but cannot fully capture the degree of substitution that occurs from a focal product to other goods. This is especially true when a product's most important characteristics are less easy to measure (e.g., packaging differences, or possibly unobserved advertising campaigns).

The paper proceeds as follows. We describe the vertical arrangements used in the vending industry, and the design of our field experiments and data in section 2. Section 3 describes the non-parametric results from the field experiments including the implications of the experimental results for firm profitability. In section 4, we describe two structural demand models commonly used to estimate substitution patterns (i.e., nested-logit and random-coefficients logit models), and a method for estimating the rate of consumer arrivals at each machine. In section 5 we compare the predicted substitution patterns of those models during the treatment and control weeks analyzed in section 3; section 6 uses the demand models to simulate the removal of focal products (holding all other conditions of the market fixed). Section 7 provides the results of hold-out analyses, in which we estimate the demand model on subsets of the data, and section 8 concludes.

2 The Vending Industry and Experimental Data

2.1 Vertical Arrangements in the Vending Industry

Vertical arrangements are widely used in the vending industry and apply to several of the upstream relationships of the firm with whom we worked. The most commonly used vertical arrangement in the industry is referred to as a "rebate program." Under a rebate program, a manufacturer refunds a portion of a vending operator's wholesale cost at the end of a fiscal year if the vending operator meets an annual sales goal, typically expressed as a percentage of last year's sales. The sales goal for an operator is typically set for the combined sales of a manufacturer's products, rather than for individual products. Some manufacturers also require a minimum number of product "facings" in an operator's machines. The amount of the rebate and the precise threshold of the sales goal or facing requirement is specific to an individual vending operator, and these terms are closely guarded by participants in the industry.

While the rebate programs share features with vertical tying (particularly when facing requirements apply), the primary function of the rebate programs is to lower the wholesale cost through an operator-specific mechanism.[5] The benefit of rebate programs for manufacturers is the ability to more closely align the downstream operator's incentive to carry and re-stock the manufacturer's products with the manufacturer's own incentives. Specifically, at any wholesale cost greater than the cost of production, the downstream firm chooses to stock fewer units of inventory than the upstream manufacturer would choose. This inventory stocking problem is well understood, and is referred to as the "newsvendor" problem in the case when prices are fixed and demand is stochastic. The intuition is formalized as follows.

Consider a single product with stochastic demand denoted by D, with distribution function F. The downstream vendor purchases q units of inventory at cost c and sells at a fixed price $p > c$. Demand lasts for one period, and inventory is purchased at the beginning of the period. For the sake of illustration, assume no salvage value for the downstream firm and a production cost for the upstream firm of zero.[6] Realized profit is:

$$\pi(q, D) = p \min(q, D) - cq \tag{1}$$

and expected profit is:

$$E[\pi(q, D)] = (1 - F(q))\pi(q, q) + \int_0^q f(x)\pi(q, x)dx. \tag{2}$$

The first term captures profits when the firm is understocked (i.e., no incremental profit is earned from consumers who arrive after the product is sold out), and the second term captures profits when the firm is overstocked. The solution to the newsvendor problem maximizes expected profit at quantity:

$$q^* = F^{-1}\left(\frac{p - c}{p}\right) \tag{3}$$

Thus, q^* is chosen so that the expected marginal return equals the marginal cost to the downstream firm of an additional unit of inventory.[7] Note that the newsvendor problem is analogous to a "fixed price/stochastic demand" version of the double-marginalization problem, in that the downstream vendor only accounts for his own mark-up $(p - c)$ rather than the full difference between p and production cost (assumed here to be zero) when stocking inventory.

Rebates lower the wholesale cost for downstream operators, leaving them with a higher expected return from stocking an additional unit of inventory. By structuring this as a

[5] Robinson-Patman prevents manufacturers from directly price discriminating across competing downstream firms when selling "inputs."

[6] This is without loss of generality for a production cost and/or salvage value for the downstream firm less than c.

[7] Alternatively, this ratio (referred to as the "critical fractile") may be viewed as relating the cost of understocking (a lost sale is worth $(p-c)$) to the total cost of being either understocked or overstocked (i.e., the cost of understocking $(p-c)$ plus the cost of overstocking (c)).

rebate, rather than directly reducing wholesale price, manufacturers are able to tailor the amount of the cost reduction to each individual operator, and to match it to targets that are retailer specific (e.g., 90 percent of his previous year's sales).

2.2 Experimental Design

We ran eight experimental treatments with the help of Mark Vend Company, which is a medium-sized independent vending operator in the Chicago area. We identified 60 snack machines located in office buildings, for which demand was historically quite stable.[8] Most of the customers at these sites are 'white-collar' employees of law firms and insurance companies. Our goal in selecting the machines was to choose machines that could be analyzed together, in order to be able to run each experiment over a shorter period of time across more machines.[9] We selected snack machines because beverage machines have extremely large capacities and a small number of products. This made the logistics of stocking out beverages more difficult (removing and storing 100 large heavy bottles vs. 20 candy bars), and also made the outcomes less interesting, because the demand system only includes around six products. Finally, we selected machines on routes that were staffed by experienced drivers, so that the implementation of the experiments would be successful. The 60 machines used for each experiment were distributed across five of Mark Vend's clients, which had between 3 and 21 machines each. The largest client had two sets of floors serviced on different days, and we divided this client into two sites. Generally, each site is spread across multiple floors in a single high-rise office building, with machines located on each floor.

Implementation of each product removal was fairly straightforward; we removed either one or two top-selling products from all machines for a period of roughly 2.5 to 3 weeks. Six of the experiments stocked-out a single top-selling product: Snickers, Peanut M&Ms, Zoo Animal Crackers, Famous Amos Chocolate Chip cookies, Doritos, or Cheetos. Two of the experiments removed two products simultaneously: Snickers plus Peanut M&Ms, or Doritos plus Cheetos. Whenever a product was experimentally stocked-out, poster-card announcements were placed at the front of the empty product column. The announcements read "This product is temporarily unavailable. We apologize for any inconvenience." The purpose of the card was two-fold: first, we wanted to avoid dynamic effects on sales as much as possible, and second, the firm wanted to minimize the number of phone calls received in response to the stock-out events.

The dates of the interventions range from June 2007 to September 2008, with all removals run during the months of May - October. We collected data for all machines for just over three years, from January of 2006 until February of 2009. During each 2-3 week experimental period, most machines receive service visits about three times. However, the length of service visits varies across machines, with some machines visited more frequently than others.

[8]More precisely, demand at these sites is "relatively" stable compared to the population of sites serviced by the vending operator.

[9]Many high-volume machines are located in public areas (e.g., museums or hospitals), and have demand that varies enormously from one day to the next, so we did not use machines of this nature. In contrast, the work-force populations at our experimental sites are relatively homogenous.

The cost of the experiment consisted primarily of driver costs. Drivers had to spend extra time removing and reintroducing products to machines, and the driver dispatcher had to spend time instructing the drivers, tracking the dates of each experiment, and reviewing the data as they were collected. Drivers are generally paid a small commission on the sales on their routes, so if sales levels fell dramatically as a result of the experiments, their commissions could be affected. Tracking commissions and extra minutes on each route for each driver would have been prohibitively expensive to do, and so drivers were provided with $25 gift cards for gasoline during each week in which a product was removed on their route to compensate them for the extra time and the potential for lower commissions. With the exception of an individual site on each of two experimental runs, implementation was successful.[10]

We faced a few limitations when designing the experiment. For example, some removals were scheduled "back-to-back." In these cases, we selected products that seemed ex-ante less likely to be close substitutes for adjacent runs. For example, the Doritos stock-out was followed by the Peanut M&Ms stockout. Due to more difficult logistics associated with experimental price changes, we were not able to implement any pricing experiments.[11] Finally, throughout our analyses, we focus on static effects. We do not see much evidence of dynamic effects in the data, but this is not something for which we are able to test directly. We note that demand for a focal product tends to remain fairly stable (and demand for other products returns to previous levels) after it is replaced.

2.3 Data Description

Data on the number and price of all products vended are recorded internally at each vending machine used in our experiments. The data track vends and revenues since the last service visit (but do not include time-stamps for each sale). Any given machine can carry roughly 35 products at one time, depending on configuration. We observe prices and variable costs for each product at each service visit during our 38-month panel. There is relatively little price variation within a site, and almost no price variation within a category (e.g., candy) at a site. Very few "natural" stock-outs occur at our set of machines.[12] Over all sites and months, we observe 162 unique manufacturer products. We organize these products into 417 site-product pairs (approximately 70 unique manufacturer products per site) by consolidating low-selling products over time within each site.[13] This set of 417 site-product pairs is our base dataset for all analyses and estimation.

In addition to the data from Mark Vend, we also collect data on the characteristics of

[10]In the two unsuccessful runs, the driver at one site forgot to remove the focal product, so no intervention took place.

[11]The firm does change some prices at some sites late in the dataset; we do not analyze this variation.

[12]Mark Vend commits to a low level of stock-out events in its service contracts.

[13]For example, we combine Milky Way Midnight with Milky Way. In a small number of cases, the consolidated products vary slightly in their wholesale costs or combine products of different manufacturers. For these products, we use the modal wholesale cost, and we apportion revenues paid by the vending operator to manufacturers on the basis of the vends of each individual product.

each product online and through industry trade sources.[14] For each product, we note its manufacturer, as well as the following set of product characteristics: package size, number of servings, and nutritional information.[15] Summary statistics at the manufacturer level are reported in the Appendix. One variable that the data do not measure is the number of people who walk away from a machine. We considered the possibility of adding video cameras or pressure mats to the machines, but neither of these options would have provided clean information on market size.[16] We discuss the issue of market size in detail when we describe the structural models of demand.

3 Nonparametric Analyses of the Experimental Outcomes

3.1 The Matching Estimator

In order to calculate changes in purchasing patterns, sales during treatment weeks are compared with sales during control weeks. We measure substitution from product k to product j as:

$$\Delta q_j = E[q_j|A_{J\setminus k}] - E[q_j|A_J]$$

where q_j denotes weekly sales, J is the full set of products, and A_J denotes availability of all products in J.

In principle, this calculation is straightforward. In practice, however, there are three challenges in implementing the experiments and interpreting the data generated by them. First, service visits vary in length across machines and over time. Second, overall sales levels vary over time, due to exogenous changes in the rate of consumer arrivals. For example, a law firm may have a large case going to trial in a given month, and vend levels will increase at the firm during that period. Third, the product mix presented in a machine is not necessarily fixed across machines, or within a machine over long periods of time (e.g., as manufacturers change their product lines, or Mark Vend changes stocking decisions). Variation in the product mix across sites and machines increases the number of outcomes that the experiment attempts to measure (consider that we start with 162 unique products, roughly 70 of which are carried at any particular site).[17] Changes in the product mix that occur over time for a given machine affect the comparability of the observational control weeks to the weeks in which treatment occurs.

[14]For consolidated products, we collect data on product characteristics at the disaggregated level. The characteristics of the consolidated product are computed as the weighted average of the characteristics of the component products, using vends to weight. In many cases, the observable characteristics are identical.

[15]Nutritional information includes weight, calories, fat calories, sodium, fiber, sugars, protein, carbohydrates, and cholesterol.

[16]Pressure mats were not workable because potential customers can see the product facings without standing close enough to be registered on the mat. Video cameras would have introduced issues of human subjects approval into the experiments, and also suffer from the problem of consumers being able to see the product facings in some machines without standing close enough to show up on the video.

[17]Recall that a single machine may be able to stock roughly 35 different products at one time, depending on configuration.

We take three steps to address these complicating factors. First, we consolidate data collected at service visits to weekly observations. This allows us to make direct comparisons across machines that are visited at different frequencies.[18] Second, we create classes of products for reporting the results of the experiments. This reduces the number of outcomes that each experiment attempts to measure. Third, we select a set of matched control weeks using nearest-neighbor matching techniques. This adjusts for the fact that the treatment weeks may belong to periods of low or high demand, due to exogenous variation in market size over time. We describe each of these steps in turn.

The first step for analyzing the experimental outcomes is the assignment of sales data to weekly units. This is done by apportioning "total vends since the previous visit" evenly across the elapsed days, and gathering groups of seven days into weeks.[19] Moving to weekly visits has a number of advantages. It smooths out variation in sales levels that occurs over very short (daily) intervals, and makes data from different service visits comparable.[20] Experiments are implemented at service visits, and all data from the experiments are also apportioned to weeks. This introduces the possibility that an experiment may be "contaminated" at the weekly level because implementation of the stock-out occurred in the middle of a week. In order to minimize this contamination, we choose start-dates for weeks at each site based on the timing of the experimental stock-out service visit.[21] This eliminates contamination, except for a very small number of cases in which different machines within a site are visited on different days. In such cases, we eliminate any treatment weeks for which more than four vends of the focal product were recorded at the site.

The second step we take for the analysis is to create product "classes" by combining the roughly 70 products at each site into groupings for which changes in purchase frequencies can be reliably measured. This is necessary because smaller products are not bought in sufficient quantity to identify changes for each one individually.[22] We allow the product classes to vary by experiment: all experiments include the six focal products, four additional major products (Twix Caramel, Salted Peanuts, Raisinets, and Skittles), and seven "assorted" classes to capture the smallest products. We choose several additional individual products to track for each product removal based on the purchase patterns observed in the data.[23]

[18]When we estimate models of demand in later sections, we preserve the visit-level detail for estimation, but report predictions and simulations of the model at the weekly level.

[19]On average across the entire dataset, service visits are approximately 2 days long. However, the time between service visits ranges from one to 21 days. The length of time between visits varies over time for a given machine, and also across machines (and sites).

[20]Daily vends would also provide comparable results, but with a significant amount of noise in overall levels of vends.

[21]For example, if Snickers bars were removed on a Wednesday at site 93, we define weeks as periods of Wednesday - Tuesday at that site. We set the start-date of weeks separately for each site and each experiment, and require that each week start on a workday (Monday - Friday).

[22]Fig Newton and Swiss Creme cookies are examples of these types of products. The stocking of these products is sometimes governed by the client (e.g., the CEO may request that Fig Newtons be available).

[23]The assorted classes are: Chocolate, Candy, Energy, Cookie, Potato Chips, Pretzels, and Salty Snacks. As an example of how the additional individual products vary across experiments, consider the Snickers and Cheetos removals. When stocking out Snickers, the additional individual products are Ruger Vanilla Wafer and Grandma's Chocolate Chip Cookie. For the Cheetos stock-out, the additional individual products are

The third step is to select matched control weeks from the three-year panel of observational data. Before selecting control weeks, we sum vends across the machines at a given site.[24] Levels of demand at a site that change over time affect our ability to compare sales during treatment and control periods, and the selection of matched control weeks focuses on choosing weeks in which the level of demand is similar to that during the relevant treatment period. In our particular setting, many of the experiments were run during the summer of 2007, which was a high-point in demand at these sites, most likely due to macroeconomic conditions. In order to select weeks of similar demand levels, we identify a set of product classes that we believe are ex-ante unlikely to be substitutes to the focal product, and we use nearest-neighbor matching methods, matching on the site-level sales of these "non-substitute" product classes.

The use of matching on non-substitute product classes may be motivated in the following way. Substitution from product k to product j in market t is described as the change in the probability of purchasing j when k is not available. Excluding cases in which products are complementary in consumption, this implies that

$$p_{jt}(A_{J\setminus k}) \geq p_{jt}(A_J).$$

In the data, we observe sales rather than choice probabilities, given by:

$$q_{jt} = M_t p_{jt}(A_t),$$

where A_t denotes the set of available products in market t, and M_t denotes market size.[25] The challenge for identifying substitution is that M_t is unobserved. The matching procedure attempts to control for changes in M_t by matching on sales levels of non-substitute products. For non-substitute product l,

$$p_{lt}(A_{J\setminus k}) \approx p_{lt}(A_J).$$

Thus,

$$\frac{q_{lt'}(A_{t\setminus k})}{q_{lt}(A_t)} \approx \frac{M_{t'}}{M_t},$$

so that by matching on sales levels of the non-substitute goods, we try to obtain a ratio of $M_{t'}/M_t$ that is close to one.[26]

Sun Chips, Frito, Farley's Mixed Fruit, Cheez-It Original, and Hot Stuff Jays.

[24]There is nothing to be gained in the nonparametric analysis from variation across different machines at the same office building, given that our experiment removes each focal product from all machines at the site at the same time.

[25]A market in our context generally has both geographic and temporal characteristics (ie., the week of March 4th at site 93).

[26]We investigated a number of alternative approaches. For example, we analyzed a "difference-in-difference" type estimator in which, instead of matching the treatment week to a set of control weeks, one matches the week before an experiment to potential control weeks on the basis of focal-product sales and calculates changes in vends over time. This avoids the problem of having to choose products on which

Matching is done within each site. For each treatment week we select the four closest control weeks based on sales of the non-substitute product classes.[27] The set of products that are used for matching are shown in table 16 of the Appendix. We grouped the salty snack experiments together, and the candy and Chocolate Chip Famous Amos experiments together for defining sets of products for matching.[28] We use different subsets of products on which to match at different sites, due to changes in availability or the product mix of the assorted classes at particular sites.

These three steps enable us to examine the results of each experiment. Note that for each experiment, we have one outcome for each non-focal product class at a site. Matching estimators are usually discussed in the context of a single outcome of interest, such as in Lalonde (1986), or Dehejia and Wahba (1999).[29] In our context, the "average treatment effect" is a vector of outcomes because we have multiple outcomes of interest. Our approach, therefore, is to use the matching methods developed in the treatment effects literature to generate matched observations.[30] We then use this matched sample to report mean outcomes for all products for the treatment and control weeks, along with the percentile of the distribution of all control outcomes to which the mean results relate. Thus, we measure substitution from product k to product j as:

$$\Delta q_j = E_i[q_{ij}|A_{J\setminus k}, T_i = 1] - E_i[q_{ij}|A_J, T_i = 0, N_i \leq 4]$$

where expectations are taken over weeks, indexed by i, and J denotes the full set of product classes. The variable T_i is an indicator variable denoting whether week i belongs to the treatment period in which product k was exogenous removed, and N_i is the distance rank

to match based on ex-ante notions of substitutability. In practice, however, week-to-week (or visit-to-visit) variation is quite noisy, and this resulted in difference-in-difference estimators that primarily captured random week-to-week fluctuations. We also ran all analyses using a set of "admissible" control weeks that were identified on the basis of a site's product mix during the experimental period. In this method, we admitted control weeks for which each product carried at a site during the experimental weeks was available in at least 80 percent of the machines at each site. This method yielded qualitatively similar results to the matching estimates that we report, but resulted in more experiments for which total vends increased when a product was removed. We report the nearest-neighbor matching estimator (using non-substitute products as the matching variables) as our baseline estimates because the statistical properties of these estimators are well understood. (See Abadie and Imbens (2006).) In contrast, admitting control weeks on the basis of a non-linear function of a vector of availability dummies for a set of products is (to us) less well understood. Finally, we also ran all analyses using the full control set of approximately 120 weeks. The matched estimators perform significantly better, particularly with respect to levels of total vends.

[27] All estimates were also run using ten matched control weeks for each treatment week; results were qualitatively similar to the baseline estimates reported here for four matches.

[28] The Chocolate Chip Famous Amos cookie experiment was grouped with the candy experiments because of the presence of chocolate. The set of products for matching in the Zoo Animal Cracker experiment was allowed to differ from the candy and Famous Amos experiments because the vending operator identified that product *ex-ante* as potentially having a different set of substitute products.

[29] Abadie and Imbens (2006) work out the large sample properties of matching estimators for average treatment effects in this context, and Imbens (2004) provides a review of this literature.

[30] We use the nnmatch command in Stata, described in Abadie, Drukker, Herr, and Imbens (2004), and choose the Mahalanobis metric for measuring the distance between the treatment and control vectors of covariates.

of each potential control week produced by the matching procedure.[31] We also estimate the effect of a product removal on the inside market shares of all remaining products. Thus, we measure:

$$\Delta s_j = E_i\left[\frac{(q_j|A_{J\backslash k}, T_i = 1)}{(\sum_j q_j|A_{J\backslash k}, T_i = 1)}\right] - E_i\left[\frac{(q_j|A_J, T_i = 0, N_i \leq 4)}{(\sum_j q_j|A_J, T_i = 0, N_i \leq 4)}\right]$$

where s_j denotes the inside market share of product j.

We estimate outcomes for each site and each experiment, which generates a set of 48 tables of outcomes (six sites times eight experiments). In order to capture the overall effect of an experiment, we sum over the average weekly rates at each site during the treatment and control periods, and compute the difference, as well as the percentile of the distribution of vends at all sites to which the rate corresponds.[32] This adds eight more tables to the set of results.

3.1.1 Inference

Whenever we report the effect of a product removal, we report the quantile of the distribution of overall weekly sales for each outcome (treatment and matched control weeks). The quantile corresponding to the matched control weeks gives a sense of how the matched sample compares to the full distribution across all weeks. Similarly the quantile corresponding to the treatment outcome allows one to compare the outcome to the full distribution of weeks. However, neither statistic gives a direct way to infer statistical significance for the effect of the removal, according to the matching estimator. For this, we use a falsification study, similar to a bootstrap-type procedure. For each product removal, we choose three control weeks at random from each site, and assign these to be "treatment" weeks. For these false treatment weeks, we perform the same matching procedure, and construct the same outcomes of interest (vends, difference in vends, revenues, etc.), as in the baseline analysis. We repeat this for 1000 trials, and report the values of this falsification procedure at the 5th and 95th percentiles as a 90% Confidence Interval for the matching results.[33]

[31] Thus, we denote the closest match as having rank $N_i = 1$, so that $N_i \leq 4$ selects the four nearest neighbors using the method in Abadie, Drukker, Herr, and Imbens (2004).

[32] Note, this essentially weights the contribution of each site to the total impact of the experiment by its overall sales level. Different sites may run a particular experiment for shorter or longer periods relative to other sites. For example, one site may stock out a product for two weeks, while another stocks out the product for three weeks. An alternative weighting across sites would be to sum all vends together and then compute weekly rates. This would weight the average based on both sales levels at each site and the number of weeks for which a site ran a particular experiment. We prefer weighting by site because the length of time for a given experiment at a given site varies for unhelpful reasons. For example, slower sites may be visited less often and have longer runs but lower weekly sales.

[33] Note that this procedure will sometimes result in a treatment mean that is outside the 90% confidence interval when the difference is not significant, and vice versa, depending on the variance in the matched control observations. When only one confidence interval is reported (e.g. tables 2 and 3), we report the significance level associated with the difference.

3.2 Changes in Purchasing Patterns

Table 1 reports results from one of the 56 tables–namely, the overall changes in purchasing patterns (summed over sites) of the Snickers removal.[34] The top panel of the table reports vends, and the bottom panel reports inside market shares. The first column in each table reports average weekly vends for the matched control weeks. The second column reports the percentile at which the mean of the matched control weeks falls relative to the full distribution of sales for all control weeks.[35] The third column reports average total vends for the treatment weeks, and the fourth column reports the percentile of the distribution of sales for all control weeks with which the treatment outcome is associated. If a product's average total weekly sales during the treatment weeks exceeds total weekly sales for all control weeks, we report the 100th percentile. The fifth column reports the difference in the two means, and the last column reports the percentage increase in sales for the substitute good. For example, during the Snickers experiment, table 1 shows that Peanut M&Ms sold 118.4 more units in an average treatment week; its mean total weekly sales during the control period were 359.9, and the percentage increase was 32.9. Sales of Peanut M&Ms in the matched control sample exceeded sales in 73.6% of all control weeks, and the average treatment outcome of 478.3 sales exceeded sales in 99% of all control weeks. Both the treatment mean and the mean difference are outside the 90% Confidence Interval of the overall distribution for Peanut M&M sales. The magnitudes of the percentile changes among the 'top substitutes' are quite striking across all experiments, and it is common to see very large changes in sales percentiles between treatment and control periods for the top products.

Table 1 also shows that total vends during the matched control (treatment) period correspond to the 74th (73rd) percentile of the overall distribution of total vends. Overall, total vends are only 0.1% lower during the treatment weeks when Snickers is removed. This likely reflects at least two factors: first, most consumers purchase another product when Snickers is not available (as opposed to walking away), and second, demand was relatively high when the Snickers experiment was run. The rows in both panels are sorted by the mean difference in vends, so products toward the top of the list are those whose sales increased the most when Snickers was removed. Sales of the top five products (Peanut M&Ms, Twix Caramel, Assorted Chocolate, Assorted Energy, and Zoo Animal Cracker) increased by a total of 370 vends during the treatment period, which exceeds the average level of Snickers vends during the matched control weeks of 323. The products with the largest percentage change are found by examining the last column.[36] Examining the lower panel of table 1 allows one to normalize by overall sales levels by comparing changes in inside market shares.

Tables 2 and 3 summarize the results for all eight product removals. Table 2 reports the top five substitutes for each focal product(s) based on level changes in sales summed across

[34] All 56 tables are available online in our online appendix. See the link for this paper at http://mortimer.fas.harvard.edu.

[35] With no matching, most of these percentiles will be near 50%. The matched control weeks admits greater variance across products in these percentiles.

[36] In the Snickers experiment, the top five products in terms of percentage changes are: Twix Caramel, Peanut M&Ms, Ruger Vanilla Wafer, Assorted Chocolate, and Zoo Animal Cracker.

sites; table 3 reports the top five substitutes for each focal product(s) based on percentage changes in sales. The last three rows in each panel report changes in the sales of: focal product(s), top 5 substitutes, and total vends between the treatment and matched control weeks. For all but one removal (Peanut M&Ms), we observe a reduction in total vends. Substitution to the top five products exceeds the average number of vends of the focal product during the matched control weeks in four removals in table 2.[37] The fact that vends to the top five substitutes exceed the vends of the focal product in this analysis implies that matching cannot fully control for changes in overall levels of demand across treatment and control weeks.

3.3 Profit Impacts

We observe prices and wholesale costs of Mark Vend, as well as its participation in rebate programs, so we can compute the total variable profit for each week with and without rebate payments. For upstream manufacturers, we observe revenue, but not production costs. We observe which products are co-owned by individual manufacturers, and we can calculate the revenue impacts across manufacturers that result from each product removal.

Table 4 reports the weekly profit impact of the Snickers removal for the downstream firm at the level of individual product classes, without manufacturer rebate payments, which are made annually based on Mark Vend's total manufacturer-level sales.[38] Manufacturers are listed for each product class, with assorted product classes noting multiple manufacturers. The first column reports Mark Vend's margin; his margin for Snickers (ignoring his rebate payments) is 21 cents, and he loses \$68.40 of variable profit per week on this product when it is removed. Sales of Assorted Salty Snacks are also down during the treatment week, and he loses \$51.60 of variable profit per week from this product class. He gains from other products: for example, increased sales of Peanut M&Ms contribute an additional \$25.20 in weekly profit. Overall, total vends are down, but the net profit impact for Mark Vend is positive, with an increase in profit of \$2.65 per week. This is generated in part by consumers' willingness to purchase other products when Snickers is removed (rather than leaving empty-handed), and in part by the relatively low margin that Mark Vend receives on vends of Snickers (e.g., \$0.21 vs. \$0.48 for Assorted Energy products).

Table 5 summarizes the total weekly profit impact of each experiment for Mark Vend, with and without rebate payments. Two experiments (Snickers and Peanut M&Ms) result in a profit increase for Mark Vend when rebate payments are not included.[39] All three Mars & Co. product removals result in higher average margins for Mark Vend. The most striking example of this effect is seen for the double removal of Snickers and Peanut M&Ms, in which

[37]Table 3 reports changes in the inside market share of the top five substitutes based on percentage changes. The combined change in the inside market shares of these five products exceeds the market share of the focal product in three experiments.

[38]Products are listed in the same order as in table 1 (based on their increase in sales when Snickers is removed). The full set of 56 profit impact tables are provided in the online appendix.

[39]However, the Peanut M&Ms result differs from the Snickers result because overall vends are higher during the treatment weeks than the control weeks in this case.

average margins increase by 1.2 cents per vend. The effect of this increase in margin is that Mark Vend loses only \$15.00 per week from the removal of these two products, despite a reduction in overall weekly sales levels of 218 units. The last two columns of table 5 report the change in Mark Vend's average margin and profit impact for each removal, taking rebate payments into account. Mark Vend receives relatively generous rebate payments from Mars & Co., with the result that the profit impact of removing Snickers is negative, at -\$2.50, with his average downstream margin falling by 0.05 cents per unit, once rebate payments are considered.

Table 6 reports the revenue impact of each removal for all manufacturers, with and without rebate payments.[40] Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15. In all but one case, the manufacturer of the focal product had lower revenues during the treatment period in which its product was removed.[41] The largest effect was for Mars Inc. during the joint stock-out of Snickers and Peanut M&Ms, for which its revenues declined by \$220.52 per week.[42]

4 Structural Models of Demand

In this section we describe two models of demand commonly used for analyzing markets for differentiated products. Both models assume knowledge of market size, M. We first describe our method for determining market size, and then specify the structural demand models.

4.1 Calculating Market Size

Unlike many settings, in which market size is assumed to be constant over time, the number of people considering a purchase from a vending machine may change significantly from one week to the next. For example, the employees at one site may have an important deadline one week, which increases demand for vending products temporarily. Unfortunately, we do not observe how many people pass by each vending machine in our sample in any given week who are considering making a purchase.[43] Thus, we estimate market size using data on total

[40]Revenues to manufacturers are calculated as the wholesale cost paid by Mark Vend to the manufacturer. When rebate payments are included, these reduce the amount that Mark Vend pays to each manufacturer by a fixed percentage, which was provided to us by Mark Vend.

[41]The exception is that PepsiCo's revenues increase when Cheetos are removed. This is likely due to the presence of week-to-week variation in relative market shares, market size, and the availability of products in some of the Assorted classes that are not fully controlled for by the matching procedure. However, the result may also partly reflect the fact that PepsiCo owns a very large fraction of the available substitute salty snacks.

[42]The decline in revenues for Snyder's, which did not have a focal product in any of the experiments, is due primarily to changes in the set of products that Mark Vend stocked over time. In particular, several Snyder pretzel products were replaced with products manufactured by 'Minor' manufacturers and PepsiCo.

[43]This is more than a simple data limitation, in the sense that even additional data monitoring (e.g., pressure mats or video cameras on the machines) would not provide perfect information on how many consumers may be considering making a purchase as opposed to simply being located near a vending machine.

vends over time at the machine level. Our baseline model specifies market size on the basis of the following regression:

$$y_{rv} = \eta_{rd} + \tau_{rt} + \epsilon_{rv} \tag{4}$$

where y_{rv} is five times total vends at visit v for machine r, η_{rd} is a full set of $machine \times day\_of\_week$ fixed effects (six for each machine), and τ_{rt} is a full set of $machine \times month \times year$ fixed effects (38 for each machine).[44] We specify market size as the predicted value $\hat{y}_{rv}$ from this regression, with two additional restrictions.[45] We investigated an alternative specification for market size and ran all prediction exercises using the alternative estimate. The results were qualitatively the same as our baseline estimates.[46]

4.2 Nested Logit and Random-coefficients Logit Specifications

We specify two models of demand: nested logit and random-coefficients logit, which are estimated from the full dataset. We consider a model of utility where consumer i receives utility from choosing product j in market t of:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \tag{5}$$

The parameter δ_{jt} is a product-specific intercept that captures the mean utility of product j in market t, and μ_{ijt} captures individual-specific correlation in tastes for products.

Nested Logit

In the case where $(\mu_{ijt} + \varepsilon_{ijt})$ is distributed generalized extreme value, the error terms allow for correlation among products within a pre-specified group, but otherwise assume no correlation. This produces the well-known nested-logit model of McFadden (1978) and Train (2003). In this model consumers first choose a product category l composed of products g_l, and then choose a specific product j within that group. The resulting choice probability for product j in market t is given by the closed-form expression:

$$p_{jt}(\delta, \lambda, a_t) \quad = \quad \frac{e^{\delta_{jt}/\lambda_l}(\sum_{k \in g_l \cap a_t} e^{\delta_{kt}/\lambda_l})^{\lambda_l - 1}}{\sum_{\forall l}(\sum_{k \in g_l \cap a_t} e^{\delta_{kt}/\lambda_l})^{\lambda_l}} \tag{6}$$

where the parameter λ_l governs within-group correlation, and a_t is the set of available products in market t.[47]

[44] The R^2 from this regression is 0.66.

[45] The restrictions are that M must be at least 30 people per visit and must be greater than the total vends for the machine-visit observation.

[46] The alternative model specifies a daily rate of arrival. For each machine-visit observation we denote Δt as the elapsed time since the previous service visit in days and estimate a least-squares regession of $y_{rv}/\Delta t$ on a series of machine fixed effects and a full set of 38 month*year dummies.

[47] Note that this is not the IV regression/"within-group share" presentation of the nested-logit model in Berry (1994), in which σ provides a measure of the correlation of choices within a nest. Roughly speaking,

Random-coefficients Logit

The random-coefficients logit allows for correlation in tastes across observed product characteristics.[48] This correlation in tastes is captured by allowing the term μ_{ijt} to be distributed according to $f(\mu_{ijt}|\theta)$. A common specification is to allow consumers to have independent normally distributed tastes for product characteristics, so that $\mu_{ijt} = \sum_l \sigma_l \nu_{ilt} x_{jl}$ where $\nu_{ilt} \sim N(0,1)$ and σ_l represents the standard deviation of the heterogeneous taste for product characteristic x_{jl}. The resulting choice probabilities are a mixture over the logit choice probabilities for many different values of μ_{ijt}, shown here:

$$p_{jt}(\delta,\theta,a_t) = \int \frac{e^{\delta_{jt}+\sum_l \sigma_l \nu_{ilt} x_{jl}}}{1+\sum_{k\in a_t} e^{\delta_{kt}+\sum_l \sigma_l \nu_{ilt} x_{kl}}} f(v_{ilt}|\theta) \tag{7}$$

Additional Parametrizations

In both the nested-logit and random-coefficient models δ_{jt} consists of product-site intercepts, so that the average taste for an individual product varies from site to site.[49] For the nested-logit model, we allow for heterogeneous tastes across four major product categories or nests: chocolate, cookie, energy, and salty snack.[50] For the random-coefficients specification, we allow for three random coefficients, corresponding to consumer tastes for salt, sugar, and fat.[51] An observation for estimation groups machine visits into unique choice sets of the 417 product-site pairs. We report and discuss the estimated parameter values from these models in the section describing the results of the hold-out analyses.

5 Comparing Predicted and Nonparametric Results

Sales are predicted by both models at the machine-visit level using the estimated parameter values and market size y_{rv}. We describe the predicted vends during the treatment and

in the notation used here, $\lambda = 1$ corresponds to the plain logit, and $(1-\lambda)$ provides a measure of the "correlation" of choices within a nest (as in McFadden (1978)). The parameter λ is sometimes referred to as the "dissimiliarity parameter."

[48] See Berry, Levinsohn, and Pakes (1995).

[49] These correspond to the 417 site-product pairs described in the data section. In addition, one can add a market-level demand shifter ξ_t.

[50] The vending operator defines categories in the same way. "Candy" includes both chocolate and non-chocolate candy items. "Energy" includes products such as peanuts, fruit rolls, crackers, and granola bars.

[51] We do not allow for a random coefficient on price because of the relative lack of price variation in the vending machines. We also do not include random coefficients on any discrete variables (such as whether or not a product contains chocolate). As we discuss in Conlon and Mortimer (2009), the lack of variation in a continuous variable (e.g., price) implies that random coefficients on categorical variables may not be identified when product dummies are included in estimation. We did estimate a number of alternative specifications in which we include random coefficients on other continuous variables, such as carbohydrates. In general, the additional parameters were not significantly different from zero, and they had no appreciable effect on the results of any prediction exercises.

matched control weeks and compare them to the actual vends. This exercise does not predict substitution patterns *per se*. Rather, it predicts vends in the environment of the treatment weeks, and again in the environment of the control weeks. Thus, non-experimental changes that occur during the control weeks are included in these predictions. There are two main sources of these non-experimental changes. The first is changes in the characteristics of other products (e.g., a manufacturer may change the fat content of its pretzels, or run a national advertising campaign in a particular week). The second is changes in the availability of other products (e.g., a manufacturer discontinues a particular product, or Mark Vend changes the set of products it carries at a site). The models will adjust for these changes as well as the removal of the focal product. In this sense, neither the nonparametric results nor the predicted vends presented here can be interpreted as the "pure" treatment effect of a product removal.

5.1 Predicted and Actual Sales

Table 7 reports predicted vends from the nested-logit model during treatment and control weeks next to the results of the nonparametric analyses for the Snickers experiment; table 8 repeats the exercise using the random-coefficients logit model.[52] Predicted vends for the matched control weeks fit well in both tables 7 and 8. While both demand models will necessarily fit average vends of all products quite closely (through the choice of the δ_{jt} parameters), predictions for vends in the matched control weeks need not fit well, because both demand models are estimated off of the full dataset. In this sense, the predicted vends for the matched control weeks provide some insight into the importance of week-to-week variation in market shares in this setting.

Predicted vends fit less well for the treatment weeks than for the matched control weeks. Both models predict more substitution to the outside good than the actual data display: total predicted vends go down by roughly two percent, compared to only 0.1 percent in the actual data. Table 7 shows how the predicted sales patterns in the nested-logit model rely on the nesting structure specified for consumer tastes. Snickers is in the "candy" nest, along with Twix Caramel, Peanut M&Ms, Assorted Chocolate, Assorted Candy, Raisinets, and Skittles. Sales of all of these products are predicted to increase by eight to 17 percent when Snickers is not available. Three of these products (Twix Caramel, Peanut M&Ms, and Assorted Chocolate) are also identified as top substitutes in the nonparametric analyses. Actual vends of Skittles were lower during the treatment weeks than during the matched control weeks. However, it seems unlikely that Snickers is complementary to Skittles in consumption, and the fact that the model does not predict lower sales of Skittles may be viewed as a desirable feature. Assorted Potato Chips are predicted to have relatively high sales during the treatment period. This results from changes in the set of products

[52] Two full sets of 56 tables (one set for each model) are available in the online appendix. At the individual site level, these tables also report changes in the availability of each product class, as well as any changes to the inclusive value of the assorted product classes that occur between the treatment and control periods due to changes in the individual products that comprise these classes. Further discussion is provided in the online appendix.

included in the Assorted Potato Chip class at two individual sites.[53] Two assorted product classes (salty snack and cookie) are predicted by both models to have lower sales during the treatment period compared to the matched control weeks; actual vends are also lower during the treatment period for these classes. This also results from changes in the component products at individual sites: more-popular brands were carried during the matched control weeks than during the treatment weeks.[54] Relative to the nested-logit model, the random-coefficients model in table 8 predicts smoother sales patterns across product categories. Thus, it successfully predicts more vends of cookie and energy products, which are closer to Snickers in some of the observable product characteristics (i.e., fat and sugar).

5.2 Discussion of Predicted Sales

Both models predict fewer sales of the products identified as 'top substitutes' in the matching exercise, and predict more sales to other products. This may be explained partly by noise in the matched data, which we discussed in section 3. However, another explanation is that the parametric models are misspecified. We discuss two potential sources of misspecification.

The first potential source of misspecification is that we have not chosen the correct form for $f(\mu_{ijt}|\theta)$. For the nested-logit model, misspecification is well understood, and occurs when the pre-specified groupings of products that determine substitution patterns do not fully capture consumer behavior. For example, the experiments suggest that many consumers substitute from Snickers to Salted Peanuts, but Salted Peanuts also appear as a substitute for Doritos. Since Doritos and Snickers do not display cross-substitution effects with each other, this creates a dilemma regarding the nest to which Salted Peanuts belong.

There may be several reasons why $f(\mu_{ijt}|\theta)$ is misspecified in the random-coefficients model. The one most strongly suggested by our experimental results is that there are omitted product characteristics for which consumers have heterogeneous tastes. For example, sales of Peanut M&Ms and Salted Peanuts are higher during the Snickers removal than the model suggests, and sales of Salted Peanuts are higher during the Peanut M&M's removal. However, Peanut M&Ms and Salted Peanuts are not particularly similar along observable dimensions. To formalize this intuition, suppose that the true utility model is given by:

$$\tilde{u}_{ijt} = \delta_{jt} + \sum_l \sigma_l \nu_{ilt} x_{jl} + \gamma_i z_{jt} + \varepsilon_{ijt} \qquad \gamma_i \sim N(0, \sigma_z) \tag{8}$$

where z_{jt} is an unobserved characteristic for which consumers have heterogenous tastes. One can safely exclude the omitted characteristic only when $\sigma_z = 0$. There are two restrictions

[53] Less-popular brands of chips were stocked during the matched control weeks than during the treatment weeks. The change is most noticeable at sites 5055 and 5655.

[54] The nested-logit model also predicts slightly lower vends during the treatment period for the assorted energy product class, and for one individual product (Grandma's Chocolate Chip Cookie). The latter prediction is due to the absence of Grandma's Chocolate Chip Cookie during the treatment period at one site, and the fact that the nested-logit model does not predict much substitution from Snickers to the cookie nest (so predicted vends at other sites do not increase very much). Detailed information on the effects of changes in the component products of the assorted classes, and the availability of all products, is included in the full set of tables in the online appendix.

that the parametric models of u_{ijt} impose. The first is that correlation among tastes is parametrized by μ_{ijt}, which is projected onto the space of observable characteristics x_{jt}.[55] This approach will always leave some residual correlation, unless the unrestricted $J \times J$ matrix can be estimated. The second restriction is that, conditional on the consumer type μ_{ijt}, all substitution follows a standard logit, including the IIA (Independence of Irrelevant Alternatives) property. The random-coefficients model is a mixture over many different IIA logit models. However, traces of the IIA property remain which may lead to underprediction of substitution to the best substitutes and overprediction of substitution to the worst substitutes.[56]

The second source of difficulty for parametric models is that the set of available products a_t is assumed to be exogenous to the decision-making process. In reality, firms choose product mix carefully in order to maximize profits. The endogeneity of the choice set (or product mix) is something that the researcher cannot generally control, and it is precisely this variation that is used to identify these models. This creates a challenge when we use the models to extrapolate out of sample. For example, Snickers and Twix are both nearly always in stock, and substitution between the two is identified in part by differences in observed substitution to both products from a third product (such as Raisinets) which is sometimes, but not always, stocked. The machines or time periods in which Raisinets are stocked are chosen by Mark Vend to maximize profits, and may reflect locations or times when the taste for Raisinets is especially high. This leads the model to over-value Raisinets and to misspecify substitution between Snickers and Twix. The ability to exogenously alter choice sets reduces our reliance on endogenous forms of choice-set variation, and allows us to test the sensitivity of the models to the presence of exogenous variation, which we do in Section 7.

There is a third source of misspecification that is more special to our setting, and that is the absence of price variation.[57] The ability to identify substitution patterns in the types of demand models estimated here benefits greatly from the presence of price variation, particularly when a strong instrument can be found for addressing the endogeneity of variation in price.[58]

5.3 Predicted Profit Impacts

Differences in predicted vends across products may not have much impact on firms' bottom lines, particularly in a setting with limited price variation. Table 9 reports the predicted profit difference for the downstream firm in the treatment vs. matched control weeks using

[55]McFadden and Train (2000) show that any random utility model can be represented as a mixture of logits as in (5). However, this depends on having a sufficient space of x_{jt}'s.

[56]There are several approaches to dealing with these problems in the literature. One example is the pure characteristics model of Berry and Pakes (2007) which avoids the logit error altogether, but restricts the substitution matrix to vary only according to observable characteristics. Another approach is found in Bajari, Fox, Kim, and Ryan (2010) which proposes a method to non-parametrically recover $f(\mu_{ijt}|\theta)$.

[57]There are, however, many markets with limited price variation, and product availability is often a key consideration in these contexts (e.g., movie theaters, iTunes, and subscription services such as Netflix).

[58]For a discussion of the role of price variation in identifying discrete-choice models of demand, see Ackerberg and Rysman (2005), Berry and Haile (2008), and Fox and Gandhi (2010).

the random-coefficients logit model. A comparison of the results to those in table 5 shows that the change in the downstream firm's average margin is predicted quite well, implying that discrepencies between predicted and actual vends for individual products matters little for the effect on the firm's average profitability on a per-unit basis. Indeed, the most important determinant for the change in Mark Vend's average margin is the difference between the wholesale cost of the focal product compared to other goods: Snickers and Peanut M&Ms have very high wholesale costs, and Mark Vend's average margin increases when those products are removed. This effect is offset somewhat by the rebate payments. The model predicts a more negative overall profit impact for each removal relative to the non-parametric estimates, because total vends during the treatment period are predicted to be lower than the actual data. However, the patterns are correlated with those from the non-parametric analyses: the two most positive/least negative predicted differences correspond to the Peanut M&Ms and Snickers interventions, which were also estimated to have the most positive impacts in the non-parametric results.

Table 10 reports the impact of the product removal on manufacturer revenue predicted by the random-coefficients model, with and without rebate payments. Relative to the non-parametric estimates in table 6, some experiments have predicted impacts that are quite similar (e.g., the cookie experiments), while others predict a larger negative impact for the manufacturer of the focal product and corresponding larger positive impacts for competing manufacturers. This reflects the fact that many of the non-focal products with under-predicted sales are owned by the manufacturer of the focal product (e.g., both Snickers and Peanut M&Ms are owned by Mars Inc.). As expect, the impact of each product removal is less negative for manufacturers after accounting for rebate payments.

6 Simulating Stockout Events

The exercise in which we predict vends for the treatment and control weeks is an exercise in which we are rarely interested from an economic point of view. Policy makers are generally interested in understanding the impact of a control variable (in this case, the decision to stock a product), rather than understanding the joint effect of a control variable and other factors. Both the non-parametric approach and the prediction exercise above conflates these things, modeling the product removal simultaneously with changes in other factors that shift consumer tastes (such as national manufacturer advertising campaigns). For that reason, the prediction exercise is also a strenuous test on which to expect the models to perform well, because of the need to adjust for multiple factors simultaneously. In this section, we simulate the effect of a product removal directly, using the estimated model parameters and holding all other factors fixed.

Both the nested logit and random-coefficients logit specifications have straightforward predictions regarding how demand should respond to a product removal. The expected change in sales of product j when product k is removed is:

$$\Delta q_{jt}^{(k)} = M_t \left(p_{jt}(\delta, \theta, a_{t \setminus k}) - p_{jt}(\delta, \theta, a_t) \right). \qquad (9)$$

We calculate $\Delta q_{jt}^{(k)}$ for the set of products available during the treatment weeks at each site. Table 11 reports the results of this calculation for the Snickers removal, summing over all sites.[59] We choose M_t so that the sales levels of the focal product(s) match the sales levels in the actual data during the matched control weeks. The left panel reports the change in sales for a simulated removal of Snickers based on the nested logit and random-coefficients logit models. Under the random-coefficients specification, Snickers has sales of 323 units when it is available. Of these 323 "Snickers consumers," 73 leave the machine and 250 purchase another product (e.g., 34 purchase an Assorted Energy product, 33 purchase Peanut M&Ms, etc.). The downstream firm loses $69 in variable profit from the sales of Snickers, but gains $95 in variable profit across all substitute goods. Thus, in spite of the fact that 73 Snickers consumers leave the machine, the removal of Snickers is profitable for the downstream firm.

Table 12 summarizes the overall profit impacts of simulations in which we remove each of the focal products. As expected, vends decrease in every case because of substitution to the outside good. As in the nonparametric analyses and the prediction exercise, average margins increase in the three candy removals as consumers substitute to products with lower wholesale costs for the downstream firm, although this effect is ameliorated by the rebate payments. Mark Vend's variable profits are still forecast to increase when Snickers is removed, although the rebate payments provide an incentive to carry Peanut M&Ms. All other product removals result in lower profits for Mark Vend, even in the absence of rebate payments. The revenue impacts for upstream firms are reported in table 13. The manufacturer of each focal product loses revenue when the product is removed, while other manufacturers gain. The gains to other manufacturers are spread out fairly widely.

7 Hold-out Analyses

In most applications, the type of exogenous variation in product availability that we induce through our experiments does not exist in the data. Demand models are thus identified entirely from naturally-occuring variation in choice sets. This variation may arise from many sources (e.g., changes in stocking decisions, price, and other characteristics), but is often endogenously determined by firms. Our setting permits us to examine the sensitivity of our demand estimates to the presence of exogenous variation in choice sets.

Table 14 reports the estimated coefficients that govern correlation in consumer tastes when we estimate our demand models on subsets of the data that exclude different sets of our exogenous product removals. The random-coefficients model is shown in the top panel, and the nested logit model in the second panel. The bottom panel reports the number of products and choice sets for each sample of the data. The parameter values from the base model, in which the full set of data is used for estimation, are reported in the first column.[60] The second column reports the estimated parameters when the model is estimated on only control observations (i.e., we exclude all experimental periods). The estimates of σ_{sugar} and

[59] The full set of 56 tables is provided in the online appendix.

[60] The log likelihood values are reported in the last row of each panel. They are the same for both models, but are not comparable across the different hold-out sub-samples because they apply to a different number of observations.

λ_{candy} change significantly (from 5.25 to 2.91 for σ_{sugar} and from 0.57 to 0.66 for λ_{candy}). The other parameters are more robust to the exclusion of the treatment observations. The third column of table 14 reports the estimated parameters when the models are estimated without the benefit of the data from the candy experiments. We see the same effect on the parameter values: the estimated correlation in tastes for sugar and the candy nest are lower (i.e., σ_{sugar} is lower and λ_{candy} is higher).[61] The remaining columns repeat the exercise by withholding data from the salty snack and cookie experiments in turn. The estimated parameter values are much more robust to these hold-out exercises.

The bottom panel provides some guidance for assessing these results. The total number of products is reported next to each category name (e.g., there are 417 products total, 115 of which are candy products). The experimental interventions increase the total number of unique choice sets from 1096 to 1734. Among candy products alone, there are 427 unique choice sets in the full dataset, but only 262 in the subset of the data that includes only control observations. The salty snack category has about 60 percent more products than the candy category (187 vs. 115), but 120 percent more unique within-category choice sets when using only the control observations (578 vs. 262). Removing data from the salty snack interventions does reduce the number of within-category choice sets, but not as dramatically (578 vs. 794).

The choice-set data help to better understand the challenges of identification in demand models when there is relatively little variation in stocking (or other product characteristics) within a group of closely-related products. In our setting, the experimentally stocked-out candy bars have large market shares relative to their salty-snack counterparts, and are nearly always available simultaneously to consumers during the control observations. Furthermore, there is limited variation (relative to the number of products) in consumer choice sets generated by other candy products. Thus, identifying correlation in tastes for the characteristics that are important to these goods is greatly aided by our ability to exogenously create choice-set variation.[62]

8 Conclusion

Our analysis of product availability makes use of a field experiment in which top-selling products are removed from a set of 60 vending machines. We analyze changes in purchasing patterns and profits in two ways: first, using nonparametric techniques common to the treatment effects and experiments literature; second, using structural demand estimation. We find substitution patterns that seem quite sensible, and we note that in the short-run, relatively few people leave the market when a product is out of stock. We find evidence of the incentive problems facing firms in vertically-separated markets: some product removals result in lower revenues for upstream manufacturers but higher profits for the downstream vending

[61] Recall that the λ parameters give an indication of the 'dissimilarity' of products, so that a higher estimate of λ indicates less substitution between two products within a nest.

[62] As discussed earlier, price variation will be helpful in many settings for generating additional variation in the choice sets facing consumers.

operator in the absence of rebate payments because consumers substitute to products with lower wholesale costs.

We discuss the trade-offs of both empirical approaches. The treatment-effects approach places no parametric restrictions on substitution patterns, but noisy estimates occasionally imply nonsensical effects of a product's removal. The structural approach avoids this problem, but at the cost of imposing restrictions on the substitution patterns and profit impacts associated with a product's removal. Using hold-out analyses, we find that the structural model is sensitive to the absence of the experimental data in parts of product space where relatively little non-experimental variation occurs in consumers' choice sets.

Product availability is a critical feature of many markets; firms make both long-run and short-run decisions about the capacity or inventory of different products to stock, which brands to carry, and how to respond to changes in the product availability of rivals. Product availability can vary over time due to mergers, foreclosure, or other factors. Despite the prevalence of these issues, relatively little empirical evidence exists in the IO literature on the importance of product stocking decisions for firm profits, consumer choices, or vertical relationships. Using vending machines, we are able to exogenously alter a firm's product mix in order to shed light on two central outcomes of interest (i.e., substitution patterns and profit impacts for both upstream and downstream firms), which apply broadly to more complex settings. The experimental approach also provides an opportunity to compare the trade-offs and assess the sources of identification of different empirical methods.

References

Abadie, A., D. Drukker, J. Herr, and G. Imbens (2004): "Implementing Matching Estimators for Average Treatment Effects in Stata," *The Stata Journal*, 4(3), 290–311.

Abadie, A., and G. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235–267.

Ackerberg, D., and M. Rysman (2005): "Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects," *RAND Journal of Economics*, 36(4).

Aguirregabiria, V. (1999): "The Dynamics of Markups and Inventories in Retailing Firms," *Review of Eocnomic Studies*, 66, 278–308.

Anderson, E. T., G. J. Fitzsimons, and D. Simester (2006): "Measuring and Mitigating the Costs of Stockouts," *Management Science*, 52(11), 1751–1763.

Angrist, J., and J.-S. Pischke (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *The Journal of Economic Perspectives*, 24(2), 3–30.

Anupindi, R., M. Dada, and S. Gupta (1998): "Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products," *Marketing Science*, 17(4), 406–423.

Arrow, K., T. Harris, and J. Marschak (1951): "Optimal Inventory Policy," *Econometrica*, 19(3), 250–272.

Bajari, P., J. Fox, K. I. Kim, and S. Ryan (2010): "A Simple Nonparametric Estimator for the Distribution of Random Coefficients," Working Paper.

Berry, S. (1994): "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, 25(2), 242–261.

Berry, S., and P. Haile (2008): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Working Paper.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S., and A. Pakes (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48(4), 1193–1125.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010): "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment*," *Quarterly Journal of Economics*, 125(1), 263–305.

Brownell, K., T. Farley, W. Willett, B. Popkin, F. Chaloupka, J. Thompson, and D. Ludwig (2009): "The public health and economic benefits of taxing sugar-sweetened beverages," *New England Journal of Medicine*, 361, 1599–1605.

Brownell, K., and T. Frieden (2009): "Ounces of prevention: The public policy case for taxes on sugared beverages," *New England Journal of Medicine*, 360, 1805–1808.

Bruno, H. A., and N. Vilcassim (2008): "Structural Demand Estimation with Varying Product Availability," *Marketing Science*, 27(6), 1126–1131.

Cai, H., H. Fang, and C. Yuyu (2009): "Observational Learning: Evidence from a Randomized Natural Field Experiment," *American Economic Review*, 99(3), 864–82.

Conlon, C., and J. H. Mortimer (2009): "Demand Estimation Under Incomplete Product Availability," Working Paper.

Dehejia, R. H., and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluation of the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062.

Edgeworth, F. (1888): "The Mathematical Theory of Banking," *Journal of the Royal Statistical Society*, 53, 113–127.

Einav, L., and J. Levin (2010): "Empirical Industrial Organization: A Progress Report," *The Journal of Economic Perspectives*, 24(2), 145–162.

Engber (2009): "Let Them Drink Water! What a fat tax really means for America.," `http://www.slate.com/id/2228713/`.

Engelbrecht-Wiggans, R., J. A. List, and D. H. Reiley (2005): "Demand Reduction in Multi-Unit Auctions: Evidence from a Sportscard Field Experiment: Reply," *American Economic Review*, 95(1), 472–476.

Eriv, I., and A. Roth (1998): "Predicting how People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 88(4), 848–881.

Fitzsimons, G. J. (2000): "Consumer Response to Stockouts," *Journal of Consumer Research*, 27(2), 249–266.

Fox, J., and A. Gandhi (2010): "Nonparametric Identification and Estimation of Random Coefficients in Nonlinear Economic Models," Working Paper.

Heckman, J. J. (2010): "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48(2), 356–398.

Imbens, G. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *Review of Economics and Statistics*, 86(1), 4–29.

Karlan, D., and J. A. List (2007): "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment," *American Economic Review*, 97(5), 1774–1793.

Keane, M. (2010): "A Structural Perspective on the Experimentalist School," *The Journal of Economic Perspectives*, 24(2), 47–58.

Kraiselburd, S., V. Narayanan, and A. Raman (2004): "Contracting in a Supply Chain with Stochastic Demand and Substitute Products," *Production and Operations Management*, 13(1), 46–62.

Lalonde, R. (1986): "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review*, 76(4), 604–620.

Leamer, E. (1983): "Let's Take the Con Out of Econometrics," *American Economic Review*, 75(3), 308–313.

——— (2010): "Tantalus on the Road to Asymptopia," *The Journal of Economic Perspectives*, 24(2), 31–46.

Levitt, S. D., and J. A. List (2009): "Field experiments in economics: The past, the present, and the future," *European Economic Review*, 53(1), 1–18.

Matsa, D. (2010): "Competition and Product Quality in the Supermarket Industry," Working Paper.

McFadden, D. (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, L. Lundsqvist, F. Snickars, and J. Weibull. North-Holland.

McFadden, D., and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470.

Musalem, A., M. Olivares, E. Bradlow, C. Terwiesch, and D. Corsten (2010): "Structural Estimation of the Effect of Out-of-Stocks," *Management Science*, 52(7), 1180–1197.

Nevo, A., and M. Whinston (2010): "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference," *The Journal of Economic Perspectives*, 24(2), 69–82.

Ostrovsky, M., and M. Schwarz (2009): "Reserve Prices in Internet Advertising Auctiosn: A Field Experiment," Working Paper.

Schweitzer, M., and G. Cachon (2000): "Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence," *Management Science*, 46(3), 404–420.

SIMS, C. (2010): "But Economics Is Not an Experimental Science," *The Journal of Economic Perspectives*, 24(2), 59–68.

SMITH, S. (2010): "City may curb sales of sugary beverages," *The Boston Globe*, September 20th.

SMITH, V. (1962): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Journal of Political Economy*, 70(2), 111–137.

SPENGLER, J. (1950): "Vertical Integration and Antitrust Policy," *Journal of Political Economy*, 58, 347–352.

STOCK, J. (2010): "The Other Transformation in Econometric Practice: Robust Tools for Inference," *The Journal of Economic Perspectives*, 24(2), 83–94.

TRAIN, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

TVERSKY, A., AND D. KAHNEMAN (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106(4), 1039–1061.

Table 1: Changes in Purchasing Patterns for Snickers Removal, All Sites

Product	Control Mean	Control %ile	Treatment Mean	Treatment %ile	Mean Difference	% Δ
Vends						
Peanut M&Ms	359.9	73.6	478.3*	99.4	118.4*	32.9
Twix Caramel	187.6	55.3	297.1*	100.0	109.5*	58.4
Assorted Chocolate	334.8	66.7	398.0*	95.0	63.2*	18.9
Assorted Energy	571.9	63.5	616.2	76.7	44.3	7.8
Zoo Animal Cracker	209.1	78.6	243.7*	98.1	34.6*	16.5
Salted Peanuts	187.9	70.4	216.3*	93.7	28.4	15.1
Choc Chip Famous Amos	171.6	71.7	193.1*	95.0	21.5*	12.5
Ruger Vanilla Wafer	107.3	59.7	127.9	78.6	20.6*	19.1
Assorted Candy	215.8	43.4	229.6	60.4	13.7	6.4
Assorted Potato Chips	279.6	64.2	292.4*	66.7	12.8	4.6
Assorted Pretzels	548.3	87.4	557.7*	88.7	9.4	1.7
Raisinets	133.3	66.0	139.4	74.2	6.1	4.6
Cheetos	262.2	60.1	260.5	58.2	-1.8	-0.7
Grandmas Choc Chip	77.9	51.3	72.5	37.8	-5.4	-7.0
Doritos	215.4	54.1	203.1	39.6	-12.3*	-5.7
Assorted Cookie	180.3	61.0	162.4	48.4	-17.9	-10.0
Skittles	100.1	62.9	75.1*	30.2	-25.1*	-25.0
Assorted Salty Snack	1382.8	56.0	1276.2*	23.3	-106.7*	-7.7
Snickers	323.4	50.3	2.0*	1.3	-321.4*	-99.4
Total	5849.6	74.2	5841.3	73.0	-8.3	-0.1
Shares						
Peanut M&Ms	6.2	60.4	8.2*	100.0	2.0*	33.1
Twix Caramel	3.2	41.5	5.1*	96.9	1.9*	58.6
Assorted Chocolate	5.7	53.5	6.8*	96.9	1.1*	19.0
Assorted Energy	9.8	49.1	10.5	75.5	0.8	7.9
Zoo Animal Cracker	3.6	65.4	4.2*	98.1	0.6*	16.7
Salted Peanuts	3.2	56.6	3.7*	89.9	0.5	15.3
Choc Chip Famous Amos	2.9	64.8	3.3*	90.6	0.4*	12.7
Ruger Vanilla Wafer	1.8	51.6	2.2	74.8	0.4*	19.3
Assorted Candy	3.7	35.8	3.9	49.7	0.2	6.5
Assorted Potato Chips	4.8	62.9	5.0*	63.5	0.2	4.7
Assorted Pretzels	9.4	75.5	9.5*	82.4	0.2	1.9
Raisinets	2.3	58.5	2.4	67.9	0.1	4.7
Cheetos	4.5	38.0	4.5	36.1	-0.0	-0.5
Grandmas Choc Chip	1.3	39.7	1.2	23.1	-0.1	-6.8
Doritos	3.7	37.7	3.5	20.8	-0.2*	-5.6
Assorted Cookie	3.1	55.3	2.8	44.0	-0.3	-9.8
Skittles	1.7	56.6	1.3*	20.1	-0.4*	-24.9
Assorted Salty Snack	23.6	32.7	21.8*	4.4	-1.8*	-7.6
Snickers	5.5	44.0	0.0*	0.6	-5.5*	-99.4
Total	100.0	50.3	100.0	50.3	0.0	0.0

Notes: Control weeks are selected through nearest-neighbor matching using four control observations for each treatment week. Percentiles are relative to the full distribution of control weeks. Outcomes outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 2: Top 5 Substitutes (Based on Change in Sales), All Removals

Candy Experiments			
Focal Product:	Snickers	Peanut M&Ms	Snickers + Peanut M&Ms
Top 5 Substitutes:	Peanut M&Ms* Twix Caramel* Ass. Chocolate* Ass. Energy Zoo Animal Cracker*	Snickers* Plain M&Ms* Twix Caramel* Salted Peanuts* Raisinets*	Twix Caramel* Plain M&Ms* Ass. Chocolate* Raisinets* Ass. Cookie*
Effects:	Focal (-321.4) Top 5 (+370.0) Total (-8.3)	Focal (-317.7) Top 5 (+265.2) Total (+54.1)	Focals (-287.6,-305.9) Top 5 (+305.5) Total (-218.4)
Salty Snack Experiments			
Focal Product:	Doritos	Cheetos	Doritos + Cheetos
Top 5 Substitutes:	Ass. Potato Chips* Peanut M&Ms* Salted Peanuts* Sun Chips* Ass. Energy	Sun Chips* Ass. Potato Chips* Ass. Energy Frito* Salted Peanuts*	Ass. Salty Snack* Ass. Energy Snickers* Cheez-It Original* Hot Stuff Jays*
Effects:	Focal (-210.6) Top 5 (+183.6) Total (-75.1)	Focal (-273.2) Top 5 (+284.6) Total (-122.5)	Focals (-172.7,-215.5) Top 5 (+324.1) Total (-35.9)
Cookie Experiments			
Focal Product:	Zoo Animal Crackers	Choc Chip Famous Amos	
Top 5 Substitutes:	Peanut M&M* Ass. Energy Snickers* Twix Caramel* Ruger Vanilla Wafer*	Sun Chips* Salted Peanuts* Ass. Potato Chips Raspberry Knotts* Grandma Choc Chip*	
Effects:	Focal (-210.6) Top 5 (+216.8) Total (-28.7)	Focal (-141.6) Top 5 (+155.7) Total (-179.5)	

Notes: Control weeks are selected through nearest-neighbor matching using four control observations for each treatment week. Effects report the change in average vends of the treatment and control weeks for: the focal product, the Top 5 substitutes based on changes in sales levels, and total vends. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 3: Top 5 Substitutes (Based on Percentage Change in Own Sales), All Removals

Candy Experiments			
Focal Product:	Snickers	Peanut M&Ms	Snickers + Peanut M&Ms
Top 5 Substitutes:	Twix Caramel* Peanut M&Ms* Ruger Vanilla Wafer* Ass. Chocolate* Zoo Animal Crackers*	Plain M&Ms* Snickers* Raisinets* Farleys Mixed Fruit* Salted Peanuts*	Plain M&Ms* Twix Caramel* Ass. Chocolate* Raisinets* Reeses Peanut Butter Cups*
Effects:	Focal (-5.5%) Top 5 (+6.0%) Total (-0.1%)	Focal (-5.9%) Top 5 (+4.5%) Total (+1.0%)	Focal (-5.4%,-5.7%) Top 5 (+6.4%) Total (-4.1%)
Salty Snack Experiments			
Focal Products:	Doritos	Cheetos	Doritos + Cheetos
Top 5 Substitutes:	Salted Peanuts* Ass. Potato Chips* Sun Chips* Peanut M&Ms* Choc Chip F A	Sun Chips* Frito* Salted Peanuts* Ass. Potato Chips* Doritos*	Hot Stuff Jays* Cheez-It Original* Frito* Ass. Salty Snack* Smartfood
Effects:	Focal (-3.6%) Top 5 (+3.3%) Total (-1.2%)	Focal (-4.6%) Top 5 (+4.8%) Total (-2.1%)	Focal (-3.5%,-4.4%) Top 5 (+5.9%) Total (-0.7%)
Cookie Experiments			
Focal Products:	Zoo Animal Crackers	Choc Chip Famous Amos	
Top 5 Substitutes:	Ruger Vanilla Wafer* Raspberry Knotts Twix Caramel* Peanut M&Ms* Snickers*	Sun Chips* Salted Peanuts* Raspberry Knotts* Grandma Choc Chip* Ass. Potato Chips	
Effects:	Focal (-3.4%) Top 5 (+3.0%) Total (-0.5%)	Focal (-2.5%) Top 5 (+3.2%) Total (-3.2%)	

Notes: Control weeks are selected through nearest-neighbor matching using four control observations for each treatment week. Effects report the change in average vends of the treatment and control weeks for: the focal product, the Top 5 substitutes based on percentage changes in sales levels, and total vends. Percentage changes for Focal and Top 5 products refer to changes in their average inside market share. Percentage change for Total reports the percentage change in the number of total vends. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 4: Impact of Snickers Removal on Downstream Firm Profit, All Sites

Product	Manufacturer	Retail Margin	Difference in Vends	Difference in Profit
Peanut M&Ms	Mars	0.21	118.4*	25.2*
Twix Caramel	Mars	0.21	109.5*	23.3*
Assorted Chocolate	Hershey's / Mars / Other	0.25	63.2*	13.2*
Assorted Energy	Various	0.48	44.3	17.6
Zoo Animal Cracker	Kellogg's	0.52	34.6*	17.6*
Salted Peanuts	Kraft	0.49	28.4	13.2*
Choc Chip Famous Amos	Kellogg's	0.53	21.5*	12.7*
Ruger Vanilla Wafer	Other	0.50	20.6*	10.5
Assorted Candy	Various	0.41	13.7	5.9
Assorted Potato Chips	PepsiCo	0.35	12.8	4.7
Assorted Pretzels	Snyder's / PepsiCo	0.38	9.4	1.6
Raisinets	Other	0.31	6.1	1.9
Cheetos	PepsiCo	0.41	-1.8	-0.7
Grandmas Choc Chip	PepsiCo	0.59	-5.4	-3.4
Doritos	PepsiCo	0.41	-12.3*	-5.0*
Assorted Cookie	Various	0.48	-17.9	-10.3
Skittles	Mars	0.21	-25.1*	-5.2*
Assorted Salty Snack	Various	0.41	-106.7*	-51.6*
Snickers	Mars	0.21	-321.4*	-68.4*
Total		0.38	-8.3	2.6

Notes: Retail margin is reported in cents per unit. Difference in vends is reported in units per week, and difference in revenue is reported in dollars per week. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 5: Impact of All Product Removals on Total Downstream Firm Profit

	Difference in:	Before Rebate		After Rebate	
Product(s) Removed:	Vends	Margin	Profit	Margin	Profit
Snickers	-8.27	0.02	2.65	-0.05	-2.50
Animal Crackers	-28.68	-0.80	-60.40	-0.64	51.20
Doritos	-75.08	-0.48	-53.64	-0.40	-51.44
Peanut M&Ms	54.13	0.47	43.78	0.35	39.79
Cheetos	-122.49	-0.16	-55.06	-0.07	-53.95
Choc Chip Famous Amos	-179.47	-0.34	-89.21	-0.28	-91.40
Cheetos + Doritos	-35.93	-0.21	-12.76	-0.15	-12.68
Snickers + Peanut M&Ms	-218.37	1.21	-15.00	0.89	-42.14

Notes: Difference in margin is reported in cents per unit. Difference in vends is reported in units per week, and difference in profits is reported in dollars per week. No profit difference is outside the 90% Confidence Interval described in section 3.1.1.

Table 6: Impact on Manufacturer Revenue

	Before Rebate						
Product(s) Removed:	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	-5.85	22.37	11.42	**-28.27**	19.31	-28.11	25.94
Animal Crackers	-3.93	**-34.18**	10.89	81.37	33.45	-23.60	2.60
Doritos	-2.65	22.01	15.27	41.37	**-2.27**	-33.52	2.66
Peanut M&Ms	0.65	9.63	10.42	**-49.11**	56.98	-5.74	23.74
Cheetos	0.84	4.03	15.99	13.66	**21.41**	-32.16	-37.92
Choc Chip Famous Amos	-9.55	**-51.41**	19.10	2.96	65.62	-34.80	-35.47
Cheetos + Doritos	12.60	14.63	-1.43	12.48	**-49.71**	3.83	-2.26
Snickers + Peanut M&Ms	45.93	18.19	5.78	**-218.43**	0.99	-9.57	49.09
	After Rebate						
Product(s) Removed:	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	-4.99	20.70	10.37	**-23.82**	17.53	-25.75	26.40
Animal Crackers	-3.42	**-31.33**	10.07	68.24	30.24	-21.71	3.87
Doritos	-2.30	20.46	13.90	34.64	**-1.52**	-30.66	3.60
Peanut M&Ms	0.29	8.95	9.64	**-41.23**	51.39	-5.37	25.10
Cheetos	0.51	3.70	14.52	11.36	**19.41**	-29.53	-35.62
Choc Chip Famous Amos	-7.99	**-47.31**	17.16	2.37	59.00	-31.89	-32.90
Cheetos + Doritos	10.37	13.49	-1.44	10.45	**-44.84**	3.45	-2.39
Snickers + Peanut M&Ms	37.92	16.68	5.59	**-183.38**	0.77	-8.70	47.94

Notes: Revenues to manufacturer are calculated as the wholesale cost paid by Mark Vend to the manufacturer, not including any potential rebate payments. Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15.

Table 7: Predicted Sales During the Snickers Treatment and Matched Control Weeks (All Sites), Nested Logit Model

	Results of Matching (Means)			Fitted Values (Means)		
	Treatment	Control	%Δ	Treatment	Control	%Δ
Twix Caramel	297.1*	187.6	58.4	208.6*	190.7	9.4
Peanut M&Ms	478.3*	359.9	32.9	388.6*	349.1	11.3
Ruger Vanilla Wafer	127.9	107.3	19.1	114.3	111.5	2.5
Assorted Chocolate	398.0*	334.8	18.9	371.0*	328.6	12.9
Zoo Animal Cracker	243.7*	209.1	16.5	206.9	205.2	0.8
Salted Peanuts	216.3*	187.9	15.1	190.7*	184.8	3.2
Choc Chip Famous Amos	193.1*	171.6	12.5	171.4	167.5	2.4
Assorted Energy	616.2	571.9	7.8	582.7	584.3	-0.3
Assorted Candy	229.6	215.8	6.4	253.1*	217.0	16.6
Raisinets	139.4	133.3	4.6	142.8	129.5	10.3
Assorted Potato Chips	292.4*	279.6	4.6	348.4*	290.9	19.8
Assorted Pretzels	557.7*	548.3	1.7	533.2*	522.4	2.1
Cheetos	260.5	262.2	-0.7	271.8	270.1	0.6
Doritos	203.1	215.4	-5.7	229.3	226.5	1.3
Grandmas Choc Chip	72.5	77.9	-7.0	83.7	83.9	-0.3
Assorted Salty Snack	1276.2*	1382.8	-7.7	1356.3	1410.5	-3.8
Assorted Cookie	162.4	180.3	-10.0	162.5	178.4	-8.9
Skittles	75.1*	100.1	-25.0	100.3	93.2	7.6
Snickers	2.0*	323.4	-99.4	4.4*	336.2	-98.7
Total	5841.3	5849.6	-0.1	5719.9	5880.2	-2.7

Notes: Products are sorted by percentage change in actual values. Predicted results report sales during the treatment weeks and control weeks using the parameters estimated in the baseline nested-logit model. Just as for the actual data, predicted vends of the focal product result from the assignment of service visits to weeks. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 8: Predicted Sales During the Snickers Treatment and Matched Control Weeks (All Sites), Random Coefficients Model

	Results of Matching (Means)			Fitted Values (Means)		
	Treatment	Control	%Δ	Treatment	Control	%Δ
Twix Caramel	297.1*	187.6	58.4	206.2*	189.8	8.7
Peanut M&Ms	478.3*	359.9	32.9	382.4*	347.9	9.9
Ruger Vanilla Wafer	127.9	107.3	19.1	120.7	110.2	9.5
Assorted Chocolate	398.0*	334.8	18.9	367.7*	327.6	12.2
Zoo Animal Cracker	243.7*	209.1	16.5	211.8*	204.3	3.7
Salted Peanuts	216.3*	187.9	15.1	193.8*	183.9	5.4
Choc Chip Famous Amos	193.1*	171.6	12.5	178.9*	165.9	7.9
Assorted Energy	616.2	571.9	7.8	604.6	579.6	4.3
Assorted Candy	229.6	215.8	6.4	246.0*	217.9	12.9
Raisinets	139.4	133.3	4.6	138.7	129.4	7.2
Assorted Potato Chips	292.4*	279.6	4.6	349.6*	289.8	20.6
Assorted Pretzels	557.7*	548.3	1.7	525.7*	519.4	1.2
Cheetos	260.5	262.2	-0.7	264.6	266.4	-0.7
Doritos	203.1	215.4	-5.7	229.6	225.5	1.8
Grandmas Choc Chip	72.5	77.9	-7.0	90.5*	83.2	8.7
Assorted Salty Snack	1276.2*	1382.8	-7.7	1353.1	1403.2	-3.6
Assorted Cookie	162.4	180.3	-10.0	169.1	178.5	-5.2
Skittles	75.1*	100.1	-25.0	97.7	92.9	5.1
Snickers	2.0*	323.4	-99.4	4.4*	334.1	-98.7
Total	5841.3	5849.6	-0.1	5735.1	5849.6	-2.0

Notes: Products are sorted by percentage change in actual values. Predicted results report sales during the treatment weeks and control weeks using the parameters estimated in the baseline random-coefficients logit model. Just as for the actual data, predicted vends of the focal product result from the assignment of service visits to weeks. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 9: Impact of All Product Removals on Total Downstream Firm Profit (Predicted, Random Coefficients)

	Difference in:	Before Rebate		After Rebate	
Product(s) Removed:	Vends	Margin	Profit	Margin	Profit
Snickers	-114.46	0.45	-15.82	0.31	-30.43
Animal Crackers	-152.71	-0.85	-109.71	-0.67	-104.85
Doritos	-193.68	-0.48	-98.08	-0.36	-99.34
Peanut M&M's	-82.42	0.69	1.70	0.50	-9.93
Cheetos	-205.94	-0.19	-94.33	-0.19	-96.85
Choc Chip Famous Amos	-243.73	-0.52	-130.73	-0.53	-133.28
Cheetos + Doritos	-187.34	-0.06	-74.43	-0.11	-79.75
Snickers + Peanut M&M's	-369.02	1.98	-33.76	1.58	-73.94

Notes: Difference in margin is reported in cents per unit. Difference in vends is reported in units per week, and difference in profits is reported in dollars per week. No profit difference is outside the 90% Confidence Interval described in section 3.1.1.

Table 10: Impact on Manufacturer Revenue (Predicted, Random Coefficients)

Product(s)	Before Rebate						
Removed:	Hershey's	Kellogg	Kraft	Mars Inc.	PepsiCo	Snyder's	Minor
Snickers	-0.18	10.46	9.33	**-109.08**	41.52	-19.70	2.54
Animal Crackers	-1.19	**-49.44**	6.15	43.09	44.73	-19.02	-27.52
Doritos	-6.39	3.74	13.95	31.28	**-16.68**	-31.36	-26.42
Peanut M&Ms	3.95	9.95	9.48	**-96.56**	49.02	-18.13	12.46
Cheetos	-6.83	-0.65	17.13	17.29	**-10.67**	-31.69	-36.35
Choc Chip Famous Amos	-9.52	**-48.99**	6.01	30.75	13.51	-30.69	-17.96
Cheetos + Doritos	11.60	1.69	6.46	2.01	**-81.60**	2.24	-2.67
Snickers + Peanut M&Ms	26.45	16.53	8.22	**-291.40**	19.39	-10.53	29.21
	After Rebate						
	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	-0.32	9.64	8.39	**-91.68**	37.40	-18.06	3.70
Animal Crackers	-1.22	**-45.49**	5.54	36.14	40.25	-17.50	-25.79
Doritos	-5.37	3.47	12.53	26.25	**-14.82**	-28.66	-24.50
Peanut M&Ms	3.01	9.21	8.64	**-81.17**	44.20	-16.64	13.30
Cheetos	-5.77	-0.61	15.37	14.38	**-9.55**	-29.11	-33.62
Choc Chip Famous Amos	-7.96	**-45.08**	5.36	25.73	12.13	-28.12	-15.65
Cheetos + Doritos	9.49	1.58	5.90	1.66	**-73.48**	2.00	-2.45
Snickers + Peanut M&Ms	21.71	15.26	7.68	**-244.73**	17.42	-9.60	29.10

Notes: Revenues to manufacturer are calculated as the wholesale cost paid by Mark Vend to the manufacturer, not including any potential rebate payments. Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15.

Table 11: Simulated Effects of Removing Snickers (All Sites)

	Change in Sales		Change in Profit	
	Nested Logit	RC Logit	Nested Logit	RC Logit
Peanut M&Ms	39.35*	32.87*	8.37*	6.99*
Assorted Chocolate	34.96*	29.73	10.14	8.47
Assorted Candy	24.20	16.16	10.45*	6.96
Twix Caramel	16.47	16.38	3.50	3.48
Raisinets	10.78	8.14	3.34	2.52
Skittles	10.29	10.51	1.89	1.98
Assorted Salty Snack	10.02	27.67	4.11	11.31
Assorted Energy	4.52	34.35	2.20	16.57
Assorted Pretzels	3.29	7.61	1.24	2.88
Cheetos	1.85	4.80	0.76	1.97
Doritos	1.53	4.65	0.63	1.91
Assorted Potato Chips	1.43	3.70	0.53	1.37
Assorted Cookie	1.41	12.31	0.65	5.59
Zoo Animal Cracker	1.36	8.89	0.69	4.43
Choc Chip Famous Amos	1.12	12.13	0.61	6.53
Grandmas Choc Chip	0.67	12.00	0.39	7.02
Ruger Vanilla Wafer	0.63	8.03	0.34	4.33
Salted Peanuts	0.10	0.42	0.06	0.23
Snickers	-320.00*	-323.42*	-68.06*	-68.79*
Total	-156.02	-73.11	-18.18	25.77

Notes: The left panel reports the change in sales for each product of a simulated removal of Snickers. The simulated profit change of each product for the downstream firm is reported in the right panel. Products with a mean difference outside of the 90% Confidence Interval described in section 3.1.1 are denoted with *'s.

Table 12: Simulated Effects of Stockouts on Retailer Profit (Random Coefficients)

	Difference In:	Before Rebate		After Rebate	
Product(s) Removed:	Vends	Margin	Profit	Margin	Profit
Snickers	-73.11	0.98	25.77	0.71	8.19
Animal Crackers	-106.88	-0.48	-70.42	-0.42	-70.24
Doritos	-104.03	-0.05	-43.07	-0.05	-46.88
Peanut M&Ms	-113.89	1.08	8.48	0.78	-10.39
Cheetos	-72.34	-0.01	-28.74	-0.02	-32.09
Choc Chip Famous Amos	-50.20	-0.48	-43.84	-0.42	-42.93
Cheetos + Doritos	-153.59	-0.04	-61.81	-0.06	-68.11
Snickers + Peanut M&Ms	-189.36	2.24	39.22	1.61	0.78

Notes: Difference in margin is reported in cents per unit. Difference in vends is reported in units, and difference in profits is reported in dollars. No profit difference is outside the 90% Confidence Interval described in section 3.1.1.

Table 13: Simulated Effects of Stockouts on Manufacturer Revenue (Random Coefficients)

	Before Rebate						
	Hershey's	Kellogg	Kraft	Mars	PepsiCo	Snyder's	Minor
Snickers	3.55	12.82	3.74	**-132.67**	13.70	2.31	18.06
Animal Crackers	1.24	**-46.43**	1.87	15.45	7.64	2.31	6.93
Doritos	0.78	4.62	1.71	7.11	**-59.26**	2.46	5.26
Peanut M&Ms	3.31	8.41	3.35	**-135.83**	10.75	1.75	15.54
Cheetos	0.89	8.86	3.05	9.28	**-65.59**	3.68	9.91
Choc Chip Famous Amos	1.24	**-36.23**	1.31	15.44	8.22	1.40	5.61
Cheetos + Doritos	1.50	11.32	4.41	14.04	**-109.23**	5.64	13.40
Snickers + Peanut M&Ms	7.03	23.23	7.27	**-283.18**	25.72	4.45	37.46
	After Rebate						
Snickers	2.91	11.80	3.45	**-111.44**	12.33	2.11	17.91
Animal Crackers	1.02	**-42.72**	1.72	12.98	6.88	2.10	6.84
Doritos	0.64	4.25	1.58	5.98	**-53.33**	2.24	5.14
Peanut M&Ms	2.72	7.74	3.10	**-114.10**	9.67	1.60	15.42
Cheetos	0.73	8.15	2.80	7.80	**-59.03**	3.35	9.62
Choc Chip Famous Amos	1.02	**-33.33**	1.21	12.97	7.40	1.28	5.54
Cheetos + Doritos	1.23	10.42	4.06	11.79	**-98.30**	5.14	13.05
Snickers + Peanut M&Ms	5.78	21.37	6.72	**-237.86**	23.16	4.05	37.20

Notes: Revenues to manufacturer are calculated as the wholesale cost paid by Mark Vend to the manufacturer, not including any potential rebate payments. Revenue impacts for the manufacturer of the focal product of any given experiment are shown in bold typeface (e.g., Mars Inc. manufactures Snickers). Minor manufacturers include all 17 manufacturers listed in table 15.

Table 14: Parameter Estimates and Holdout Analyses

	Base Model	Hold Out Experimental Data from:				
		All Experiments	Candy	Salty Snack	Animal Crackers	Choc Chip F. Amos
Random Coefficients						
σ_{fat}	2.10 (0.091)	2.42	2.39	2.06	2.17	2.14
σ_{salt}	3.49 (0.179)	3.93	4.65	3.21	3.47	3.52
σ_{sugar}	5.25 (0.281)	2.91	2.94	5.44	5.08	5.11
$LL * 10^6$	-5.206	-4.259	-4.806	-4.861	-5.101	-5.111
Nested Logit						
λ_{candy}	0.57 (0.013)	0.66	0.65	0.57	0.58	0.58
λ_{cookie}	0.72 (0.021)	0.74	0.72	0.72	0.73	0.72
λ_{energy}	0.86 (0.015)	0.81	0.84	0.84	0.85	0.86
$\lambda_{saltysnack}$	0.62 (0.020)	0.57	0.62	0.60	0.62	0.61
$LL * 10^6$	-5.206	-4.259	-4.806	-4.861	-5.101	-5.111
Number of Choice Sets						
All (417)	1734	1096	1485	1586	1674	1653
Candy (115)	427	262	265	425	427	427
Cookie (53)	184	122	184	184	152	154
Energy (62)	166	166	166	166	166	166
Salty Snack (187)	794	578	792	582	794	793

Notes: Base model is estimated off the full dataset, including all experimental periods. Candy excludes data from the Snickers, Peanut M&M's, and joint Snickers/Peanut M&M Experiments. Salty Snack excludes data from the Cheetos, Doritos, and joint Doritos/Cheetos Experiments. The third panel reports the number of products in each category in parentheses after the category name. The number of unique choice sets is reported in the row labeled "All." The number of choice sets by category is calculated for the products within the category. Category-level choice sets do not sum to the total number of choice sets because variation in choice sets across categories contributes to the total number of choice sets. Standard errors for the base model estimates are reported in parentheses.

Appendix: Manufacturer Statistics and Matching Detail

Table 15 provides summary statistics by manufacturer. The six major snack manufacturers are listed separately, followed by a column for all minor manufacturers.[63] In the first two panels, we report statistics for each manufacturer's full portfolio of products: the inside-good market share, average daily sales per machine, average number of products per machine, and product counts by category. PepsiCo produces the largest number of products (30), with the largest inside market share (over 33%).[64] The smallest major manufacturer is Kraft, producing 8 products with an inside market share of about 3%. Minor manufacturers account for just over 0.9% of all vends each (15.9% combined), spread across 40 products. The average number of products per machine ranges from 1.6 (out of 8) for Kraft to nearly 11 (out of 30) for PepsiCo. The average daily sales per machine is higher for the major manufacturers than for the combined minor manufacturers, at 1.2 on average. Major manufacturers segment across different product categories, with Mars and Hershey's focusing on chocolate and candy, PepsiCo and Snyder's focusing on salty snacks, and Kellogg and Kraft in the cookie and energy categories.

Maximum, minimum and median retail and wholesale prices are listed in the second panel of table 15. The largest spread in prices exists among the combined minor manufacturers. In the last panel, we report retail and wholesale prices and the average inside good share for the "best," "median," and "worst" products, defined as the products with the highest/median/lowest average daily sales rate in our three-year sample period. The range of inside good shares is very similar for major and minor manufacturers. Neither retail nor wholesale prices are correlated with sales performance.

Table 16 reports the products used in the matching procedure described in section 3.

[63] We classify 17 manufacturers as 'minor' manufacturers based on the availability and sales of their products at the six experimental sites. These are: Barton's Confectioners, Biscomerica, Brother's Kane, California Chips, ConAgra, Farley's & Sathers Candy Company, Frontera Foods, General Mills, Genisoy, Inventure Group, Just Born Inc., Kar's Nuts, Nestle, Procter & Gamble, Sherwood Brands, Snak King, and United Natural Foods.

[64] All Frito-Lay and Quaker brands are owned by PepsiCo.

Table 15: Manufacturer Ownership

	Manufacturer						
	Hershey's	Kellogg	Kraft	Mars Inc.	PepsiCo	Snyder's	Minor
Data on Manufacturer Portfolios							
Avg. Inside Good Share	3.36	12.74	4.76	21.11	33.20	8.97	0.93
Avg. Daily Sales per Machine	0.43	1.66	0.62	2.74	4.32	1.16	0.23
Avg. Products per Machine	1.56	5.22	2.77	5.72	10.48	2.80	1.24
# of Products	8	10	14	18	30	10	40
# Chocolate Products	4	0	0	10	0	0	6
# Candy Products	4	0	3	6	0	0	5
# Cookie Products	0	4	7	0	4	0	3
# Energy Products	0	5	4	2	3	0	8
# Salty Snack Products	0	1	0	0	23	10	18
Price Data for Individual Products							
Price (Maximum)	0.85	1.00	0.85	0.85	0.95	1.00	2.00
Price (Median)	0.75	0.75	0.76	0.75	0.74	0.63	0.78
Price (Minimum)	0.75	0.50	0.60	0.60	0.50	0.45	0.50
Wholesale Price (Maximum)	0.57	0.49	0.40	0.67	0.49	0.77	0.77
Wholesale Price (Median)	0.46	0.28	0.28	0.53	0.33	0.23	0.33
Wholesale Price (Minimum)	0.38	0.15	0.16	0.38	0.15	0.14	0.02
Marketshare and Price Data for Individual Products, Based on Performance							
Avg. Inside Share (Best)	5.39	5.33	5.00	6.63	5.85	6.55	5.34
Avg. Inside Share (Median)	1.76	2.48	1.64	1.98	2.93	3.40	1.76
Avg. Inside Share (Worst)	0.02	0.02	0.01	0.09	0.04	1.72	0.03
Price (Best)	0.85	0.85	0.75	0.75	0.50	0.50	0.75
Price (Median)	0.75	0.85	0.85	0.65	0.75	0.75	0.75
Price (Worst)	0.75	0.75	0.85	0.60	0.75	0.60	0.75
Wholesale Price (Best)	0.49	0.39	0.27	0.54	0.23	0.21	0.44
Wholesale Price (Median)	0.49	0.24	0.32	0.54	0.34	0.32	0.02
Wholesale Price (Worst)	0.38	0.28	0.26	0.38	0.28	0.22	0.47

Notes: All calculations performed using the full dataset from Mark Vend Company at six experimental sites. There are 17 minor manufacturers: Barton's Confectioners, Biscomerica, Brother's Kane, California Chips, ConAgra, Farley's & Sathers Candy Company, Frontera Foods, General Mills, Genisoy, Inventure Group, Just Born Inc., Kar's Nuts, Nestle, Procter & Gamble, Sherwood Brands, Snak King, and United Natural Foods. The second panel, "Price data for individual products," ranks products by the reported variable. The third panel ranks products based on their average inside market share and reports statistics for all variables based on this ranking.

Table 16: List of Products Used for Matching at Each Site

Candy and Choc Chip Famous Amos Experiments	
93	Ass. Pretzels, Cheetos, Doritos, Ass. Potato Chips
5055	Cheetos, Doritos
5655	Doritos, Ass. Potato Chips
6056	Ass. Pretzels, Cheetos, Doritos, Ass. Salty Snack
6263	Ass. Pretzels, Cheetos
7277	Ass. Pretzels, Cheetos, Doritos, Ass. Potato Chips, Ass. Salty Snack
Salty Snack Experiments	
93	Ass. Candy, Ass. Chocolate, Raisinets
5055	Raisinets, Twix
5655	Ass. Candy, Ass. Chocolate, Twix
6056	Ass. Candy, Skittles, Twix
6263	Skittles, Twix
7277	Ass. Candy, Ass. Chocolate, Raisinets
Zoo Animal Cracker Experiment	
93	Ass. Pretzels, Ass. Potato Chips, Ass. Salty Snack
5055	Ass. Pretzels, Ass. Potato Chips
5655	Ass. Pretzels, Ass. Potato Chips
6056	Ass. Pretzels, Ass. Potato Chips, Ass. Salty Snack
6263	Ass. Pretzels, Ass. Potato Chips, Ass. Salty Snack
7277	Ass. Pretzels, Ass. Potato Chips

Notes: Candy experiments are Snickers, Peanut M&Ms, and the joint removal of both. Salty Snack experiments are Doritos, Cheetos, and the joint removal of both.

Market-based Emissions Regulation and the Evolution of Market Structure

Meredith Fowlie, Mar Reguant, and Stephen P. Ryan*

Preliminary and Incomplete – Please Do Not Cite

July 13, 2011

Abstract

We assess the long-run dynamic implications of market-based regulation for mitigating carbon dioxide emissions in the US Portland cement industry. We consider several policy designs, including mechanisms that partially offset the cost of compliance through rebating. Our results highlight two general countervailing market distortions that face regulators of trade-exposed, concentrated industries. First, echoing a point first made by Buchanan (1969), reductions in product market surplus due to market power counteract the social benefits of carbon abatement. Second, import-exposed cement producers face competition from unregulated foreign competitors, leading to emissions "leakage" which offsets domestic emissions reductions. We find that a combination of these forces leads to social welfare *losses* for low social costs of carbon. At higher social costs of carbon, policies with production subsidies are efficient and welfare dominate more standard policy designs.

*UC Berkeley and NBER, Stanford GSB, and MIT and NBER. The authors gratefully acknowledge the support of NSF grant SES-0922401.

1 Introduction

With the passage of the 1990 Amendments to the Clean Air Act, Congress gave the United States Environmental Protection Agency (EPA) a mandate to implement market-based strategies for reducing harmful ambient emissions. Specifically, Title IV of the Amendments encourages the EPA to transition from prescriptive, "command and control" emissions regulations to more decentralized, market-based mechanisms, such as emissions taxes and trading programs.[1] Market-based incentives now play a crucial role in incentivizing emissions abatement among large industrial sources.

Traditionally, economic analysis of market-based emissions regulations focused exclusively on perfectly competitive industries free of pre-existing distortions or other market failures. In this "first-best" context, policy design is relatively straightforward. A Pigouvian tax, or an emissions trading program designed to equate marginal abatement costs with marginal damages, will generally achieve the socially optimal outcome. However, policy makers rarely, if ever, work in this first-best setting. Emissions intensive industries are generally characterized by several imperfections that complicate the design of efficient policy.

First, the majority of emissions regulated under existing and planned emissions regulations come from industries that are highly concentrated.[2] In an imperfectly competitive industry, a first-best emissions policy that completely internalizes external damages will incentivize pollution abatement, but it will also exacerbate the pre-existing distortion associated with the exercise of market power. In a seminal paper, Buchanan (1969) asserts that the implementation of Pigouvian taxes should be limited to "situations of competition" because taxing an emissions externality further restricts already sub-optimal levels of output. In contrast, Oates and Strassman (1984) argue that the case for Pigouvian taxes "is not seriously compromised by likely deviations from competitive behavior" because the welfare gains from pollution control likely dwarf the potential losses from the various imperfections in the economy.

In the context of global pollutants, such as greenhouse gases, a second consideration

[1] The CAAA legislation authorized the use of "economic incentive regulation" for the control of acid rain, the development of cleaner burning gasoline, the reduction of toxic air emissions, and for states to use in controlling carbon monoxide and urban ozone.

[2] Emissions from restructured electricity markets represent the majority of emissions currently targeted by existing cap-and-trade programs in the United States and Europe. Numerous studies provide empirical evidence of the exercise of market power in these industries, such as Borenstein et al. (2002); Joskow and Kahn (2002); Wolfram (1999); Puller (2007); Sweeting (2007); Bushnell et al. (2008). Other emissions intensive industries being targeted by regional emissions trading programs, such as cement and refining, are also highly concentrated.

further complicates the welfare analysis of market-based emissions policy interventions. If emissions regulations apply to only a subset of the sources that contribute to the environmental problem, firms may respond to regulation by substituting production to the unregulated jurisdiction. This "emissions leakage" may substantially offset, or paradoxically even reverse, the reductions in emissions achieved in the regulated sector. Concerns about leakage and adverse competitiveness impacts have led to a series of policy proposals designed to penalize emissions while also providing incentives to mitigate adverse competitiveness impacts.

In this paper, we use the Markov-perfect Nash equilibrium (Maskin and Tirole, 1988; Ericson and Pakes, 1995) dynamic oligopoly framework developed in Ryan (2011) as the foundation for an analysis of market-based regulations limiting industrial emissions. Our approach allows us to assess the welfare implications of a market-based policy intervention in an industrial context characterized by both imperfect competition and exposure to competition from unregulated imports.

This paper analyzes the efficiency and distributional properties of several policies designed to reduce carbon dioxide emissions in the domestic Portland cement industry. For a number of reasons, this industry has been at the center of the debate about domestic climate change policy and international competitiveness. First, the industry is environmentally important: cement is one of the largest manufacturing sources of domestic carbon dioxide emissions (Kapur et al, 2009). Second, carbon regulation could result in major changes to the industry's cost structure; complete internalization of the estimated social cost of carbon would increase average variable operating costs by more than 50 percent.[3] Third, the industry is highly concentrated in regionally-segregated markets, making the industry potentially susceptible to the Buchanan critique. Finally, import penetration in the domestic cement market has exceeded 20 percent in recent years, giving rise to concerns about the potential for emissions leakage (Van Oss, 2003 ENV; USGS Mineral Commodity Summary 2010). For these reasons, the cement industry is an interesting and important setting to study the complex interactions between industrial organization and environmental policy design.

A distinguishing feature of our analysis is our emphasis on industry dynamics. For a number of reasons, a static, short-run analysis is ill-suited to the domestic cement industry. First, capital stock turnover is expected to play an essential role in improving the environmental performance of this industry (Worrell et al., 2001; Sterner, 1990). This is partly due to the limited opportunities to reduce carbon intensity through process changes and

[3]On average, domestic cement producers emit approximately one ton of carbon for each ton of cement produced. Marginal costs of cement production are estimated to be in the range of \$30-\$40/ton (Ryan, 2011).

disembodied capital change, and partly due to fact that some very old and inefficient kilns are still in operation in the United States. It is estimated that replacing these with newer and more efficient technologies could yield emissions reductions in excess of 15 percent (Mahasenan et al., 2005). Second, an exclusive focus on short run outcomes would likely fail to capture the extent of emissions leakage. Although leakage can manifest immediately as firms adjust variable input and output decisions such that less (more) stringently regulated production assets are used more (less) intensively, it can also occur gradually as firms accelerate the retirement of older production technologies in more stringently regulated jurisdictions and invest in new facilities and equipment in less stringently regulated jurisdictions. Static modeling cannot capture this second leakage channel.

Our analysis begins with the specification of a theoretical model of dynamic oligopoly in which strategic domestic cement producers compete in spatially-segregated regional markets. Some of these markets are trade exposed, whereas other landlocked markets are sheltered from foreign competition. Firms make entry, exit, and investment decisions in order to maximize their expected stream of profits conditional on the strategies of their rivals. Given capital investments, producers compete each period in homogeneous quantities. Regional market structures evolve as firms enter, exit, and adjust production capacities in response to changing market conditions.

Building on the parameter estimates from Ryan (2011), we then turn to our investigation of the static and dynamic implications of market-based emissions regulation design decisions. We use the econometrically estimated model to simulate industry response to a series of counterfactual emissions regulations. The basic intuition underlying our counterfactual simulations is quite simple. In the benchmark model that we estimate, emissions are unconstrained. Firms invest at the level where marginal costs equal expected marginal benefits subject to covering their fixed costs. The expected benefits are a function of the period payoffs, as firms with larger capacities are able to compete over a larger segment of the market. The market-based emissions regulations we consider affect firms' production and investment choices through changes in operating cost and revenue incentives. Importantly, we assume that cement producers' past response to changes in operating costs and revenues mimics what we would observe in response to policy-induced changes.

In addition to more standard carbon tax and emissions trading programs, we are interested in analyzing policy designs that incorporate both an emissions penalty (i.e. an obligation to pay a tax or hold a permit to offset emissions) and a production incentive in the form of a rebate. Under an emissions tax regime, tax revenues can be recycled (or rebated) to producers on the basis of lagged production. In the context of cap-and-trade programs, dynamic permit allocation updating schemes make future free permit allocations contingent on a firm's output or emissions shares in the previous period. In a first-best setting, these contingent rebates would undermine the efficiency of permit market outcomes because the implicit subsidy conferred by allocation updating encourages firms to increase output to economically inefficient levels (Bohringer and Lange, 2005; Sterner and Muller, 2008).[4] However, in second-best settings, these rebates can be used to mitigate pre-existing distortions and regulatory imperfections.

Given the uncertainty surrounding estimates of the social cost of carbon, we simulate outcomes over a range of carbon dioxide (CO_2) damages. We follow the lead of a landmark interagency process which recommends a range of social cost of carbon (SCC) values for use in policy analysis (Greenstone et al., 2011).[5] In this working paper, we simulate outcomes for approximately half of the regional markets that comprise the industry. Future versions of the paper will include all domestic cement markets.

We find that the imposition of a carbon tax or emissions trading program that fully internalizes the social cost of carbon could have negative welfare impacts for SCC values at or below the central SCC value of $21/ton. Two primary market forces drive theis result. The first intuition follows Buchanan's insights with regards to balancing distortions from market power against those induced by pollution externalities; the US Portland cement industry is highly concentrated. The second contributing factor stems from the incompleteness of the emissions regulation which creates the potential for emissions leakage.

As the assumed value of the negative emissions externality increases, the benefits from the emissions regulation (in the form of avoided damages from emissions) exceeds the costs, emissions leakage and the constriction of economic surplus notwithstanding. Notably, policy designs that couple a carbon tax with a production subsidy (in the form of a tax rebate or contingent permit allocation) welfare dominate more standard designs. The rebate works to mitigate leakage in trade exposed cement markets and the distortion associated with the exercise of market power.

This paper makes substantive contributions to three areas of the literature. First, this paper is germane to the literature that considers the dynamic efficiency properties of market-

[4]Here, "first-best" refers to a regulatory environment in which the only market distortion or imperfection is the environmental externality that the emissions regulation is designed to internalize.

[5]The U.S. Government recently concluded a year-long process to estimate the monetized damages caused per ton of CO_2 emissions. For 2010, the central social cost of carbon (SCC) estimate is $21, although sanctioned estimates range from approximately $5 to $65.

based emissions regulations. By their very nature, long-run policy effects are very difficult to identify empirically. During the time it takes for policy outcomes to manifest, a host of other potentially confounding factors and processes change and evolve. The conventional approach to analyzing these long run relationships has been to use either highly stylized theoretical models (Conrad and Wang, 2003; Lee, 1999; Requate, 2005; Sengupta, 2010; Shafter, 1999) or large, deterministic, optimization-based simulation models (Jensen and Rasmussen, 2000; Fischer and Fox, 2007; Szaboe et al, 2006; US EPA, 1996).[6] In a recent review of the literature, Millimet et al. (2009) suggest that the failure to bring the rich literature on dynamic industry models to bear on analyses of long-term consequences of environmental regulation constitutes "the most striking gap in the literature on environmental regulation." This paper starts to fill that gap.

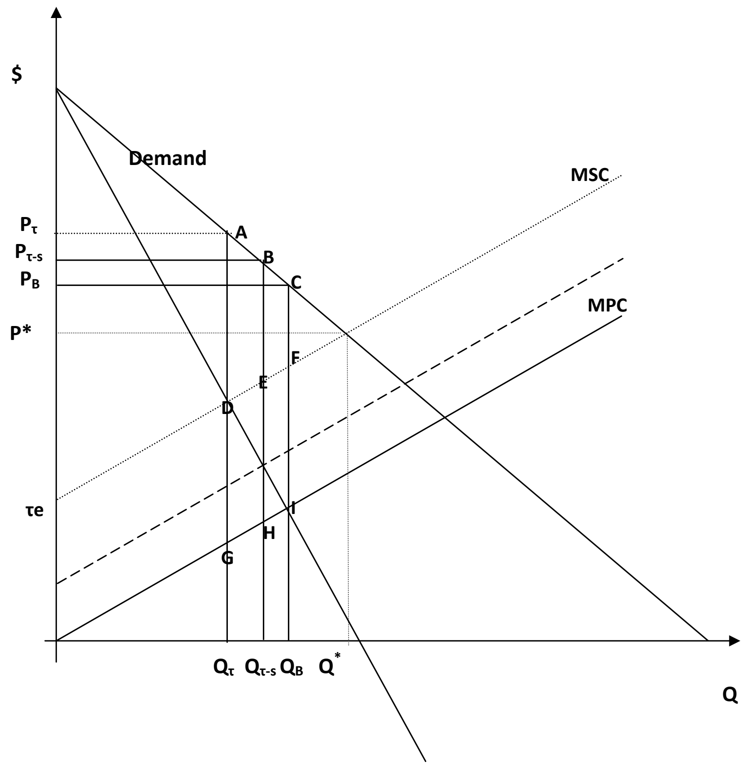
Second, we are not aware of any other paper that investigates the impacts of market-based emissions regulations in the domestic cement industry. This industry has an important role to play in efforts to reduce industrial CO_2 emissions. Ponssard and Thomas (2010) provides some indirect evidence to suggest that unilateral climate change policy would negatively impact investment in the domestic cement industry, thus amplifying the short run production impacts captured by static modeling approaches. In this paper, we investigate this dynamic industrial response in detail.

Finally, the paper makes an important methodological contribution in its application of parametric value function methods to a dynamic game. We make use of interpolation techniques to compute the equilibrium of the counterfactual simulations. This allows us to treat the capacity of the firms as a continuous state. Even though parametric methods have been used in single agent problems, its application to dynamic industry models with discrete entry, exit and investment decisions have not been very successful to date (Doraszelski and Pakes, 2007).

The paper is organized as follows. Section 2 introduces the conceptual framework for our applied policy analysi. Section 3 provides some essential background on the US Portland cement industry. We introduce the model and a detailed description of the alternative policy designs we consider in Section 4. We present the estimation and computational methodology in Section 5. The counterfactual simulations are introduced in Section 6. Simulation results are summarized in Section 7. We conclude with a discussion of the results and directions for

[6]One limitation of these numerical simulation models is that they must rely on the extant econometric literature to provide "off-the-shelf" estimates of important structural parameters (such as the fixed costs of entry or the elasticity of import supply). It is often the case that the econometric literature is not up to the task; models are often parameterized using outdated values or educated guesses.

Figure 1: Emissions-intensive Monopoly: Static case



future research in Section 8.

2 Market-based emissions regulation in a second-best setting

A simple conceptual framework helps to lay the foundation for the applied welfare analysis that is the central focus of the paper. Figure 1 illustrates, among other things, the static welfare consequences of an emissions externality in an industry that is monopolized by a single producer.

The curve labeled MPC measures the marginal private costs of production (i.e. fuel costs, labor costs, etc.) net of any environmental compliance costs. Absent any emissions regulation, this monopolist will produce output Q_B and receive a price P_B. This is the baseline (B) against which we will compare the alternative policy outcomes.

Production generates harmful emissions. We assume a constant emissions rate per unit of output (e) and a constant marginal social cost of emissions τ. The curve labeled MSC captures both private marginal costs and the monetized value of the damages from the firm's emissions: $MSC = MPC + \tau e$. The social welfare maximizing level of output is Q^*. The corresponding price is P^*.

We first consider a case in which the monopolist is required to pay a Pigouvian tax of τ per unit of emissions. This increases the monopolist's variable operating costs by τe. The monopolist will choose to produce Q_τ. The equilibrium price is P_τ.

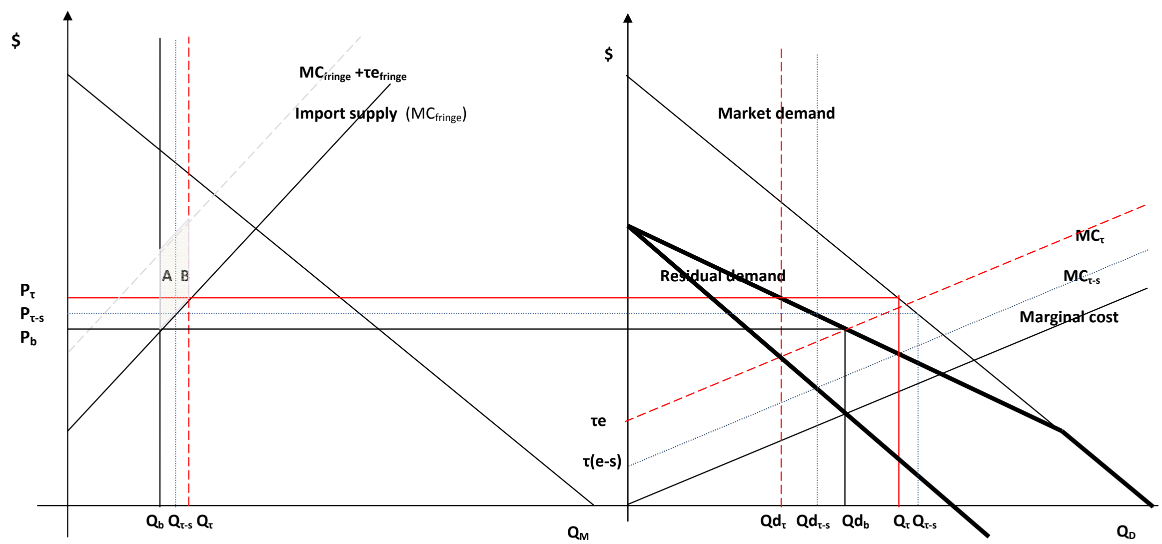
Alternatively, consider an emissons trading program in which permits are auctioned off to the highest bidder or freely distributed in lump-sum to regulated sources based on predetermined, firm-specific characteristics (i.e. "grandfathered"). If the monopolist is sufficiently small relative to the larger emissions trading program, changes in monopolist's net supply or demand for permits will not affect the equilibrium permit price. Within our framework, a large scale emissions trading program with an equilibrium permit price of τ is functionally equivalent to the emissions tax described in the previous paragraph.

In Figure 1, these market-based emissions regulations will *reduce* welfare because the costs associated with further restricting already sub-optimal levels of output outweigh the benefits associated with emissions abatement. This need not always be the case. If the social cost per unit of emissions is sufficiently large, the benefits from full internalization of the emissions externality will offset the costs associated with reductions in output.

If the emissions regulation were coupled with a production subsidy equal to the difference between marginal cost and marginal revenue at the socially optimal level of output, the efficient outcome could achieved. Traditionally, it has been assumed that environmental regulators do not have the authority to subsidize the production of the industries they regulate (Cropper and Oates, 1992). However, policy makers have started to experiment with rebating tax revenues (in the case of an emissions tax) or allocating emissions permits (in the case of a cap-and-trade program) on the basis of production. These contingent rebates affect marginal production incentives, and can thus be used to mitigate—or eliminate—the distortion introduced by the exercise of market power.

The equilibrium outcome under a market-based emissions regulation that incorporates an output-based rebate (or subsidy) s is denoted $\tau - s$ in Figure 1. The monopolist's profit maximizing choice of output under contingent rebating is $Q_{\tau-s}$.In this case, the subsidy does not achieve the first best outcome, although it does mitigate the negative welfare impact of the policy.

Figure 2: Emissions Intensive, Trade Exposed Monopoly



The policy setting we are concerned with is characterized by both imperfect competition and incomplete emissions regulation. Figure 1 captures only the first consideration. A simple extension of this graphical analysis serves to demonstrate the potential implications of incomplete emissions regulation. In Figure 2, the domestic, emissions-intensive monopolist is exposed to competition from producers in jurisdictions that are exempt from the emissions regulation. In the right panel, the thick line represents the residual demand curve (i.e. market demand less import supply) faced by the monopolist. The left panel depicts import supply which is modeled as a competitive fringe.

In the absence of any regulation, import supply is given by q_0^m. The equilibrium output price is P_b. The introduction of market-based emissions regulation increases the operating costs of the monopolist vis a vis its import competition. In the case of an emissions tax or a cap-and-trade program with no rebating, import market share increases to q_τ^m and the difference ($q_0^m - q_\tau^m$) represents leakage in production. Rebating permits or tax revenues to the monopolist based on output reduces this leakage by ($q_\tau^m - q_{\tau-s}^m$).

2.1 Welfare decomposition

Expositionally, it will be useful to decompose the net welfare effects of the emissions policy interventions we analyze into three parts:

1. *Changes in economic surplus.* The first part is comprised of producer and consumer surplus plus any tax revenues or auction revenues earned through the government sale of emissions permits. In Figure 1, the introduction of a carbon tax or an emissions trading program that incorporates auctioning or grandfathering reduces producer and consumer surplus by area $ACIG$. Under a carbon tax or auctioning regime, area $DFIG$ are transferred from producers to the government as auction or tax revenues. Contingent rebating reduces the reduction in consumer and producer surplus by an amount equal to area $ABGH$. Thus, the rebate serves to partially mitigate the distortion associated with the exercise of market power.

2. *Changes in damages from emissions.* An emissions tax or cap-and-trade program reduces economic surplus in the product market, but also reduces damages associated with industrial emissions. Market-based emissions regulations with no rebating reduce emissions damages by an amount equal to area $DFIG$ in Figure 1. Under a tax regime, the introduction of the rebate increases damages from emissions by area $DEHG$.

   Under a cap-and-trade program, the introduction of the rebate does not increase emissions in aggregate because emissions are constrained to equal the cap (assuming the cap binds). However, the introduction of the rebate increases emissions in this monopolized industry, thus shifting more of the compliance burden to other industries and sources subject to the cap. We assume a constant permit price, equivalent to assuming that the abatement supply curve facing the monopolist is locally flat. The additional abatement costs which must be incurred outside this industry in order to offset the emissions increase is area $DEHG$.

3. *Emissions leakage.* If the introduction of an emissions regulation increases production—and thus emissions—among producers in unregulated jurisdictions, this emissions "leakage" will offset some of the emissions reductions achieved among regulated sources. In Figure 2, the shaded parallelogram (area A+B) denotes the monetary cost of this leakage under the market-based regulation that does not incorporate rebating. This cost is reduced to area A under rebating.

Of course, the domestic cement industry is considerably more complex than the stylized cases depicted in Figures 1 and 2. First, regional cement markets are served by more than one firm. Much of the intuition underlying the simple static monopoly case should apply in the case of a static oligopoly (Ebert, 1992). However, the oligopoly response to market-based emissions regulation can be more nuanced in certain situations.[7]

We are particularly interested in how market-based emissions regulations affect welfare via industry dynamics which are not represented in the analytical framework introduced above. Over a longer time frame, firms can alter their choice of production scale, technology, entry, exit, or investment behavior in response to an environmental policy intervention. An important objective of the paper is to explicitly capture the implications of these dynamic industry responses.

The welfare impacts of a market-based emissions policy can look quite different across otherwise similar static and dynamic modeling frameworks. On the one hand, incorporating industry dynamics into the simulation model can improve the projected welfare impacts of a given emissions regulation. Intuitively, the short run economic costs of meeting an emissions constraint can be significantly reduced once firms are able to re-optimize production processes, adjust investments in capital stock, and so forth.

On the other hand, incorporating industry dynamics may result in estimated welfare impacts that are strictly smaller than those generated using static models. In the policy context we consider, there are two primary reasons why this can be the case. First, in an imperfectly competitive industry, emissions regulation may further restrict already sub-optimal levels of investment, thus exacerbating the distortion associated with the exercise of market power. Second, a dynamic model captures an additional channel of emissions leakage. In a static model, firms may adjust variable input and output decisions such that less stringently regulated production assets are used more intensively. This leads to emissions leakage in the short run. In our dynamic modeling framework, the emissions regulation can also accelerate exit and retirement of regulated production units. This further increases the market share claimed by unregulated imports, thus increasing the extent of the emissions leakage to unregulated jurisdictions or entities.

3 The Portland cement industry

Portland cement is an inorganic, non-metallic substance with important hydraulic binding properties. It is the primary ingredient in concrete, an essential construction material used widely in building and highway construction. Demand for cement comes primarily from the ready-mix concrete industry, which accounts of over 70 percent of cement sales. Other major

[7] For example, if firms are highly asymmetric and the inverse demand function has an extreme curvature, it is possible (in theory) for the optimal tax rate to exceed marginal damage (Levin, 1985).

consumers include concrete product manufacturers and government contractors.

Because of its critical role in construction, demand for cement tends to reflect population, urbanization, economic trends, and local conditions in the cement industry. Cement competes in the construction sector with substitutes such as asphalt, clay brick, rammed earth, fiberglass, steel, stone, and wood (Van Oss, 2003, ENV). Another important class of substitutes are the so called supplementary cementitious materials (SCMs) such as ferrous slag, fly ash, silica fume and pozzolana (a reactive volcanic ash). Concrete manufacturers can use these materials as partial substitutes for clinker.[8]

The US cement industry is fragmented into regional markets. This fragmentation is primarily due to transportation economies. The primary ingredient in cement production, limestone, is ubiquitous and costly to transport. To minimize input transportation costs, cement plants are generally located close to limestone quarries. Land transport of cement over long distances is also not economical because the commodity is difficult to store (cement pulls water out of the air over time) and has a very low value to weight ratio. It is estimated that 75 percent of domestically produced cement is shipped less than 110 miles (Miller and Osborne, 2010).[9]

3.1 Carbon dioxide emissions from cement production

Cement producers are among the largest industrial emitters of airborne pollutants, second only to power plants in terms of the criteria pollutants currently regulated under existing cap-and-trade programs (i.e. NOx and SO_2). The cement industry is also one of the largest manufacturing sources of domestic carbon dioxide emissions (Kapur et al, 2009). Worldwide, the cement industry is responsible for approximately 7 percent of anthropogenic CO2 emissions (Van Oss, 2003, ENV).

Cement production process involves two main steps: the manufacture of clinker (i.e. pyroprocessing) and the grinding of clinker to produce cement. Carbon dioxide emissions from cement manufacturing are generated almost exclusively in the pyroprocessing stage. A fuel mix comprised of limestone and supplementary materials is fed into a large kiln lined with refractory brick. The heating of the kiln is very energy intensive (temperatures

[8]The substitition of SCM for clinker can actually improve the quality and strength of concrete. Substitution rates range from 5 percent in standard portland cement to as high as 70 percent in slag cement. These blending decisions are typically made by concrete producers and are typically based on the availability of SCM and associated procurement costs (Van Oss, 2005, facts; Kapur et al, 2009).

[9]Most cement is shipped by truck to ready-mix concrete operations or construction sites in accordance with negotiated contracts. A much smaller percent is transported by train or barge to terminals and then distributed.

reach temperatures of 1450°C) and carbon intensive (because the primary kiln fuel is coal). Carbon dioxide is released as a byproduct of the chemical process that transforms limestone to clinker. Once cooled, clinker is mixed with gypsum and ground into a fine powder to produce cement.[10] Trace amounts of carbon dioxide are released during the grinding phase.

Carbon dioxide emissions intensities, typically measured in terms of metric tons of emissions per metric ton of clinker, vary considerably across cement producers. Much of the variation is driven by variation in fuel efficiency. The oldest and least fuel efficient kilns are "wet-process" kilns. As of 2006, there were 47 of these wet kilns in operation (all built before 1975) (PCA PIS, 2006). "Dry process" kilns are significantly more fuel efficient, primarily because the feed material used has a lower moisture content and thus requires less energy to dry and heat. The most modern kilns, dry kilns equipped with pre-heaters and pre-calciners, are more than twice as fuel efficient as the older wet-process kilns.

Because plants with different emissions intensities will respond differently to the policy interventions we analyze, it is important to capture this variation as accurately as possible. Although data limitations prevent us from estimating emissions intensities specific to each kiln in the data set, we can estimate technology-specific emissions rates. Both the IPCC and the World Business Council for Sustainable Development's Cement Sustainability Initiative (WBC, 2005) have developed protocols for estimating emissions from clinker production. We use these protocols to generate technology-specific estimates of carbon dioxide emissions rates. Appendix A explains these emissions rate calculations in more detail.

There have been several recent studies commissioned to assess the potential for carbon emissions reductions in the cement sector.[11] Using different scenarios, baseline emissions and future demand forecasts, all reach broadly similar conclusions. Although there is no one "silver bullet" on the horizon, there are four key levers for carbon emissions reductions. We summarize these here. We postpone the discussion of how these abatement options are captured by our modeling framework to section 4.

The first set of strategies involve energy efficiency improvements. The carbon intensity of clinker production can by replacing older equipment with current state of the art technologies. In the United States, it is estimated that converting wet installed capacity to dry kilns could reduce annual emissions by approximately 15 percent. Converting from wet to the semi-wet

[10]The US cement industry is comprised of clinker plants (kiln only operations), grinding-only facilities, and integrated (kiln and grinding) facilities.Almost all of the raw materials and energy used in the manufacture of cement are consumed during pyroprocessing. We exempt grinding only facilities from our analysis.

[11]A comprehensive list of studies can be found at http://www.wbcsdcement.org/pdf/technology/References%20FINAL.pdf

process would deliver an additional 3 percent reduction (Mahasenan et al., 2005).

A second set of carbon mitigation strategies involve substitution. One approach is to simply increase the use of substitute construction materials such as wood or brick, thus reducing demand for cement. Alternatively, the amount of clinker needed to produce a given amount of cement can be reduced by the use of supplementary cementitious materials (SCM) such as coal fly ash, slag, and natural pozzolans.[12] It is estimated that the increased use of blended cement could feasibly reduce carbon emissions by a third over the time frame we consider (Mahasenan et al., 2005).

Fuel switching offers a third emissions abatement strategy. Less carbon intensive fuels, such as waste derived fuels or natural gas, could replace coal as the primary kiln fuel. The potential for CO_2 mitigation by fuel switching to lower carbon fuels and fuels qualifying for emissions offsets in North America has been estimated to be on the order of 5 percent of current emissions (Humphreys and Mahasenan, 2001).

Finally, carbon dioxide emissions can be separated or captured during or after the production process and subsequently sequestered. This abatement option is unlikely to play a significant role in the near term given that sequestration technologies are in an early stage of technical development or acceptance and are relatively costly.

3.2 Trade Exposure

Whereas overland transport of cement is very costly, sea-based transport of clinker is relatively inexpensive. In the 1970s, technological advances made it possibly to transport cement in bulk qantities safely and cheaply in large ocean vessels. Since that time, U.S. imports have been growing steadily. The United States now absorbs approximately one quarter of the total global cement trade (Van Oss, 2003 ENV). In the recent past, import penetration rates have averaged around 20 percent (USGS Mineral Commodity Summary 2010). China is currently the largest supplier of imported cement (accounting for 22 percent of imports), followed by Canada, Korea, and Thailand (USGS, 2010 fact sheet).

Exposure to import competition in regional markets has given rise to growing concerns about unilateral climate policy. For example, an industy trade group has warned that, in the absence of measures that either relieve the initial cost pressure or impose equivalent costs of imports, California's proposed cap on greenhouse gas emissions will "render the

[12] When part of the cement content of concrete is replaced with supplementary cementitious materials, the extent of the emissions reduction is proportional to the extent to which SCM replaces clinker. Substitution rates as high as 75 percent are possible.

California cement industry economically unviable, will result in a massive shift in market share towards imports in the short run, and will precipitate sustained disinvestment in the California cement industry in the long run."[13]

4 Model

4.1 Baseline model

The basic building block of the model is a regional cement market.[14] Each market is fully described by the $\overline{N} \times 2$ state vector, s_t, where s_{it} describes the productive capacity of the i-th firm at time t and its associated emissions rate. We set $\overline{N}$ to be the maximal number of firms. Firms with zero capacity are considered to be potential entrants. Time is discrete and unbounded. Firms discount the future at rate $\beta = 0.9$.

Each decision period is one year. In each period, the sequence of events unfolds as follows: first, incumbent firms receive a private draw from the distribution of scrap values, and decide whether or not to exit the industry. Potential entrants receive a private draw from the distribution of both investment and entry costs, while incumbents who have decided not to exit receive private draws on the fixed costs of investment and divestment. All firms then simultaneously make entry and investment decisions. Third, incumbent firms compete over quantities in the product market. Finally, firms enter and exit, and investments mature. We assume that firms who decide to exit produce in this period before leaving the market, and that adjustments in capacity take one period to realize. We also assume that each firm operates independently across markets.[15]

Firms obtain revenues from the product market and incur costs from production, entry, exit, and investment. Firms compete in quantities in a homogeneous goods product market. Firms in trade-exposed regional markets face an import supply curve:

$$\ln M_m = \rho_0 + \rho_1 \ln P_m, \tag{1}$$

where M_m measures annual import supply in market m and ρ_1 is the elasticity of import supply. Here we assume that the elasticity of import supply is an exogenously determined

[13] Letter from the Coalition for Sustainable Cement Manufacturing and Environment to Larry Goulder, Chair of the Economic and Allocation Advisory Committee. Dec. 19, 2009.

[14] This section borrows heavily from Ryan (2011).

[15] This assumption explicitly rules out more general behavior, such as multimarket contact as considered in Bernheim and Whinston (1990) and Jans and Rosenbaum (1997).

parameter.[16] In future work, we hope to explore the potential implications of the strategic use of imports by dominant market players.

After netting out imports, firms face a constant elasticity residual demand curve:

$$\ln Q_m(\alpha) = \alpha_{0m} + \alpha_1 \ln P_m, \tag{2}$$

where Q_m is the aggregate market quantity, P_m is price, α_{0m} is a market-specific intercept, and α_1 is the elasticity of demand. For clarity, we omit the m subscript in what follows.

There are essentially five variable inputs used in cement production: labor, fuel (primarily coal), electricity, feedstocks, and maintenance. These factor inputs are not substitutable (Das, 1994). The majority of variable operating costs are energy related. Because frequent heating and cooling damages the firebrick lining, kilns typically operate continuously at full capacity for 24 hours a day. Annual output is adjusted by varying the length of time the kiln is shut down for annual maintenance. In the model, each firm chooses the level of annual output that maximizes their static profits given the outputs of the competitors, subject to capacity constraints that are determined by dynamic capacity investment decisions:

$$\max_{q_i} \; P\left(q_i + \sum_{j \neq i} q_j; \alpha\right) q_i - C_i(q_i; \delta) - \varphi(q_i, e_i, \tau), \tag{3}$$

where $P(Q;\alpha)$ is the inverse of Equation 2. In the presence of fixed operation costs the product market may have multiple equilibria, as some firms may prefer to not operate given the outputs of their competitors. However, if all firms produce positive quantities then the equilibrium vector of production is unique, as the best-response curves are downward-sloping.

We will use this model to evaluate the impacts of alternative approaches to allocating emissions permits in an emissions trading program. Firm-specific compliance costs will be determined by kiln-specific emissions rates, e_i, production quantity, and the number of permits the firm receives free of charge. While postponing the discussion of the policy designs we consider until Section 4.2, we note here that the introduction of a tax or emissions trading program modifies the profit function in Equation 3 through the term $\varphi(q_i, e_i, \tau)$, where τ

[16] In fact, firms that own a majority of the domestic production capacity in the United States are also among the largest importers. These dominant producers presumably use imports to supplement their domestic production as needed, and to compete in markets where they do not own production facilities. Domestic cement producers have noted that increased domestic ownership of import facilities has contributed to a "more orderly flow of imports into the U.S."

Grancher, Roy A. "U.S. Cement: Record Performance and Reinvestment", Cement Americas, Jul 1, 1999 is the price paid to offset one metric ton of carbon dioxide. The precise nature of the modification will vary across policy designs.

The cost of output, q_i, is given by the following function:

$$C_i(q_i;\delta) = \delta_1 q_i + \delta_2 1(q_i > \nu s_i)(q_i - \nu s_i)^2. \tag{4}$$

Variable production costs consist of two parts: a constant marginal cost, δ_1, and an increasing function that binds as quantity approaches the capacity constraint. We assume that costs increase as the square of the percentage of capacity utilization, and parameterize both the penalty, δ_2, and the threshold at which the costs bind, ν. This second term, which gives the cost function a "hockey stick" shape common in the electricity generation industry, accounts for the increasing costs associated with operating near maximum capacity, as firms have to cut into maintenance time in order to expand production beyond utilization level ν. We denote the profits accruing from the product market by $\bar{\pi}_i(s;\alpha,\delta)$.

Firms can change their capacity through costly adjustments, denoted by x_i. The cost function associated with these activities is given by:

$$\Gamma(x_i;\gamma) = 1(x_i > 0)(\gamma_{i1} + \gamma_2 x_i + \gamma_3 x_i^2) + 1(x_i < 0)(\gamma_{i4} + \gamma_5 x_i + \gamma_6 x_i^2). \tag{5}$$

Firms face both fixed and variable adjustment costs that vary separately for positive and negative changes. Fixed costs capture the idea that firms may have to face significant setup costs, such as obtaining permits or constructing support facilities, that accrue regardless of the size of the kiln. Fixed positive investment costs are drawn each period from the common distribution F_γ, which is distributed normally with mean μ_γ^+ and standard deviation σ_γ^+, and are private information to the firm. Divestment sunk costs may be positive as the firm may encounter costs in order to shut down the kiln and dispose of related materials and components. On the other hand, firms may receive revenues from selling off their infrastructure, either directly to other firms or as scrap metal. These costs are also private information, and are drawn each period from the common distribution G_γ, which is distributed normally with mean μ_γ^- and standard deviation σ_γ^-.

Firms face fixed costs unrelated to production, given by $\Phi_i(a)$, which vary depending on their current status and chosen action, a_i:

$$\Phi_i(a_i;\kappa_i,\phi_i) = \begin{cases} -\kappa_i & \text{if the firm is a new entrant,} \\ \phi_i & \text{if the firm exits the market.} \end{cases} \tag{6}$$

Firms that enter the market pay a fixed cost of entry, κ_i, which is private information and drawn from the common distribution of entry costs, F_κ. Firms exiting the market receive a payment of ϕ_i, which represents net proceeds from shuttering a plant, such as selling off the land and paying for an environmental cleanup. This value may be positive or negative, depending on the magnitude of these opposing payments. The scrap value is private information, drawn anew each period from the common distribution, F_ϕ. Denote the activation status of the firm in the next period as χ_i, where $\chi_i = 1$ if the firm will be active next period, whether as a new entrant or a continuing incumbent, and $\chi_i = 0$ otherwise. All of the shocks that firms receive each period are mutually independent.

Collecting the costs and revenues from a firm's various activities, the per-period payoff function is:

$$\pi_i(s, a; \alpha, \rho, \delta, \gamma_i, \kappa_i, \phi_i) = \bar{\pi}_i(s; \alpha, \rho, \delta) - \Gamma(x_i; \gamma_i) + \Phi_i(a_i; \kappa_i, \phi_i). \tag{7}$$

For the sake of brevity, we henceforth denote the vector of parameters in Equation 7 by θ.

4.1.1 Transitions Between States

To close the model it is necessary to specify how transitions occur between states as firms engage in investment, entry, and exit. We assume that changes to the state vector through entry, exit, and investment take one period to occur and are deterministic. The first part is a standard assumption in discrete time models, and is intended to capture the idea that it takes time to make changes to physical infrastructure of a cement plant. The second part abstracts away from depreciation, which does not appear to be a significant concern in the cement industry, and uncertainty in the time to build new capacity.[17]

We also assume that the emissions rate of the firm is fixed. We assume that there are three discrete levels of emissions rates, corresponding to the three major types of production technology in the cement industry. Existing incumbents are modeled as having one of the three technologies, while new entrants are always endowed with the frontier technology. As a result, the emissions profile of an industry changes over time in response to firm turnover.

[17]It is conceptually straightforward to add uncertainty over time-to-build in the model, but assuming deterministic transitions greatly reduces the computational complexity of solving for the model's equilibrium.

4.1.2 Equilibrium

In each time period, firm i makes entry, exit, production, and investment decisions, collectively denoted by a_i. Since the full set of dynamic Nash equilibria is unbounded and complex, we restrict the firms' strategies to be anonymous, symmetric, and Markovian, meaning firms only condition on the current state vector and their private shocks when making decisions, as in Maskin and Tirole (1988) and Ericson and Pakes (1995).

Each firm's strategy, $\sigma_i(s, \epsilon_i)$, is a mapping from states and shocks to actions:

$$\sigma_i : (s, \epsilon_i) \to a_i, \tag{8}$$

where ϵ_i represents the firm's private information about the cost of entry, exit, investment, and divestment. In the context of the present model, $\sigma_i(s)$ is a set of policy functions which describes a firm's production, investment, entry, and exit behavior as a function of the present state vector. In a Markovian setting, with an infinite horizon, bounded payoffs, and a discount factor less than unity, the value function for an incumbent at the time of the exit decision is:

$$\begin{aligned} V_i(s; \sigma(s), \theta, \epsilon_i) = \bar{\pi}_i(s; \theta) + \max \Bigg\{ \phi_i, E_{\epsilon_i} \Bigg\{ \beta \int E_{\epsilon_i} V_i(s'; \sigma(s'), \theta, \epsilon_i) \; dP(s'; s, \sigma(s)) \\ + \max_{x_i^* > 0} \left[-\gamma_{i1} - \gamma_2 x_i^* - \gamma_3 x_i^{*2} + \beta \int E_{\epsilon_i} V_i(s'; \sigma(s'), \theta, \epsilon_i) \; dP(s_i + x^*, s'_{-i}; s, \sigma(s)) \right], \\ \max_{x_i^* < 0} \left[-\gamma_{i4} - \gamma_5 x_i^* - \gamma_6 x_i^{*2} + \beta \int E_{\epsilon_i} V_i(s'; \sigma(s'), \theta, \epsilon_i) \; dP(s_i + x^*, s'_{-i}; s, \sigma(s)) \right] \Bigg\} \Bigg\}, \end{aligned} \tag{9}$$

where θ is the vector of payoff-relevant parameters, E_{ϵ_i} is the expectation with respect to the distributions of shocks, and $P(s'; \sigma(s), s)$ is the conditional probability distribution over future state s', given the current state, s, and the vector of strategies, $\sigma(s)$.

Potential entrants must weigh the benefits of entering at an optimally-chosen level of capacity against their draws of investment and entry costs. Firms only enter when the sum of these draws is sufficiently low. We assume that potential entrants are short-lived; if they do not enter in this period they disappear and take a payoff of zero forever, never entering in the future.[18] Potential entrants are also restricted to make positive investments; firms cannot "enter" the market at zero capacity and wait for a sufficiently low draw of investment costs

[18] This assumption is for computational convenience, as otherwise one would have to solve an optimal waiting problem for the potential entrants. See Ryan and Tucker (2010) for an example of such an optimal waiting problem.

before building a plant. The value function for potential entrants is:

$$V_i^e(s;\sigma(s),\theta,\epsilon_i) = \max\{0,$$
$$\max_{x_i^*>0}\left[-\gamma_{1i}-\gamma_2 x_i^*-\gamma_3 x_i^{*2}+\beta\int E_{\epsilon_i}V_i(s';\sigma(s'),\theta,\epsilon_i)dP(s_i+x^*,s'_{-i};s,\sigma(s))\right]-\kappa_i\bigg\}. \quad (10)$$

Markov perfect Nash equilibrium (MPNE) requires each firm's strategy profile to be optimal given the strategy profiles of its competitors:

$$V_i(s;\sigma_i^*(s),\sigma_{-i}(s),\theta,\epsilon_i) \geq V_i(s;\tilde{\sigma}_i(s),\sigma_{-i}(s),\theta,\epsilon_i), \quad (11)$$

for all s, ϵ_i, and all possible alternative strategies, $\tilde{\sigma}_i(s)$. As we work with the expected value functions below, we note that the MPNE requirement also holds after integrating out firms' private information: $E_{\epsilon_i}V_i(s;\sigma_i^*(s),\sigma_{-i}(s),\theta,\epsilon_i) \geq E_{\epsilon_i}V_i(s;\tilde{\sigma}_i(s),\sigma_{-i}(s),\theta,\epsilon_i)$. Doraszelski and Satterthwaite (2010) discuss the existence of pure strategy equilibria in settings similar to the one considered here. The introduction of private information over the discrete actions guarantees that at least one pure strategy equilibrium exists, as the best-response curves are continuous. However, there are no guarantees that the equilibrium is unique, a concern we discuss next in the context of my empirical approach.

4.2 Market based emissions policy designs

We use the model to simulate both static and dynamic industry response to the introduction of both price instruments (emissions taxes) and quantity instruments (cap-and-trade programs). In the tax regimes we consider, all domestic producers must pay τ per unit of emissions. In the emissions trading programs we analyze, an emissions cap limits greenhouse gas emissions across multiple emissions-intensive sectors. To comply with the trading program, producers must hold permits to offset their uncontrolled emissions. We impose no spatial or sectoral restrictions on permit trading; permits can be traded freely among all program participants. To keep the analysis more tractable, we do not allow banking or borrowing of permits across time.

The carbon price, τ, is an exogenous parameter. In the case of the tax, this simply means that the level of the tax does not depend on the production and/or pollution decisions of the regulated firms. The tax is set by the regulator and does not change over the time horizon we consider (30 years). In the case of an emissions trading program, we assume that the aggregate marginal abatement cost curve is flat in the neighborhood of the constraint imposed by the emissions cap. This will be an appropriate assumption if the domestic cement industry is a relatively small player in the emissions market, such that changes in industry net supply/demand for permits cannot affect the equilibrium market price.[19] The policy designs we analyze can best be classified into one of four categories: standard auction design/ carbon tax; grandfathering (i.e. lump sum transfer); output-based rebating; emissions-based rebating. In the subsections that follows, these policy design alternatives are described in detail.

4.2.1 Standard design: Emissions tax or emissions trading with auctioned permits

In the wake of failed attempts to implement a federal cap-and-trade program for greenhouse gases, some are advocating for a reconsideration of a carbon tax.[20] In the context of an economy-wide greenhouse gas emissions trading program,a cap-and-trade program that incorporates auctioning also has its proponents.[21] Given our assumption about the exogeneity of the carbon price, these two market-based policy designs are, within our modeling framework, functionally eqiuvalent.

The first policy regime we analyze is indended to capture the most salient features of an emissions tax or an emissions cap-and-trade program in which all emissions permits are allocated via a uniform price auction. In the tax regime, regulated firms must pay a tax τ for each ton of emissions. In the emissions trading regime, the equilibriun permit price is τ; a change in the net supply or demand for permits from the domestic cement industry doesl not affect this price.

The per-period production profit function becomes:

$$\pi_{it} = P\left(q_{it} + \sum_{j \neq i} q_{jt}; \alpha\right) q_{it} - C_i(q_{it}; \delta) - \tau e_i q_{it}, \tag{12}$$

[19]This assumption is likely to be approximately true in the context of a federal GHG trading program that permits offsets. Keohane (2009) estimates the slope of the marginal abatement cost curve in the United States (expressed in present-value terms and in 2005 dollars) to be 8.0 x 10^7 \$/GT CO_2 for the period 2010–2050. Suppose this curve can be used to crudely approximate the permit supply function. If all of the industries deemed to be "presumptively eligible" for allowance rebates reduced their emissions by ten percent for this entire forty year period, the permit price would fall by approximately \$0.25/ ton.

[20]Blinder, Alan. January 31, 2011. "The Carbon Tax Miracle Cure". Wall Street Journal.

[21] For example, in 2007, the Congressional Budget Office Director warned that a failure to auction permits in a federal greenhouse gas emissions trading system "would represent the largest corporate welfare program that has even been enacted in the history of the United States" "Approaches to Reducing Carbon Dioxide Emissions: Hearing before the Committee on the Budget U.S. House of Representatives", November 1, 2007. (testimony of Peter R. Orszag)

where e_i is the firm's emissions rate and E represents aggregate industry emissions.

4.2.2 Grandfathering

In this policy scenario, tradable emissions permits are allocated for free to incumbent firms that pre-date the carbon trading program. Firm-specific permit allocation schedules (i.e. the number of permits the firm will receive each period) are determined at the beginning of the program and are based on historic emissions.

Several studies have demonstrated that a pure grandfathering regime would grossly over-compensate industry for the compliance costs incurred under proposed Federal climate change legislation. For example, a recent paper finds that grandfathering fewer than 15 percent of the emissions allowances generally suffices to prevent profit losses among industries that would suffer the largest percentage losses of profit absent compensation (Goulder, Hafstead, and Dworsky, 2010). Under the grandfathering regime we consider, we assume that a number of permits equal to 20 percent of annual baseline emissions are grandfathered each year to incumbent cement producers. The per period profit function becomes:

$$\begin{aligned} \pi_{it} &= P\left(q_{it} + \sum_{j \neq i} q_{jt}; \alpha\right) q_{it} - C_i(q_{it}; \delta) - \tau(e_i q_{it} - A_i), \qquad (13) \\ \text{with} & \quad \sum_i A_i = \overline{A}. \end{aligned}$$

The number of permits the firm receives for free from the regulator is A_i; $\overline{A}$ represents the total amount of emissions allocated for free to domestic cement producers. We assume that the share of emissions allowances allocated to firm i (i.e. $A_i/\overline{A}$) is equal to its share of the installed kiln capacity at the outset of the program.

Note that the first order conditions associated with static profit maximization under auctioning are identical to those under grandfathering. This highlights the so-called "independence property", which holds that firms' short run production and abatement decisions will be unaffected by the choice between auctioning permits or allocating them freely to firms in lump sum (Hahn and Stavins, 2010).

When permits are grandfathered in a cap and trade program, policy makers must decide ex ante how to deal with new entrants and firms who exit. In our simulations, we assume that a firm forfeits its future entitlements to free permits when it exits the market. We assume that new entrants are not entitled to free permits.[22] In some existing program

[22]In practice, policies regarding free permit allocations to free entrants and former incumbents vary. In

designs (including the EU ETS), some fraction of the permits to be allocated are set aside for new production capacity entering the market. Future work will explore these alternative policy designs that offer free permit allocations as incentive for new entrants.

4.2.3 Output-based allocation/rebating

The third policy regime we analyze incorporates output-based rebating. This scenario can be motivated in two ways. First, under an emissions tax, tax revenues can be rebated to regulated firms based on output. For example, Sweden has refunded revenues from a tax on nitrogen oxide emissions in proportion to output (Sterner and Isaksson, 2006). Second, our modeling approach also captures the essential features of an emissions trading program in which free permit allocations are contingent upon production levels. For example, under proposed state and federal climate change legislation, output-based updating provisions are used to address concerns about near-term competitiveness impacts, job loss, and emissions leakage. Emissions permits are allocated for free to eligible firms using a continuously updated, output-based formula.[23]

Following Bushnell and Chen (2009), we adopt a closed-loop approach to modeling of these kinds of rebating regimes. Permits are allocated/ tax revenues are recycled based on product shares (or emissions shares) in the current period. The per period profit function becomes:

$$\pi_{it} = P\left(q_{it} + \sum_{j\neq i} q_{jt}; \alpha\right) q_{it} - C_i(q_{it}; \delta) - \tau(e_i q_{it} - \theta_i(q_{it})\overline{A}), \tag{14}$$

where $\phi_i(q_{it})$ denotes the share of the total rebate allocated to to firm i. Emissions allowances are allocated (or tax revenues are rebated) according to market share:

$$\theta_i(q_i) = \frac{q_i}{\sum_i q_i}.$$

Implicit in Equation 14 are two simplifying assumptions. First, the rebate a firm receives

the EU ETS, policies governing the free allocation of permits to entrants vary across member states. Most states require forfeiture of free permit allocations upon closure.

[23]Proposed federal climate change legislation included a provision to allocate permits to eligible industries using an output-based formula. These free allocations are intended to compensate both direct compliance costs (i.e. the cost of purchasing permits to offset emissions) and indirect compliances costs (i.e. compliance costs reflected in higher electricity prices). Under California's Assemly Bill 32, implementing agencies have recommended that free allocation to industry will, "to the extent feasible, be based on output-based GHG efficiency "benchmarks" and "update" to reflect changes in production each year for industry with leakage risk" (Greenhouse Gas Cap-and-Trade Regulation Status Update May 17, 2010 California Air Resources Board).

in the current period depends on its production level in that same period. Thus, we do not explicitly account for the fact that firms will discount the value of the subsidy conferred by rebating if the rebate is paid in a future period. Second, the size of the implicit subsidy per unit of output is taken to be exogenous to firms' production decisions. More precisely, we assume that firms do not take into account how their production decisions affects the size of the implicit subsidy γ_i via the effect on aggregate production levels. Together, these assumptions simplify the dynamic problem considerably, while still allowing us to capture the dynamic implications of the grandfathering mechanism to a significant extent.

4.2.4 Output-based allocation updating/rebating

The fourth and final policy design alternative we consider incorporates emissions-based rebating. This works in precisely the same way as output-based rebating, except that rebates (in the form of recycled tax revenues or free emissions permits) are allocated based on emissions. The more emissions intensive a firm, the larger the rebate (per unit of outpu) it recieves. This design has been proposed in cases where firms owning older, less efficient kilns insist that they should be entitled to a larger allowance allocation so as to compensate them for their higher compliance costs. In this case:

$$\theta_i(q_i) = \frac{e_i q_i}{\sum_i e_i q_i}.$$

Firms receive a rebate as long as they are producing in the market. Therefore, new entrants also receive an allocation proportional to either their output or their emissions.

4.3 Modeling Emissions Abatement

Section 3.1 included a discussion of how carbon dioxide emissions reductions can be achieved in the domestic cement industry. If emissions from cement manufacturing were to be subject to a binding cap, it is anticipated that mandated emissions reductions would be achieved through a combination of factors. Chief among these are the replacement of old production processes with new state-of-the-art technology and the increased substitution of less carbon intensive materials for clinker or cement.

We explicitly model what is expected to be the most important efficiency improvement: the replacement of older kiln technology with current, state-of-the-art technology. Our modeling approach is well suited to modeling the retirement of old process equipment and entry of new firms. We assume all new entrants adopt new, state-of-the-art equipment. This assumption finds empirical support in the data. Our specific assumptions about the emissions intensities of old and new production equipment are described in Appendix A.

The substitution of SCM for clinker is also expected to play an important role in delivering emissions reductions in a carbon constrained cement industry. Supplementary cementitious materials are used widely throughout the U.S. as additives to concrete. Utilization rates have varied due to economic considerations and the availability of materials. Although we do not explicitly model the substitution of SCMs for clinker, this substitution is implicitly captured, to some extent, by our estimated demand elasticity.

Ideally, a model designed to simulate industry response to an emissions regulation would accurately capture all viable carbon abatement strategies. Unfortunately, our econometric approach is not well suited to modeling responses that have yet to be observed in the data. Consequently, fuel switching and carbon sequestration are not represented in our model. Although these options are not expected to play as significant a role as efficiency improvements or substitution, this omission will bias up our estimates of the economic costs imposed of the emissions regulations we analyze.[24]

5 Estimation and computation

The econometric estimation is based on the benchmark model, in which the price of emissions is set to zero ($\tau = 0$), i.e. there is no compliance cost due to emissions regulation. Once estimated, this model can be used to simulate the dynamic industry response to market-based emissions regulations that affect firms' production and investment choices primarily through operating costs provided certain assumptions are met. In particular, we will assume that firms' response to a given operating cost change is independent of whether the cost change is caused by emissions regulation or other exogenous factors (such as changes in energy prices or other inputs).

5.1 Estimation

Although our data sources and identification strategy are similar to Ryan (2011), there are some important differences in how the model is specified and estimated. In this section,

[24]In future work we plan to compute what would be an upper bound on the cost of fuel switching for it to be observed in equilibrium together with sensitivity analysis on how important such adoption would be for mitigating the adverse effects of carbon regulation.

significant deviations are discussed. The interested reader is referred to Ryan (2011) for additional details regarding the data and estimation.

5.1.1 Regional market definition

The USGS collects establishment-level data from all domestic Portland cement producers and publishes these data in an annual Minerals Yearbook. Cement price and sales data are aggregated to the regional market level to protect the confidentiality of the respondents. In recent years, increased consolidation of asset ownership has required higher levels of data aggregation. Conversations with the experts at USGS indicate that the current regional market definitions group plants that are unlikely to compete with each other (Van Oss, personal communication).

Rather than adopt the USGS protocols, we base our regional market definitions on the industry-accepted limitations of economic transport as well as company-specific SEC 10k filings which include information regarding markets served by specific plants. To merge the USGS cement prices with our data set, USGS prices are weighted by kiln capacity in each region. For example, if kiln capacity in the region we define as region A is equally divided between USGS defined markets B and C, we define the price in region A to be the average price reported in USGS markets B and C. We report some descriptive statistics using USGS data from 2006 for our regional markets in Table 1.

This table helps to highlight inter-regional variation in market size, emissions intensity, and trade exposure. Notably, the degree of import penetration varies significantly across inland and coastal areas. Whereas several inland markets are supplied exclusively by domestic production, imports now account for over half of domestic cement consumption in Seattle.Import penetration rates tend to be highest along the coasts versus inland waterways.

One concern with using these market definitions is that the demand estimates implied by the USGS data may overstate the residual demand faced by firms under these more restrictive market definitions. To deal with this problem, we re-estimate the intercepts of the demand curves, holding the elasticity of demand constant, by matching the predicted capacity of the market under each parameter guess to the actual observed capacities. We solve the dynamic programming problem faced by firms in each market, and check to see if the firms want to embark on an immediate investment program, as would be the case if the USGS estimates overstate demand. We then search for an intercept of the residual demand curve to match the observed equilibrium level of capacity.

Table 1: Descriptive Statistics for Regional Markets (based on 2006 data)

Market	Number of Firms	Capacity	Emissions Rate	Import Market Share
Atlanta	6	1285	0.97	0.12
Baltimore/Philadelphia	6	1497	0.99	0.12
Birmingham	5	1288	0.94	0.35
Chicago	5	972	0.98	0.04
Cincinnati	3	875	0.93	0.21
Dallas	5	1766	1.05	0
Denver	4	998	0.95	0
Detroit	3	1749	1.02	0.19
Florida	5	1297	0.93	0.35
Kansas City	4	1661	0.95	0
Los Angeles	6	1733	0.93	0.18
Minneapolis	1	1862	0.93	0.2
New York/Boston	4	1033	1.16	0.45
Phoenix	4	1138	0.93	0.13
Pittsburgh	3	614	1.08	0
Salt Lake City	2	1336	1.01	0
San Antonio	6	1318	0.95	0.3
San Francisco	4	931	0.93	0.18
Seattle	2	607	1.05	0.65
St Louis	4	1358	1.05	0

5.1.2 Import supply and residual demand elasticities

We estimate the following demand equation using two stage least squares (2SLS):

$$\ln Q_{mt} = \gamma_0 + \gamma_1 \ln P_{mt} + \gamma_{2m} + \gamma_3' \ln X_{mt} + \varepsilon_{1mt}. \tag{15}$$

The dependent variable is the natural log of the total market demand in market m in year t. The coefficient on market price, γ_1, is the elasticity of demand, and X_{mt} is a set of demand shifters.

We instrument for the potential endogeneity of cement price using supplyside cost shifters: coal prices, gas prices, electricity rates, and wage rates. Each market has a demand shifter in the intercept, γ_{2m}, using Atlanta as the baseline market. Data sources are summarized in Ryan (2011).

Given our interest in understanding how policy-induced operating cost increases could affect import penetration rates, it will be important to separate the import supply response to changes in domestic operating costs from the domestic market demand response. We estimate the following import supply schedule using 2SLS:

$$\ln M_{mt} = \phi_0 + \phi_1 \ln P_{mt} + \phi_{2m} + \phi_3' \ln Z_{mt} + \varepsilon_{2mt}. \quad (16)$$

This model is estimated using data from those markets exposed to import competition over the period 1993-2007. For inland markets supplied entirely by domestic production, all ϕ coefficients are set to zero. The dependent variable is the log of the quantity of cement shipped to market m in year t. The average customs price of cement is P_{mt}. These data are collected by the U.S. Geological Survey and are published in the annual Minerals Yearbook. These data are reported by Customs districts (i.e. groupings of ports of entry). These districts are matched to the regional markets described in the previous section.

We instrument for the import price using new residential construction building starts, gross state product, value of construction, and population. These state-level data are aggregated for all states included in the regional market area. The matrix Z_{mt} includes other plausibly exogenous factors that affect import supply. To capture transportation costs, we subtract the average customs price from the average C.I.F. price of the cement shipments. This residual price accounts for the transportation cost on a per unit basis, as well as the insurance cost and other shipment-related charges. The Z_{mt} matrix also includes coal and oil prices to capture variation in production costs. Region dummy variables capture regional differences.

To construct the residual demand curve faced by domestic producers in a trade exposed market, the import supply at a given price is subtracted from the aggregate demand at that price. The resulting residual demand does not necessarily feature a constant elasticity and potentially features a kink at the price below which importers do not supply any output at the market. In practice, in all the counterfactual simulations some positive imports are observed at coastal markets.[25]

5.2 Estimation results

Table 2 enumerates the parameter estimates used in our simulations. Overall, these estimates appear reasonable.

- The marginal cost estimate of \$30/ton of clinker falls well within the range that is typically reported for domestic production: \$27-\$44 per ton (Van Oss, 2003 ENV).

[25]This is intuitive as the costs of the domestic industry increase in the counterfactuals considered, which weakly raises the market price.

- The import supply elasticity point estimate is 2.5. When analyzing the impacts of environmental regulations, the US EPA assumes an import supply elasticity of 2 for the cement sector based on Broda et al (2008).
- The elasticity of aggregate demand is 2.96. This is higher in absolute value than some other demand elasticities reported in the literature. For example, Jans and Rosenbaum (1996) estimate a domestic demand elasticity of -0.81. On the other hand, using much higher-quality data, Foster, Haltiwanger, and Syverson (2008) estimate several similar high demand elasticities for homogeneous goods industries, such as -5.93 for ready-mixed concrete, cement's downstream industry.
- Investment costs are roughly in line with the accounting costs cited in Salvo (2010), which reports a cost of \$200 per ton of installed capacity. Our numbers are slightly higher, which in line with the idea that these costs represent economic opportunity costs as opposed to accounting costs. The implied cost of a cement plant is also in line with plant costs reported in newspapers and trade journals. For example, on October 15, 2010, it was reported that the most recent expansion of the Texas Industries New Braunfels cement plant, increasing capacity from 900 thousand tons per year to 2.3 million tons per year, was pegged at a cost of \$350M, which implies a cost of \$250 per ton of installed capacity.[26]
- The magnitudes of the fixed costs are reasonable at face value, and in conjunction with the estimated variances, are in accord with the observed rates of investment, entry, and exit in the cement industry.

Some of the parameters from the model described above are not reported. Substantial divestment is virtually never observed in the data and thus the estimates of divestment costs to be very large. Fixed costs of production and operation are also not reported, as these are set to zero. The reason is that we do not observe sufficient periods of operation without production (mothballing) which are required to separately identify those parameters from the distribution of exit costs.

5.3 Computation

Once the parameters have been estimated, the model can be computed to compare the market performance under market-based policy designs. In order to compute the equilibrium of the

[26] Source: KGNB Radio, New Braunfels, Texas.

Table 2: Simulation Parameters

Parameter	Value
Demand Parameters	
Constant	20.38
Elasticity of Demand	-2.96
Discount Factor	
β	0.9
Production Parameters	
Capacity Cost	1.157E10
Capacity Cost Binding Level	1.896
Marginal Cost	30
Investment Parameters	
Fixed Cost Mean	1,798
Fixed Cost Standard Deviation	420
Marginal Cost	233
Exit Cost	
Scrap Distribution Mean	-67,490
Scrap Distribution Standard Deviation	55,167
Entry Distribution	
Entry Cost Mean	172,680
Entry Cost Standard Variance	41,559

game, we make use of parametric approximation methods. In particular, we interpolate the value function using cubic splines. The reasons behind using parametric methods are twofold. First, the game has a continuous state space, given by the vector of capacities of the firms. By using parametric methods, we can allow firms to deterministically choose their capacity in a continuous space. Second, parametric approximation methods can be useful to improve computational speed. Previous work has already suggested the potential benefits of using parametric approximation methods (Pakes and McGuire, 1994).

Parametric value function methods have been explored in a single agent dynamic programming context.[27] However, they have not been widely used in dynamic games, particularly in games in which players take discrete actions, such as entry and exit (Doraszelski and Pakes, 2007). In our application, we find the method to perform well compared to a discrete value function method. In particular, this parametric method allows us to treat capacity as a continuous state, which improves the convergence properties of the game.[28]

The procedure we use is similar in spirit to the discrete value function iteration approach. In both methods, the value function is evaluated at a finite number of points. At each iteration and for a given guess of the value function, firms' strategies are computed optimally (*policy step*). Then, the value function is updated accordingly (*value function step*). This process is repeated until the value function and the policy functions do not change significantly.

The difference between the discrete value function iteration and our iterative approach is that we approximate the value function with a flexible parametric form. In particular, given a guess for the value function V^k at pre-specified grid points, we interpolate the value function with a multi-dimensional uniform cubic spline, which can be computed very efficiently (Habermann and Kindermann, 2007).[29] This interpolation defines an approximation of the value function in a continuous space of dimension equal to the number of active firms. For a given number of firms active N_A in the market, the value function at any capacity vector s is approximated as,

$$\hat{V}_i^k(s) = \sum_{j=1}^{(J+2)^A} \phi_{N_A,j} B_{N_A,j}(s), \tag{17}$$

[27]For a general treatment of approximation methods used in the context of dynamic programming, see Judd (1998). An assessment of these methods in a single agent model can be found in Benitez-Silva et al. (2000).

[28]This is mainly driven by the fact that firms take deterministic actions with respect to the continuous state.

[29]For a detailed treatment of splines methods, see de Boor (2001).

where J is the number of grid points, $\phi_{N_A,ij}$ are the coefficients computed by interpolating the values V^k when there are A active firms, and $B_{N_A,j}(s)$ is the spline weight given to coefficient $\phi_{N_A,j}$ when the capacity state equals s. This coefficient is the product of capacity weights for each of the incumbent firms, so that $B_{N_A,j}(s) = \prod_{i \in A} B_j(s_i)$.

In the *policy step*, optimal strategies are computed over this continuous function. For a given firm, we compute the conditional single-dimensional value function, given the capacity values of the other firms, $\hat{V}_i^k(s_i|s_{-i})$. This formulation allows us to represent the single-dimensional investment problem of the firm. The following expression defines the expected value function of the firm conditional on staying in the market and investing to a new capacity s'_i. Firms maximize,

$$\max_{s'_i} \quad \pi_i(s_i, s'_i|s_{-i}) + \sum_{s'_{-i} \in S_{-i}} Pr^k(s'_{-i}; \sigma^k(s)) \hat{V}_i^k(s'_i|s'_{-i}). \tag{18}$$

We compute the optimal strategy by making use of the differentiability properties of the cubic splines, which allows us to compute the first-order conditions with respect to investment. Given that the cubic spline does not restrict the value function to be concave, we check all local optima in order to determine the optimal strategy of the firm.[30] Conditional on optimal investment strategies, we then compute the new policy function with respect to the entry, investment and exit probabilities, which gives us an updated optimal policy σ^{k+1}. This allows us to compute a new guess for the value function V^{k+1} in the *value function step*.

The process is iterated until the strategies for each of the firms and the value function in each of the possible states do not change more than an established convergence criterion, such that $\parallel \sigma^{k+1} - \sigma^k \parallel < \epsilon_\sigma$ and $\parallel V^{k+1} - V^k \parallel < \epsilon_V$.

6 Welfare measures and metrics

In this section, we first describe the analytical framework we will use to interpret the results. We then discuss an important parameter in our analysis: the social cost of carbon.

[30] Given that the cubic spline is defined by a cubic polynomial at each of the grid intervals, this implies that at most there will be $2(J-1)+2$ candidate local optima, where J is the number of grid points.

6.1 Analytical framework

We focus exclusively on outcomes in the domestic cement industry. Within a regional cement market, static, per period welfare is defined as follows:

$$w(s,a;\alpha,\delta_i,\gamma,\tau,e)=\int_0^Q P(x;\alpha)dx-\sum_i\int_0^{q_i} C(x;\delta_i)dx-P(Q;\alpha)M(P;\gamma)-\tau\sum_i e_i q_i-\tau e^M M(P;\gamma). \tag{19}$$

The vector e includes the emissions intensity measures of both domestic producers and foreign imports. The parameter e^M denotes the emissions intensity of imports. This value is estimated using an import volume weighted average of estimated foreign cement producers' emissions intensities (Worrell et al., 2001).

This welfare measure ignores any surplus captured by the producers of domestic imports; domestic policy makers presumably ignore the economic costs and benefits accruing to producers and consumers outside their jurisdiction. In specifying this welfare function, we assume that marginal damages from carbon dioxide emissions are constant and equal to the assumed equilibrium permit price τ. Because damages from greenhouse gases are independent of where in the world the emissions occur, we penalize both domestic and foreign emissions at a rate of τ per unit. We also assume that the cement sector is small relative to the larger emissions trading program, such changes in cement industry emissions do not affect the equilibrium permit price.[31]

Each market-based policy regime we consider affects Equation 19 through its effect on firm-level production choices. Our static, single period, aggregate welfare measure sums (19) across regional markets. Welfare analysis in the dynamic simulations sums Equation 19 across markets and time periods, subtracting any entry, exit and investment costs accruing over the time horizon we consider. This measure of dynamic efficiency is somewhat unconventional insofar as it rules out innovation and technological change. For our purposes, a dynamically efficient outcome maximizes social welfare subject to the constraints imposed by existing and proven production technologies.

Expositionally, it is useful to decompose the net welfare impact of a policy intervention into the three components introduced in Section 2. We define three welfare measures:

- $W1$ captures changes in the private economic surplus accruing from domestic cement

[31] If net permit demand from the cement sector can affect the equilibrium permit price, our estimates of the costs of allocation updating, vis a vis auctioning or grandfathering, will be too low.

consumption (i.e. the first three terms in Equation 19.

- $W2$ accounts for both economic surplus changes ($W1$) plus the benefits of emissions reductions.
- $W3$ accounts for emissions leakage across policy designs by penalizing foreign increases in emissions.

In this preliminary draft, we report results from simulating outcomes in nine out of twenty regional markets: Cincinnati, Detroit, Minneapolis, Pittsburgh, Salt Lake City, San Francisco, Seattle, Phoenix, and St. Louis. Across regional markets we analyze, there is significant variation in market size, plant technology, and import presence. For example, the Salt Lake City market is not accessible by water and demand is met entirely by domestic suppliers, who have heterogeneity in their emissions rates. In contrast, a market like San Francisco is trade-exposed and incumbent producers are homogenous with respect to emissions rates.

In the future, the scope of the analysis will be expanded to include all domestic cement markets. This will allow us to provide a more comprehensive assessment of the industry-wide effects of the policies we consider, and to assess the extent to which a "one-size-fits-all" policy regime can result in differential outcomes across heterogeneous regional markets.

6.2 The Social Cost of Carbon

The τ value we use to penalize each ton of simulated $\mathrm{CO_2}$ emissions is intended to capture the monetized damages associated with an incremental (one ton) increase in carbon emissions. Given the uncertainty inherent in this kind of policy analysis, it is important to consider a range of values of τ. The range of values we choose to consider, \$5 to \$75 per ton of CO_2, is informed by a landmark interagency process which produced estimates of the social cost of carbon (SCC) for use in policy analysis(Greenstone et al., 2011).

Table 3 summarizes the four SCC schedules that were selected in this process. In light of disagreements about the appropriate choice of interest rate, three different discount rates are used (corresponding to the first three schedules). The final schedule (fourth column) corresponds to a scenario with higher than expected economic costs from climate change. The SCC increases over time because future emissions are expected to produce larger incremental damages as physical and economic systems become more stressed.[32]

[32]In our analysis, we assume the carbon price does not change over the time horizon we consider.

Table 3: Estimated social cost of carbon
($ per metric ton of carbon dioxide in $2007)

	(1)	(2)	(3)	(4)
Discount rate	5%	3%	2.5%	3%
2010	4.70	21.40	35.10	64.90
2020	6.80	26.30	41.70	80.70
2030	9.70	32.80	50.00	100.00

Source: U.S. Department of Energy (2010), "Final Rule Technical Support Document (TSD): Energy Efficiency Program for Commercial and Industrial Equipment: Small Electric Motors," Appendix 15A (by the Interagency Working Group on Social Cost of Carbon): "Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866".

Another important assumption we make is that the carbon price reflects the true social cost of carbon. Thus, the carbon tax or permit price and the social cost of carbon are assumed to be one and the same. This approach has expositinal advantages. However, it is essential to keep this assumption in mind when comparing results across scenarios. An alternative approach would hold the assumed SCC value constant across scenarios associated with different permit prices/tax levels. We leave this extension for future work.

7 Simulation results

Before turning to the simulation results, it is important to highlight some underlying assumptions and to issue some caveats with respect to interpretation.

One assumption underpinning our counterfactual simulations is that cement producers will respond to economic changes induced by a market-based emissions regulation in the same way that they have responded historically to similar changes induced by other exogenous market forces. This assumption seems quite plausible given that the policy designs we consider operate solely through altering production costs and revenues.

Another important assumption is that our structural assumptions, including assumptions about how firms respond to changes in policy incentives, will hold out of sample. We observe significant variation in plausibly exogenous supply and demand shifters across regional markets and across time; this variation is essential to identification. However, our inferences at high carbon prices are quite far from historical experience. To put this in context, consider that a carbon price of $60/ton would approximately triple the estimated marginal operating costs of the average cement producer. A higher-level concern is that plausible general equilibrium effects of high carbon prices could lead to unforeseen structural shifts in the supply and demand curves characterizing cement outcomes.

In what follows, we have elected to report simulation results for the range of SCC values that have been deemed policy relevant. However, it is important to keep in mind that the higher the carbon price we consider, the farther out counterfactual is from the data we observe, and the more sensitive our simulation results will be to our modeling assumptions. That said, we believe that our results both help illustrate the general forces shaping the interaction of market structure and carbon regulation and provide the best possible estimates of efficiency and distributional welfare effects under a range of policies.

This section begins with a summary of the simulated static and dynamic responses of the domestic cement industry to the range of counterfactual policy interventions we consider. We begin by emphasizing the results using the model that accounts for industry dynamics. We then contrast our welfare measures across the static and dynamic simulation exercises. We conclude with a discussion of how outcomes vary across regional markets with different characteristics.

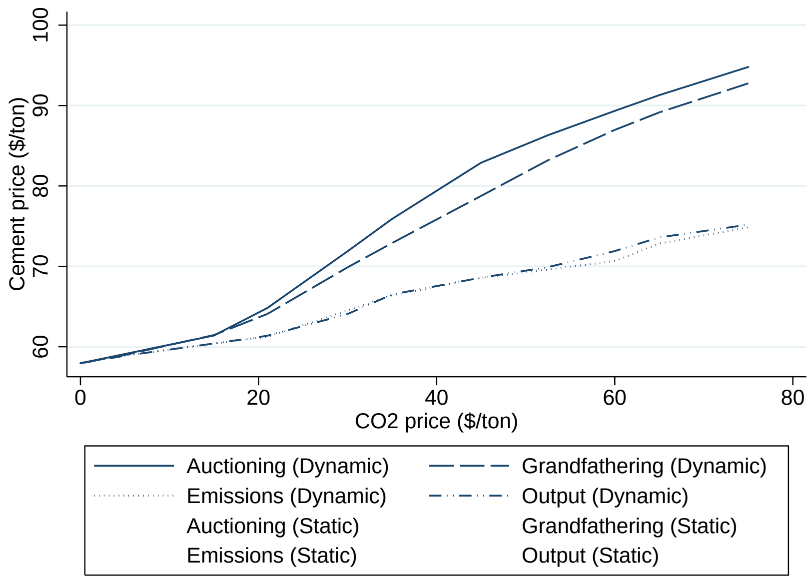
7.1 Market outcomes and aggregate welfare measures

7.1.1 Cement prices

Figure 3 plots quantity-weighted average cement prices as a function of the exogenous permit price (or emissions tax) τ. The introduction of market-based emissions regulation increases equilibrium cement prices across the range of τ values we consider. Cement price increases are most pronounced under the standard auction/tax regime. Under this policy design, firms must bear the complete cost of compliance; no compensation in the form of contingent rebates or lump sum transfers is offered.

Note that equilibrium cement prices vary across grandfathering and auctioning regimes. Thus, the so-called independence property fails to hold when industry dynamics are incorporated into the model. Under the grandfathering regime, an incumbent firm receives a lump sum transfer each period in the form of free permit allocation. An incumbent firm forfeits this entitlement when it chooses to exit. This lowers the exit threshold for incumbents such

Figure 3: Counterfactual: Cement Prices

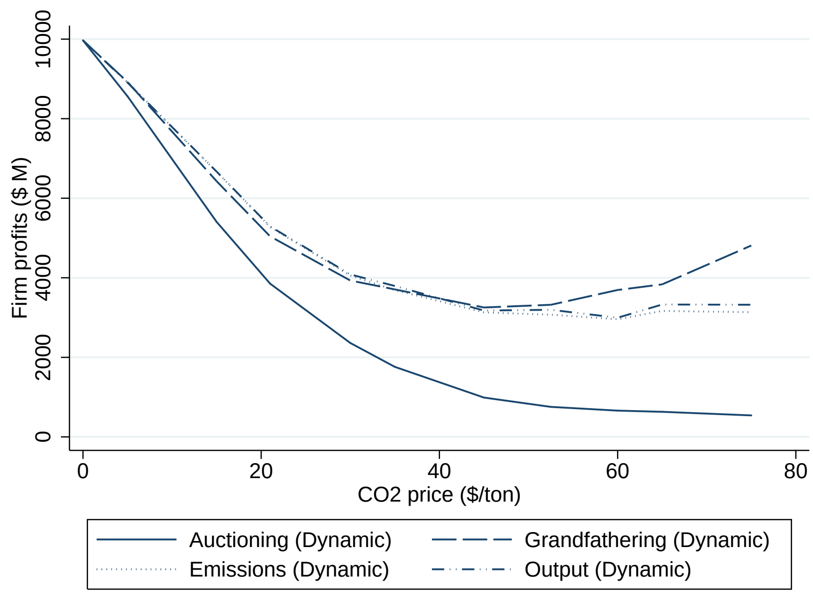


that exit rates are lower under grandfathering as compared to auctioning. As a consequence, cement markets are less concentrated at higher permit prices, and equilibrium cement prices are lower compared to the standard auctioning/emissions tax case.

A striking feature of Figure 3 is that, for a given value of τ, cement prices are much lower under policy regimes that incorporate either type of dynamic rebating. The production subsidy that is implicitly conferred by rebating partially mitigates the impact of the emissions regulation on cement prices.

Outcomes differ only slightly across the output-based and emissions-based updating regimes. The reason is two-fold: first, new firms enter at a fixed frontier emissions rate, so as the industry turns over we asymptote towards having identical outcomes under both policies. Second, among incumbents, the only margin for differences between the two is differential reorganization of production across units with different emissions intensity. This margin is relatively small compared to the overall contraction in market quantity; as such, differences between the two policies are masked by the overall market changes.

Figure 4: Counterfactual: Cement Profits



7.1.2 Industry profits

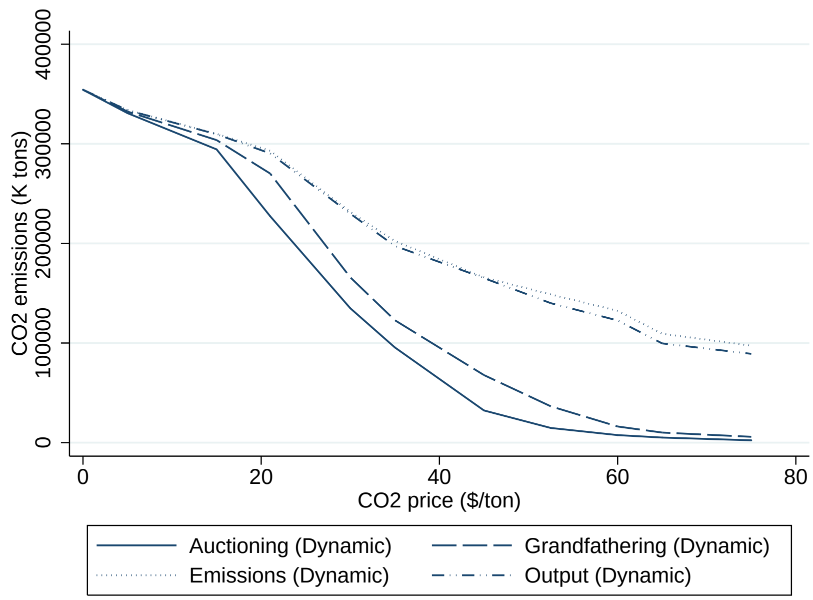
Figure 4 plots the present discounted profits earned by the regulated domestic cement producers over the range of carbon values we consider. For any given value of τ, profits are most significantly impacted by the auctioning regime because firms must pay the tax (or hold permits) to offset emissions, but receive no rebate or compensation for incurring these costs.

Note that discounted industry profits are increasing with τ over the range of higher carbon values in the grandfathering regime. As the carbon price increases, so does the value of the lump sum transfer (in the form of free permits) allocated to incumbent firms. At very high permit prices, some firms will have an incentive to sell permits versus using them to offset their own emissions. This revenue from selling unused permits explains the non-monotonic and increasing (in τ) discrepancy in profits across rebating and grandfathering regimes.

7.1.3 Domestic emissions

Policy makers are very concerned about how industry emissions will be impacted by alternative forms of market-based emissions regulation. Figure 5 shows how emissions from domestic cement producers, summed across all markets and time periods, decreases with τ.

Figure 5: Counterfactual: Domestic Emissions

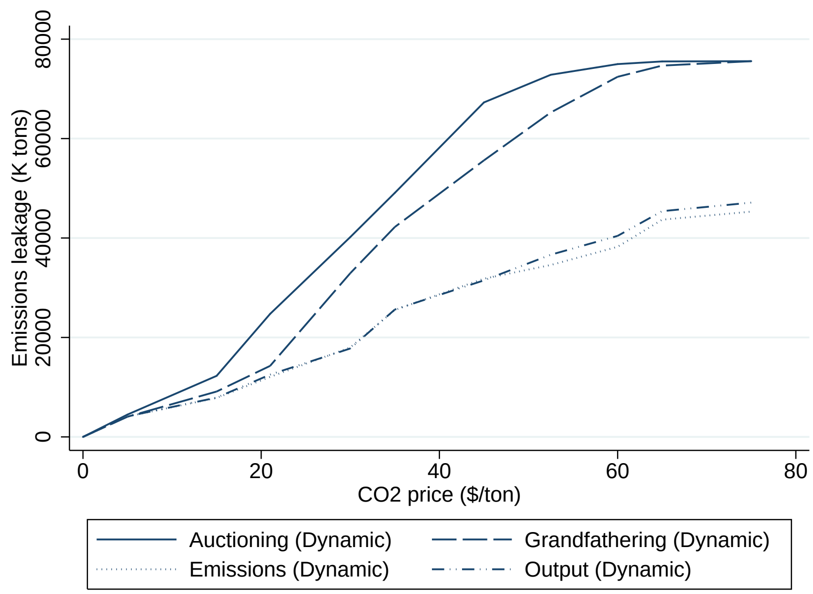


For a given carbon price, the net cost of emitting carbon dioxide (as perceived by firms), is highest under the auctioning regime and lowest under contingent rebating. Consequently, industry emissions are lowest under the auctioning regime and highest under rebating regimes. Given that the implicit subsidy per unit of cement production is higher for more emissions intensive producers under emissions-based rebating, we do see slightly elevated emissions under emissions-based, versus output-based, updating.

7.1.4 Emissions leakage

In the case of trade-exposed emissions-intensive industries, the potential for emissions leakage is a serious issue. Figure 6 plots the simulated leakage under each policy scenario. Our results suggest that there is potential for significant leakage in the US cement industry. Intuitively, auctioning leads to the highest amount of leakage because it places the highest cost burden on domestic producers. In the long-run, increased costs influence both the intensive margin through reduced production and the extensive margin as the rates of exit are highest under auctioning. In line with the earlier discussion, grandfathering slows the rate of exit vis a vis auctioning, thus slowing the rate of leakage. At very high carbon prices, the leakage rates converge across grandfathering and auctioning regimes because all domestic firms have

Figure 6: Counterfactual: Emissions Leakage



exited trade exposed markets.

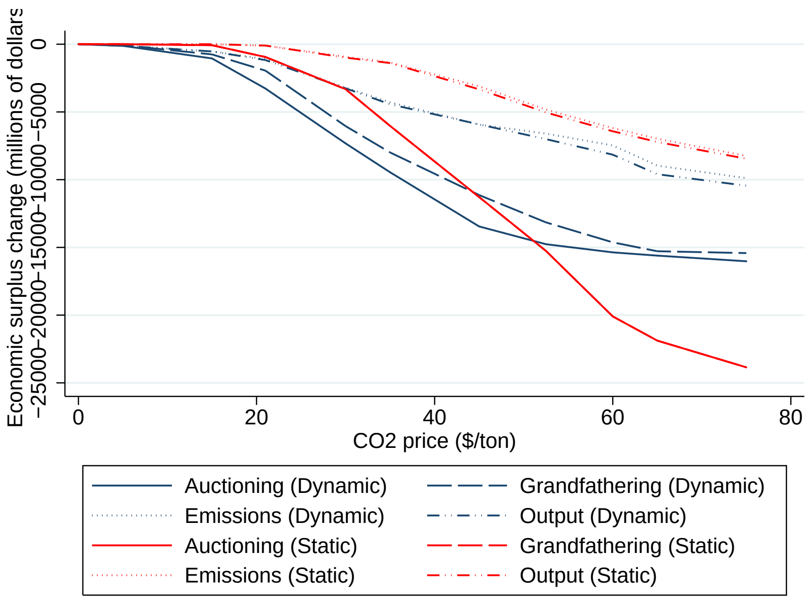
The results also demonstrate that both output and emissions-based rebating significantly mitigates emissions leakage. The rebates incentivize relatively high levels of domestic production, thus limiting the extent to which imports outcompete domestic production in trade exposed markets. At the extensive margin, incumbents are more valuable under dynamic updating in comparison to auctioning, which helps keep them active in the market, further decreasing leakage.

7.1.5 Decomposing Changes in Welfare

Our fundamental objective is to investigate the welfare implications of the alternative policy designs we consider. In what follows, the emissions unconstrained (i.e. unregulated) case serves as a benchmark. We present the three welfare metrics introduced in the previous section, decomposing along conceptually distinct lines: product market welfare consisting of producer profits, consumer surplus, and government revenues; benefits accruing to emissions reductions; and costs due to emissions leakage.

To highlight the importance of accounting for industry dynamics, we contrast the results of our dynamic simulations with a simulation exercise that holds fixed industry structure and

Figure 7: Counterfactual: W1



technology characteristics. A common practice in ex ante policy analysis involves simulating regulatory effects in a static setting, using a representative year as the basis for estimating annual regulatory impacts, and then using that test year to extrapolate outcomes over a longer time horizon (OAQPS, 1999). We adopt this approach here. To generate our "static" results, we simulate a single period market outcome in the unregulated baseline case and under the range of counterfactual policy designs we consider. To facilitate comparisons with our dynamic simulations, these results are expressed as net present values using a social discount rate of three percent. We assume the simulated annual outcomes would be observed each year of the 30 year time horizon we consider.

W1: Product Market Surplus Changes in the first welfare metric, W1, capture differences in producer and consumer surplus while also accounting for revenues raised by the government through taxation or permit sales. This is a measure of how the local market changes in response to the regulation, and is a major component of understanding welfare changes in concentrated industries.

Figure 7 shows how economic surplus is impacted across policy designs and assumed carbon prices. Given that this W1 measure captures none of the benefits from emissions abatement, these changes are all negative.

An interesting result to emerge from the static simulations (in red) is the lack of industry response to carbon prices at or below \$20. In the benchmark (unregulated) case, many firms are capacity constrained, producing at a corner solution, and thus are earning scarcity rents. When firms are required to internalize a relatively low emissions cost at a per unit cost of \$20 or less, scarcity rents are reduced, but output decisions are essentially unaffected in the short run.

In the static simulations, it is also the case that outcomes under the auctioning and grandfathering regimes exhibit the independence property: for any given carbon price, impacts on W1 are identical. Intuitively, this is because the short run incentives in production are identical across these two regimes.

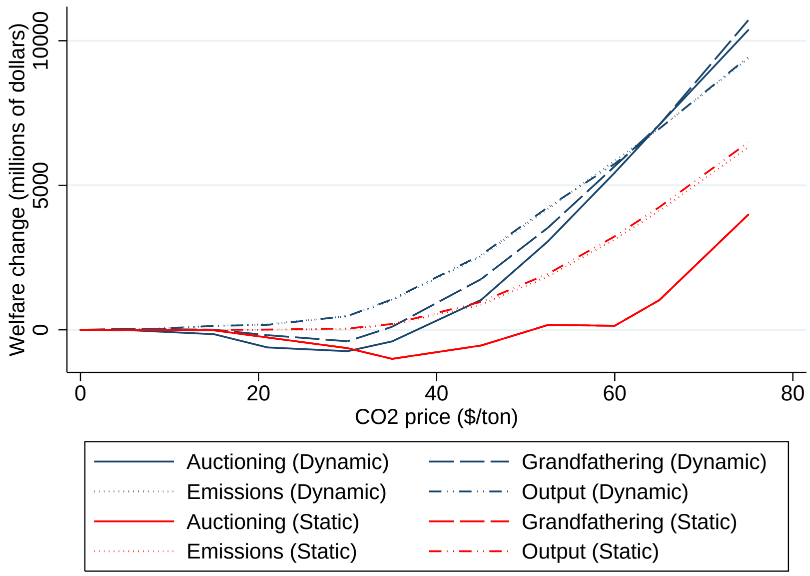
Comparisons across static and dynamic simulations highlight how the evolution of industry structure can affect policy outcomes. First, policy impacts equilibrium prices and quantities at much lower carbon prices in the dynamic simulations. This is due to the reduction in cement production capacity that the emissions regulation induces. The rate of exit is most accelerated under auctioning; firms are more likely to exit when they do not have a steady stream of freely allocated permits to look forward to. Consequently, welfare impacts are most negative under auctioning.

Second, for most of the carbon values we consider, incorporating industry dynamics leads to more pronounced negative welfare impacts. This can again be explained by the capacity/disinvestment response which is shut off in static case.

Finally, whereas outcomes under grandfathering and auctioning are indistinguishable in the static case, negative welfare impacts are mitigated somewhat under grandfathering in the dynamic simulations. This divergence is the results of two countervailing forces. On one hand, high carbon prices incentivize firms to reduce their production, which harms both consumer and producer surplus. On the other hand, grandfathered firms hold an increasingly valuable resource as carbon prices go up. This creates an incentive for firms to remain in the market (versus exiting) because they would otherwise forfeit their entitlements to free permits in the future.

W2: Accounting for Domestic Emissions Abatement Figure 7 fails to capture any of the benefits from emissions abatement. Welfare measure W2, as shown in Figure 8, adds the social benefits associated with reducing CO_2 emissions from domestic cement producers to the changes in the product market measured under W1. Recall that the value of the avoided emissions are assumed to be equal to the prevailing permit price or tax. Thus, the

Figure 8: Counterfactual: W2

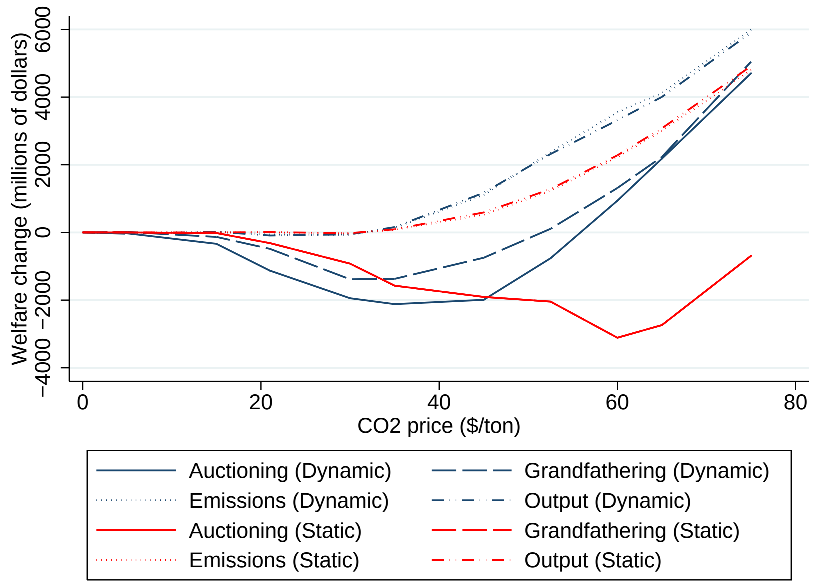


welfare adjustment per unit of emissions abated is increasing along the horizontal axis of Figure 8.

Beginning with the static simulations, the benefits from internalizing the emissions externality more than offset the economic costs under the policy regimes that incorporate rebating; net welfare impacts are weakly positive across all carbon values. The same cannot be said for the grandfathering and auctioning regimes. Over the mid-range of the carbon values we consider, the net welfare impacts are negative. This result is driven by Buchanan's observation that there are two competing distortions in concentrated markets with externalities. As the assumed social cost of carbon increases, the value of avoided damages ultimately overwhelms the value of the lost economic surplus, and the net welfare impacts turn positive.

The dynamic simulations yield somewhat different results. Welfare gains associated with rebating regimes and grandfathering in the dynamic case are larger than those generated using the static model. This is partly due to firms' ability to invest in cleaner production equipment (which reduces the emissions intensity per unit of cement produced) and partly because the quantity produced is lower, resulting in lower damages from domestic emissions. Also note that, at very high prices, grandfathering and auctioning welfare dominate the rebating regimes. The factors that made auctioning and grandfathering unattractive under metric W1—namely, output contraction and the accelerated exit of domestic producers—

Figure 9: Counterfactual: W3



make them attractive under metric W2 once we account for the benefit of carbon emissions reductions.

W3: Accounting for Emissions Leakage Carbon dioxide is a global, uniformly mixed pollutant. A comprehensive welfare analysis of policy impacts should account for any policy-induced increases in emissions in other jurisdictions. Our final welfare metric, W3, augments W2 by accounting for damages associated with emissions leakage. Emissions occurring in other jurisdictions are penalized at the same rate as domestic emissions.[33]

Figure 9 illustrates the welfare impacts of the policy regimes we consider using this more comprehensive welfare measure. In the static simulations, once leakage is accounted for, welfare impacts of the grandfathering and auctioning regimes are negative across the full range of carbon values we consider. In contrast, the benefits from domestic emissions reductions more than offset costs associated with emissions leakage and the excessive withholding of output and investment in the upper range of carbon values when rebates are incorporated in the policy design.

In the dynamic simulations, the net welfare impacts remain at or close to zero for all

[33]Ignoring co-pollutants, damages from emissions are independent of location. This contrasts to other emissions that have spatially-varying damages. See, for example, Fowlie and Muller (2010).

carbon values below \$35. Net welfare impacts of grandfathering and auctioning are negative for carbon values below \$50. The welfare ordering of policy regimes no longer reverses at higher carbon values. The policy designs that incorporate rebating welfare dominate over the range of carbon values we consider.

To help summarize this discussion, Table 4 reports key results from dynamic simulations which assume carbon values of \$21 and \$35, respectively. The auctioning regime is associated with the highest cement prices, the lowest level of installed domestic production capacity, and the lowest domestic profits of all regimes. By all welfare measures, the net welfare impacts of auctioning are negative.

For succinctness, the table reports results for output-based updating only; emissions-based updating has very similar impacts. At these carbon values, rebating regimes welfare dominate auctioning and grandfathering. Intuitively, the benefits from rebating (mitigation of the exercise of market power and emissions leakage in trade-exposed markets) outweigh the costs (dampened incentives for emissions abatement).

These qualitative results are robust to a wide range of demand elasticity estimates, which are a key determinant of the consumer gross surplus in the model and therefore an important element in our welfare measures. Appendix C presents a table with W3 welfare differences for different carbon prices and elasticities. As one would expect, we find that the negative effects are even more persistent when demand is more inelastic, but still present for more elastic demand curves.

7.2 Heterogeneous impacts of environmental regulation

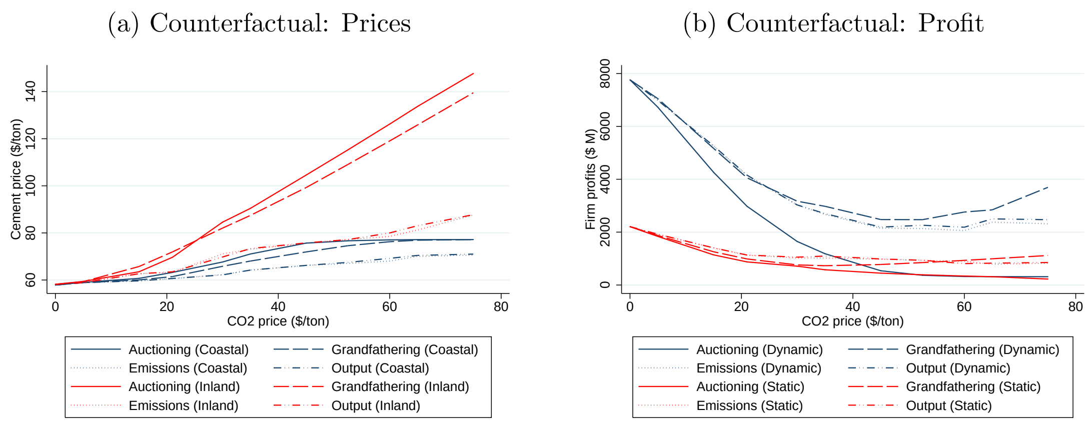
The simulation results allow us to examine the impact of a federal environmental regulation on local markets. When carbon policy is discussed, usually one-size-fits-all designs are considered. For example, in the case of the cement industry, implicit output or emissions-based updating mechanisms are considered for implementation in all markets. However, given the differences in the industry composition of local markets, as well as the differences in trade exposure, these markets can be impacted very differently due to the introduction of carbon prices. Therefore, the distribution of costs from a seemingly uniform federal policy can have heterogeneous impacts in different regions.

Figure 10 represents the average price of cement in coastal and inland markets. One can see that prices raise more rapidly in inland markets, as firms in that market do not face competition from unregulated firms and can pass-through more of the costs of the emissions. The effect is particularly striking for higher carbon prices, in which coastal markets reach a

Table 4: Dynamic simulation results: Social cost of carbon values \$21 and \$35/ton CO_2

Outcome	SCC Value	Baseline	Auction	Grand-father	Emissions-rebating
Market size	\$21	20,670	14,343	16,908	16,753
(tons per year)	\$35	20,670	8,256	15,209	12,894
Quantity-weighted price	\$21	\$58.46	\$64.06	\$62.46	\$61.33
(\$/ton)	\$35	\$58.46	\$71.26	\$69.60	\$65.39
Domestic emissions	\$21	350,000	230,000	270,000	290,000
(tons CO2)	\$35	350,000	95,588	120,000	200,000
Emissions rate	\$21	0.98	0.95	0.98	0.97
(tons CO2/ton clinker)	\$35	0.98	0.92	0.94	0.93
Domestic firm profits	\$21	\$9,969	\$3,853	\$5,043	\$5,292
(\$M)	\$35	\$9,969	\$1,761	\$3,708	\$3,795
Change in W1	\$21	\$0	-\$3,273	-\$1,950	-\$1169
(\$M)	\$35	\$0	-\$9454	-\$7,991	-\$4,433
Change in W2	\$21	\$0	-\$608	-\$184	\$173
(\$M)	\$35	\$0	-\$400	-\$106	\$1,055
Change in W3	\$21	\$0	-\$1,128	-\$484	-\$91
(\$M)	\$35	\$0	-\$2,118	-\$1,371	\$158

Figure 10: Counterfactual: Inland versus Coastal Prices and Profits



threshold price in which all local demand is served by imports from unregulated areas. Even though this is a quite extreme representation of the response of imports with respect to local prices, it highlights one of the major differences between coastal and inland markets.

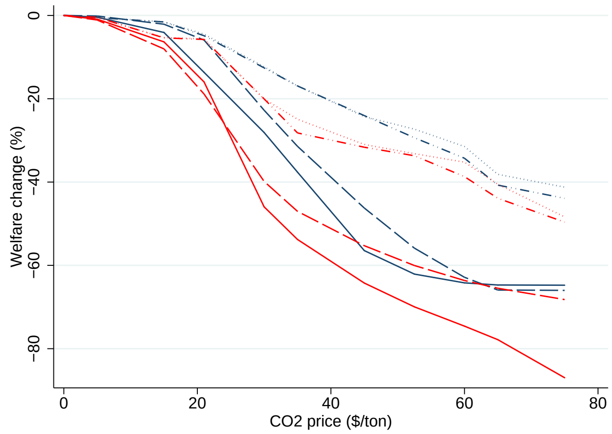
Figure 10 also includes a representation of firms profits. Note that profits in the baseline are larger in coastal markets, as these markets tend to be larger. As one would expect, firms suffer relatively more from the policy in areas in which they are exposed to trade. This is consistent with the view that emissions-intensive trade-exposed industries are the ones that will suffer more from carbon regulation.

Figure 11 represents the welfare differences among coastal and inland markets. In order to re-normalize the measures across coastal and inland markets, the welfare measure represents the percentage change in welfare with respect to the baseline. W1 highlights that industry welfare decreases relatively more in inland markets. This effect is due to the fact that demand is inelastic and there are no imports to serve the market. W2 shows that accounting for emissions reductions in the market has similar effects to both inland and coastal markets, relatively favoring grandfathering and auctioning with respect to updating mechanisms due to the full internalization of emissions costs and, thus, lower emissions.

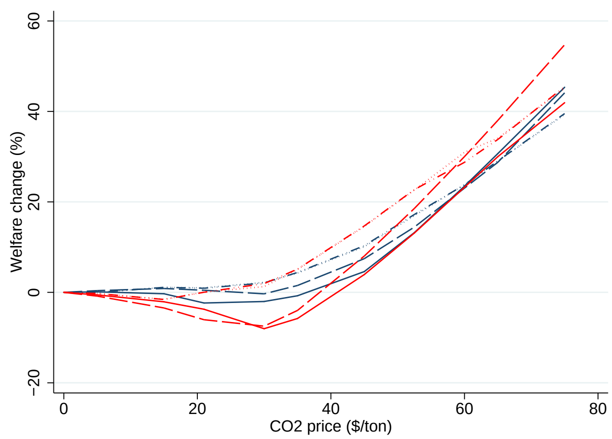
Finally, W3 shows that for sufficiently large abatement costs, coastal market underperform in terms of welfare. The intuition is that on those markets, due to the presence of imports, there is a poor internalization of emissions costs, given that imports attenuate the degree of pass-through in the market. At high carbon social costs, those emissions are particularly harmful in terms of welfare. It is also remarkable that for lower prices, both inland

Figure 11: Counterfactual: Inland versus Coastal Welfare

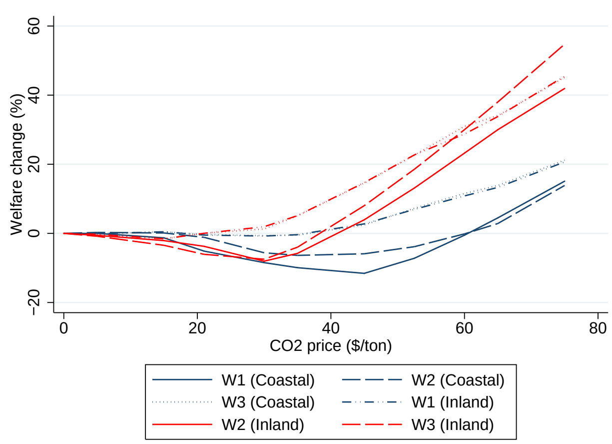
(a) Counterfactual: W1



(b) Counterfactual: W2



(c) Counterfactual: W3



and coastal markets suffer net welfare losses from the regulation and, if anything, coastal markets tend to suffer less. The intuition is that at lower prices, reductions in emissions are not valuable and therefore serving the market relatively dominates in terms of welfare. In coastal markets, the market is better served due to more flexible and cheaper production.

8 Conclusion

We present a dynamic model to evaluate the welfare impacts of market-based regulation of carbon dioxide emissions in the US cement industry. We assess the implications of several alternative policy designs, including those that incorporate both an emissions disincentive (a tax or an obligation to hold an emissions permit) and a production incentive. Simulation results reported in this working paper pertain to only a subset of regional markets. The analysis will ultimately include the entire domestic cement sector.

We find that both the magnitude and the sign of the welfare impacts we estimate depend significantly on how the policy is implemented and what we assume for the social cost of carbon. At low to moderate carbon values, our results echo Buchanan (1969). Market-based emissions regulation that internalizes the full emissions externality leads to small social losses. These losses are exacerbated by emissions leakage in trade exposed regional markets.

At higher carbon values, our results are more in line with Oates and Strassman (1984) who argue that the welfare gains from pollution control will be large relative to losses associated with output contraction and the exercise of market power.

Notably, we find that policy designs that incorporate both an emissions penalty and a production incentive in the form of a rebate welfare dominate more conventional policy designs. Intuitively, the production incentive works to mitigate leakage in trade exposed cement markets and the distortion associated with the exercise of market power.

Of course, these simulation results condition on the structural assumptions that define the underlying model. The higher the carbon price we consider, the farther out counterfactual is from the data we observe, and the more sensitive our simulation results will be to our modeling assumptions.

Policy makers are very interested in understanding how proposed climate change policies would impact strategic, emissions intensive sectors such as the cement industry. The scale and scope of these policy interventions are unprecedented, making it difficult to anticipate how industry will respond and what that response will imply for social welfare. Findings presented in this paper help illustrate the general forces shaping the interaction of market structure and proposed carbon regulations and provide important insights into the efficiency and distributional properties of leading policy design alternatives.

References

Benitez-Silva, H., Rust, J., Hitsch, G., Pauletto, G., and Hall, G. (2000). A comparison of discrete and parametric methods for continuous-state dynamic programming problems. Computing in Economics and Finance 2000 24, Society for Computational Economics.

Bernheim, B. D. and Whinston, M. D. (1990). Multimarket contact and collusive behavior. *RAND Journal of Economics*, 21(1):1–26.

Bohringer, C. and Lange, A. (2005). Economic implications of alternative allocation schemes for emission allowances. *The Scandinavian Journal of Economics*, 107(3):563–581.

Borenstein, S., Bushnell, J. B., and Wolak, F. A. (2002). Measuring market inefficiencies in california's restructured wholesale electricity market. *American Economic Review*, 92(5):1376–1405.

Buchanan, J. M. (1969). External diseconomics, corrective taxes, and market structure. *The American Economic Review*, 59(1):174–7.

Bushnell, J. and Chen, Y. (2009). Regulation, allocation, and leakage in cap-and-trade markets for co2. Technical report, EI@Haas WP-183, Washington, DC.

Bushnell, J., Mansur, E., and Saravia, C. (2008). Vertical arrangements, Market Structure, and competition: an analysis of restructured US electricity markets. *American Economic Review*, 98(1):237–266.

Cropper, M. L. and Oates, W. E. (1992). Environmental economics: A survey. *Journal of Economic Literature*, 30(2):675–740.

de Boor, C. (2001). *A Practical Guide to Splines. Revised Edition.* Springer-Verlag New York Inc.

Doraszelksi, U. and Satterthwaite, M. (2010). Computable markov-perfect industry dynamics. *RAND Journal of Economics*, 41(2):215–243.

Doraszelski, U. and Pakes, A. (2007). A framework for applied dynamic analysis in io. In Armstrong, M. and Porter, R., editors, *Handbook of Industrial Organization*, volume 3, chapter 30, pages 1888–1966. Elsevier.

Ericson, R. and Pakes, A. (1995). Markov perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, 62(1):53–82.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Fowlie, M. and Muller, N. (2010). Designing markets for pollution when damages vary across sources: Evidence from the NO_x budget program. Technical report.

Greenstone, M., Kopits, E., and Wolverton, A. (2011). Estimating the social cost of carbon for use in u.s. federal rulemakings: A summary and interpretation. NBER Working Papers 16913, National Bureau of Economic Research, Inc.

Habermann, C. and Kindermann, F. (2007). Multidimensional spline interpolation: Theory and applications. *Comput. Econ.*, 30(2):153–169.

Humphreys, K. K. and Mahasenan, N. (2001). Towards a sustainable cement industry: substudy 8. Technical report.

Jans, I. and Rosenbaum, D. I. (1997). Multimarket contact and pricing: Evidence from the U.S. cement industry. *International Journal of Industrial Organization*, 15:391–412.

Joskow, P. L. and Kahn, E. (2002). A quantitative analysis of pricing behavior in california's wholesale electricity market during summer 2000. *The Energy Journal*, 23:1–35.

Judd, K. (1998). *Numerical Methods in Economics*. MIT Press.

Mahasenan, N., Dahowski, R., and Davidson, C. (2005). *Greenhouse Gas Control Technologies, Volume I*, chapter The Role of Carbon Dioxide Capture and Sequestration in Reducing Emissions from Cement Plants in North America. Elsevier Science.

Maskin, E. and Tirole, J. (1988). A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs. *Econometrica*, 56:549–69.

Miller, N. H. and Osborne, M. (2010). Competition among Spatially Differentiated Firms: An Estimator with an Application to Cement. EAG Working Paper 10-2.

Millimet, D. L., Roy, S., and Sengupta, A. (2009). Environmental regulations and economic activity: Influence on market structure. *Annual Review of Resource Economics*, 1:99–117.

Oates, W. E. and Strassman, D. L. (1984). Effluent fees and market structure. *Journal of Public Economics*, 24(1):29–46.

Pakes, A. and McGuire, P. (1994). Computing markov-perfect nash equilibria: numerical implications of a dynamic differentiated product model. *RAND Journal of Economics*, 25(4):555–589.

Ponssard, J.-P. and Thomas, C. (2010). Capacity investment under demand uncertainty. an empirical study of the us cement industry, 1994-2006. Technical report.

Puller, S. (2007). Pricing and firm conduct in California's deregulated electricity market. *The Review of Economics and Statistics*, 89(1):75–87.

Ryan, S. P. (2011). The cost of environmental regulation in a concentrated industry. MIT Working Paper.

Ryan, S. P. and Tucker, C. E. (2010). Diversification and the dynamics of technology adoption. *Conditionally accepted at Quantitative Marketing and Economics*.

Sterner, T. (1990). Energy Efficiency and Capital Embodied Technical Change: The Case of Mexican Cement Manufacturing. *The Energy Journal*, 11(2):155–168.

Sterner, T. and Isaksson, L. H. (2006). Refunded emission payments theory, distribution of costs, and swedish experience of nox abatement. *Ecological Economics*, 57(1):93–106.

Sterner, T. and Muller, A. (2008). Output and abatement effects of allocation readjustment in permit trade. *Climatic Change*, 86(1-2):33–49.

Sweeting, A. (2007). Market power in the england and wales wholesale electricity market 1995–2000. *Economic Journal*, 117(520):654–85.

Wolfram, C. (1999). Measuring duopoly power in the British electricity spot market. *The American Economic Review*, 89(4):805–826.

Worrell, E., Price, L., Martin, N., Hendriks, C., and Meida, L. O. (2001). Carbon Dioxide Emissions from the Global Cement Industry. *Annual Review of Energy and the Environment*, 26(1):303–329.

A Construction of Emissions Rates

Over half of the emissions from clinker production come from the chemical reaction that occurs when the calcium carbonate in limestone is converted into lime and carbon dioxide. To measure carbon dioxide emissions from calcination accurately, emissions factors can be determined based on the volume of the clinker produced and the measured CaO and MgO contents of the clinker. In the absence of this detailed plant-level information, we assume a default rate of 0.525 metric tons of carbon dioxide/metric ton of clinker (WBC, 2005).

The other major source of carbon dioxide emissions from clinker production is fossil fuel combustion. The preferred approach to estimating CO_2 emissions from fuel combustion requires data on fuel consumption, heating values, and fuel specific carbon dioxide emission factors. Although the Portland Cement Association (PCA) does collect plant level data regarding fuel inputs and fuel efficiency (i.e. BTUs per ton of cement), these data are disaggregated data are not publicly available. We do have data aggregated by kiln type and vintage. We use these data (reported in 2006), together with average carbon dioxide emissions factors, provided by the U.S. Department of Energy, to estimate kiln technology specific emissions intensities.

We consider three classes of kilns in particular: wet process kilns (i.e. older, less efficient technology), dry process kilns with preheater/precalciner, and a best practice energy intensity benchmark (Coito et al., 2005)[34] Because of the dominant role played by coal/pet coke, our benchmark emissions calculations are based on coal/petcoke emissions factors. We assume an emissions factor of 210 lbs carbon dioxide/mmbtu.[35]

Our technology-specific emissions rate calculations are explained below. To put these numbers in perspective, the national weighted average emissions rate was estimated to be 0.97 tons carbon dioxide/ton cement in 2001 (Hanle et al, 2005).

Wet process In 2006, there were 47 wet process kilns in operation. On average, wet kilns produced 300,000 tons of clinker (per kiln) per year. The PCA 2006 Survey reports an average fuel efficiency of 6.5 mmbtu/metric ton of clinker equivalent among wet process kilns.

[34]The industry has slowly been shifting away from wet process kilns towards more fuel-efficient dry process kilns. On average, wet process operations use 34 percent more energy per ton of production than dry process operations. No new wet kilns have been built in the United States since 1975, and approximately 85 percent of U.S. cement production capacity now relies on the dry process technology.

[35]Fuel-specific emissions factors are listed in the Power Technologies Energy Data Book, published by the US Department of Energy (2006). The emissions factors (in terms of lbs CO2 per MMBTU) for petroleum coke and bituminous coal are 225 and 205, respectively. Here we use a factor of 210 lbs CO2/MMBTU. This is likely an overestimate for those units using waste fuels and/or natural gas.

The relevant conversion is then 0.095 metric tons carbon dioxide/mmbtu * 6.5 mmbtu/metric ton of clinker equivalent = 0.62 tons carbon dioxide/ton clinker. When added to process emissions, we obtain our estimate of 1.16 tons carbon dioxide/ton clinker.

Dry process In 2006, there were 54 dry kilns equipped with precalciners with an average annual output of 1,000,000 tons of clinker per year. The PCA 2006 Survey reports an average fuel efficiency of 4.1 mmbtu/metric ton of clinker equivalent among dry process kilns with precalciners. Thus, 0.095 metric tons carbon dioxide/mmbtu * 4.1 mmbtu/metric ton of clinker equivalent = 0.39 tons carbon dioxide/ton clinker. Adding this to process emissions results in the estimate for dry-process kilns: 0.93 tons carbon dioxide/ton clinker.

Frontier technology To establish estimates for new entrants, a recent study (Coito et al, 2005) establishes a best practice standard of 2.89 mmbtu/ metric ton of clinker (not clinker equivalent). The calculation is then: 0.095 metric tons carbon dioxide/mmbtu * 2.89 mmbtu/metric ton of clinker equivalent= 0.275 tons carbon dioxide/ton clinker. Adding this to process emissions obtains in 0.81 tons carbon dioxide/ton clinker for new kilns.[36]

B Abatement response

In the simulation exercise, the state space is modified such that emissions rates vary systematically across plants of different vintages and technology types. Incumbent firms are classified as either wet-process, dry-process, or dry-process with precalciner/preheaters. New kilns are assumed to be state-of-the-art. This modification allows us to crudely capture changes in embodied emissions intensity as the industry evolves.

There are four main strategies for reducing the carbon intensity of domestic cement industry. First, it is anticipated that capital stock turnover will be a major driver of emissions intensity reductions (Worrell, 1999). Replacing old wet-process kilns with state-of-the-art dry kilns could deliver significant reductions in combustion-related emissions.

Second, the carbon intensity of clinker production can also be reduced via fuel switching. Currently, coal and petroleum coke are overwhelmingly the dominant fuel used in pyroprocessing and electricity is used to grind raw materials into kiln feed. Most domestic kilns are

[36]This is very similar to the CO2 emissions rate assumed in analyses carried out by California's Air Resources Board in 2008 under a best practice scenario that does not involve fuel switching. If fuel switching is assumed, best practice emissions rates drop as low as 0.69 MT CO2/ MT cement. See NRDC Cement GHG Reduction Final Calculations.

capable of burning a variety of fuels in principle, although fuel switching can adversely affect plant performance.

Third, concrete manufacturers have the capacity to partially substitute SCMs for clinker inputs. The advantage of this emissions reduction strategy is that, by reducing the use of clinker, carbon emissions from both fuel combustion and calcination are eliminated. Finally, cement manufacturers have some capacity to substitute less carbon intensive raw materials for limestone.

Data limitations will prevent us from being able to model input and fuel substitution capabilities accurately at the plant level. In our model, these two abatement options are ignored. In the policy simulations, carbon dioxide emissions from the domestic cement industry can be reduced via four channels: accelerated capital turnover (i.e. retirement of older kilns and investment in newer, more efficient operations), a reallocation of production from more to less emissions intensive incumbents, an increased reliance on imports, and a decrease in domestic clinker consumption. To the extent that fuel and input substitution are economically viable and cost effective compliance alternatives, our results will over estimate compliance costs and thus should be interpreted as upper bounds.

C Sensitivity to elasticity of demand

Table 5: Differences in welfare with respect to baseline (W3) for different demand elasticities

elasticity and mechanism id	Carbon price ($) 5	15	21	30	35	45	52.5	60	65	75
1.5										
1	-139.478	-616.853	-1185.341	-1614.641	-1882.340	-2103.921	-2827.993	-2995.990	-2707.455	-2015.807
2	-134.382	-504.228	-980.765	-1470.348	-1654.099	-1780.400	-1790.835	-1701.828	-1581.148	-1325.378
3	-109.632	-180.570	-337.968	-667.447	-680.070	-580.581	-395.461	-334.922	-128.320	423.483
4	-117.862	-209.866	-377.389	-721.102	-750.347	-725.639	-629.951	-526.834	-277.713	114.588
2										
1	-119.119	-537.696	-758.321	-1084.368	-1297.710	-1502.503	-2197.092	-1768.460	-1378.108	-472.528
2	-119.005	-281.437	-718.176	-1103.484	-1183.358	-1149.994	-982.569	-692.665	-417.885	234.868
3	-100.749	-141.834	-152.292	-404.098	-366.405	-99.973	17.724	425.381	777.822	1434.494
4	-103.269	-151.355	-157.076	-433.629	-414.509	-110.604	-34.334	472.353	794.863	1390.767
2.5										
1	-153.158	-477.799	-465.576	-849.966	-1063.023	-1502.244	-1516.432	-869.830	-316.720	931.443
2	-122.904	-318.104	-474.434	-768.481	-808.152	-659.777	-355.004	40.098	499.066	1561.318
3	-103.813	-157.441	-243.149	-144.436	-65.827	120.996	546.881	1177.858	1487.736	2321.121
4	-106.428	-164.509	-241.925	-153.593	-81.790	102.907	603.383	1006.238	1357.862	2224.709
3										
1	-163.080	-426.622	-392.892	-726.721	-953.510	-1327.531	-761.627	98.747	713.740	2071.397
2	-144.635	-271.177	-380.650	-530.960	-660.347	-367.053	62.108	746.960	1339.337	2613.586
3	-130.472	-157.422	-195.500	42.525	160.937	502.394	985.083	1540.825	1941.244	2829.156
4	-133.591	-153.374	-194.176	-30.847	168.734	503.162	995.354	1484.733	1888.582	2753.995
3.5										
1	-187.764	-420.126	-411.107	-766.851	-766.052	-964.126	-217.858	670.163	1270.661	2569.615
2	-169.232	-302.172	-417.110	-520.374	-479.850	-59.862	432.515	1256.479	1935.232	3332.147
3	-150.619	-179.654	-203.744	-4.460	149.764	655.191	1135.874	1760.390	2215.660	2945.079
4	-151.375	-172.529	-199.176	20.120	142.931	652.627	1118.989	1724.758	2196.141	2924.332
4										
1	-133.274	-373.008	-417.088	-677.523	-630.093	-518.980	300.679	1087.422	1816.465	3348.917
2	-115.585	-262.352	-429.055	-366.156	-363.906	155.486	859.938	1801.899	2543.531	4063.433
3	-99.707	-126.429	-134.597	90.778	221.678	817.099	1366.364	2042.093	2377.690	3366.869
4	-100.352	-125.075	-134.900	80.096	208.503	814.127	1329.342	1996.302	2306.962	3365.637

Notes: Table reports average differences in welfare for a subset of regional markets with three or less firms (Cincinnati, Minneapolis, Pittsburgh, Salt Lake City, Seattle). Mechanisms id: 1) Auctioning, 2) Grandfathering, 3) Emissions updating, and 4) Output updating.

Information Disclosure as a Matching Mechanism: Theory and Evidence from a Field Experiment*

Steven Tadelis
UC Berkeley

Florian Zettelmeyer
Northwestern University and NBER

July 6, 2011

Abstract

Market outcomes depend on the quality of information available to its participants. We measure the effect of information disclosure on market outcomes using a large-scale field experiment that randomly discloses information about quality in wholesale automobile auctions. As the theoretical literature predicts, information disclosure increases expected revenues. However, in contrast with conventional theories, the biggest gains are for the best- and worst-quality cars. We argue that information disclosure causes better matching of heterogeneous buyers to different quality cars. This novel explanation both rationalizes patterns in our data and is confirmed by additional tests. Our findings have implications for the design of other markets, including online consumer auctions, procurement auctions, and labor markets.

JEL classifications C93, D44, D82, L15

*We are grateful to the management and employees of the firm that provided the data and worked cooperatively with us to implement the experiment. We thank Meghan Busse and Igal Hendel for helpful discussions, and many seminar participants for comments. Tadelis thanks the National Science Foundation for financial support.

1 Introduction

A market's efficiency critically depends on whether its participants have sufficient information about the nature of the goods and services being traded. The potential hazard a buyer faces when trading in markets with information asymmetries often leads to market imperfections and stifles efficient trade.[1] Indeed, in resale, housing, labor, health care, and corporate securities markets, sellers may have better information than buyers about the good or service being traded. Furthermore, sellers may have control over how much information to disclose, and buyers may choose how much information to acquire.

This paper studies the effects of information disclosure on market outcomes. We investigate the wholesale market for used automobiles where trade between car dealers is facilitated through auctions. Sellers typically have more information about the condition of the used vehicle than buyers do, and sellers can control the amount of information that they choose to disclose. Using a randomized field experiment, we are able to precisely document how more information affects auction outcomes. We quantify the changes in consummated trades and how these differ across quality levels of the cars sold. Preliminary findings show that ex ante information does indeed affect market outcomes, but in ways that are inconsistent with the standard theoretical literature. We then argue that information plays a role in matching buyers with goods, which not only rationalizes our preliminary findings, but is further confirmed in the data.

Studies of auction design usually focus on auction rules (open or sealed, first or second price, free entry or invited bidders, etc.) rather than on how much information a seller should disclose. A notable exception is the celebrated "Linkage Principle" identified by Milgrom and Weber (1982). They show that under sensible conditions, a seller who commits to disclose more information before his auction can expect revenues from the auction to increase. Two policy implications emerge. First, sellers benefit from committing to disclose as much credible information as they can. Second, auction formats that disclose more information (e.g., open auctions) generate higher expected revenues compared to auctions that do not (e.g., sealed-bid auctions.)

The intuition behind the Linkage Principle is subtle because information disclosure can either increase or decrease a buyer's valuation. If the information discloses bad news relative to expectations then it will cause valuations to drop, just as good news will cause them to increase. As a result, relative to the scenario without information disclosure, bids must be lower following bad news and higher following good news, and with correct expectations this should imply a wash for each individual bidder. However, Milgrom and Weber showed that when the valuations of the bidders are affiliated, information disclosure causes bidders to have more aligned views of the object's value.[2] This in turn increases com-

[1] It is also well known that in some cases, more information can hamper the efficiency of markets. See, e.g., the seminal work of Hirshleifer (1971).

[2] "Roughly, [affiliation] means that a high value of one bidder's estimate makes high values of the others' estimates more likely." (Milgrom and Weber (1982), p. 1096.)

petition (reduces bidders' "information rents"), resulting in higher expected revenues for the seller. In related work, Ottaviani and Prat (2001) explored the incentives of a monopolist to disclose information, and described market conditions where a force similar to the Linkage Principle occurs.

We measure the effect of information disclosure using a unique randomized field experiment. In a market where thousands of vehicles are sold each week with an average value of $8,500, we manipulate information disclosure while keeping all other aspects of the auction fixed. The disclosed information has a clear ranking of quality, which allows us not only to test whether average revenues change, but also measure how average revenues change for *each* quality rank. This enables us to show that the more refined predictions of standard information disclosure theories regarding the effect of good and bad news are violated by patterns in our data.

Specifically, our preliminary empirical findings show that information disclosure causes expected revenues to increase on average, and to (weakly) increase for all quality levels. More striking is the fact that the strongest positive effect follows disclosure of the *very best* and *very worst* quality scores. Further analysis shows that this holds even when the information disclosed changes initial expectations. If, given the observable characteristics of the vehicle, the information disclosed is consistent with expectations, then information disclosure has no effect on auction outcomes. However, if the information disclosed is either *better or worse* than expected, we observe a strong positive effect on expected revenues.

This surprising observation guides our theoretical contribution, which to the best of our knowledge has not been explored in the literature. We argue that in addition to potentially increasing competition *within* a given auction or market, ex ante information disclosure increases competition *across* auctions or markets. We illustrate a simple model where goods of different quality are randomly offered for sale in different auctions, and heterogeneous bidders need to choose in which auction to participate. Higher quality is more valued by all bidders, but the type of bidder who values the good most depends on the quality of the good, as shown in Board (2009). Thus, despite vertical differentiation of goods, buyer heterogeneity causes horizontal differentiation. Information disclosure then helps buyers choose which auction to participate in, effectively matching them with the goods for which they have a high value relative to other bidders.[3] This in turn intensifies the effective competition in any given auction by increasing the number of *high-value* bidders. As a consequence, both the number of efficient transactions and the expected revenue for sellers will increase.

Our proposed matching effect of information disclosure contributes to the theoretical literature described earlier. Moreover, the simple matching model we construct in section 5 explains our preliminary empirical findings. When disclosed information coincides with expectations given observables, then it does not affect the composition of bidders who bid on the vehicle, and as a consequence, the outcomes are the same as they would be without information disclosure. However, when the information disclosed

[3]As an executive in the company that provided the data commented, "one man's trash is another man's treasure."

is either a positive or negative surprise relative to expectations, it will attract bidders who are relatively strong given the disclosed information. This benefits the seller regardless of whether information is good or bad news. We conclude the analysis with a series of tests that both confirm the assumptions of our model as well as additional predictions derived from it.

We also contribute to the growing empirical literature on the effects of information disclosure on market outcomes in general,[4] and on auctions in particular. Due to the challenge of testing how variation in information disclosure affects auctions in the field, there have been few such studies. De Silva et al. (2008) exploited a policy change in the laws of the state of Oklahoma that led to the disclosure of internal costs estimates to complete highway construction projects. They showed that average bids fell after the change in policy, consistent with the prediction of the Linkage Principle (because this is a "reverse" auction, a drop in cost-bids is like an increase in revenue.) Cho, Paarsch, and Rust (2010) used a field experiment where auction formats varied and showed that, consistent with the Linkage Principle, the expected revenues of an open-outcry, English auction are higher than those of auction formats that reveal less information. They did not, however, exogenously vary the amount of information that is disclosed to sellers. There is also a body of work, including Kagel and Levin (1986), Kagel et al. (1987), and Levin et al. (1996), which implemented laboratory experiments that directly and indirectly tested the Linkage Principle. By manipulating the information that bidders receive, or the auction formats, they showed that more information disclosure results in higher average revenues.[5]

The paper proceeds as follows. Section 2 describes the industry, the details of the auctions, and the information provided to bidders. Section 3 describes the data and the experimental design, while section 4 presents preliminary findings that are inconsistent with standard information disclosure theories. Section 5 discusses existing theory and offers a simple model of information disclosure as a matching mechanism, which rationalizes the preliminary findings of section 4. Section 6 shows empirical results that further confirm the implications of our model. Robustness tests are performed in section 7, and section 8 concludes.

2 Wholesale Auto Auctions

The U.S. retail market for used-cars is sizeable. Estimates place used car sales at more than 35 million cars in 2009, most of which were sold by franchise or independent dealers.[6] Dealers of used cars sell on

[4] See, for example, Porter (1995); Jin and Leslie (2003); Cutler, Huckman, and Landrum (2004); and Lewis (2010).

[5] Goeree and Offerman (2002) explored the effect of information disclosure where there are both private and public value components. They showed that when the seller's information is relatively accurate, information disclosure increases efficiency and revenues.

[6] See the National Independent Automotive Dealer's Association (NIADA) website (http://www.niada.com/) for its 2010 annual report. Sales in 2008 and 2009 were similar, down from more than 42 million vehicles sold in 2006 due to the economic downturn.

the retail market and generally purchase their inventory of used cars either from trade-ins, or from the wholesale market for used automobiles.

Wholesale automobile auctions provide a prominent source of used cars. According to the National Automobile Dealers Association (NADA), 35 percent of all used cars sold by new car dealers in 2008 were sourced in auctions.[7] Most auctions are administered by a few prominent auction houses that specialize in this market, one of which provided the data for this study.

2.1 The Auction Process

Buyers at our auctions are exclusively dealers, while sellers mainly belong to one of three categories: dealers who sell used cars from their inventory; owners of large fleets, such as rental car agencies, who periodically turn over their inventory; and financial lease agencies who sell vehicles for which a lease contract has ended. Sellers bring their vehicles to the auction site one or more days in advance of the auction. Each vehicle is assigned "lane" and "run" numbers. Several thousands of vehicles may be auctioned off during a sale day. The vehicles are lined up in several (up to twelve) lanes, according to the lane and run numbers.[8]

Before the auction day begins, potential bidders receive a list of vehicles that will be auctioned, including the lane and run numbers, as well as basic information about the vehicle such as make, model, model-year, options, color, and mileage. This allows buyers to determine which cars they want to bid on. The information is available online before the auction commences, and a printout is prepared for buyers on the morning of the auction.

Each lane has an auction block where an auctioneer conducts the auction, one car at a time for that lane, so that up to twelve auctions can occur simultaneously. The vehicle that is next in line to be sold is driven to the auction block, where it stops amid several potential buyers and is left idling as the auctioneer begins the auction.[9] The auction is an ascending oral (English) auction that lasts for about thirty seconds and ends when no bidder is willing to raise the price.[10] If the price exceeds the seller's reserve price, the sale is consummated. About half the vehicles do not sell on any given auction day because their reserve price is not met. In many of these cases the seller keeps the vehicle at the site, which the auction house offers at no charge, to be auctioned later in the week or during following

[7]See NADA DATA (2009), available at http://www.nada.org/Publications/NADADATA.

[8]For example, a vehicle with a lane-run number of 9-132 will be auctioned in lane 9, and will be the 132$^{\text{nd}}$ vehicle in the lane. Every auction takes about thirty seconds, implying that this vehicle will be offered for sale about sixty-six minutes after the auctions starts.

[9]Some cars that are not in driving condition are towed.

[10]Interestingly, the auctioneer begins at a very high price, often above the winning bid, and then works his way down until some bidder signals his willingness to buy. This sounds like a Dutch auction but it is not: the first bid is not the winning bid, but instead determines the start of the ascending bid process. This procedure has been in place for decades (see Genesove, 1995, p.26), and we have been told that it is also common in livestock auctions.

weeks.[11]

There is a major difference between the way fleet-sellers and dealer-sellers set reserve prices. Fleet-sellers will sell a large number of cars in one sale day (we witnessed one lease agency bring in over 800 cars), and will have a representative sitting with the auctioneer and determining in real time whether or not to accept the highest bid. This suggests that the reserve price may have some real-time input. Dealer-sellers, however, bring in a handful of cars and are seldom present at their cars' auctions. They determine their reserve prices in advance and convey them secretly to the auction house. The auction house then informs the high bidder if the sale is accepted.

There are two distinct classes of bidders at the auction. "Lane" bidders are those bidders who are physically present at the auction and can visually inspect the car up close. Prior to the bidding, vehicles are parked outside so that potential bidders who arrive early enough can examine their exterior condition. The second class of bidders are "online" bidders who are able to participate in the auction through an Internet webcast, which provides streaming audio and video of the auction in real-time. These bidders have online access to basic information about the vehicle, e.g., make, model, year, color, mileage, and other features.

2.2 Information and Standardized Condition Reports

As the description above suggests, buyers have some information about the vehicle at the time of the auction, including basic information and, for the lane bidders, the potential to visually inspect the car and listen to the engine of those cars that can be driven. Because potential buyers cannot perform a serious inspection of the vehicles (not to mention the disadvantage of the online bidders, who cannot see the vehicles in any detail), there is residual uncertainty about a vehicle's quality. As a response, many auction houses offer some form of condition reports that describe in more detail what the condition of the vehicle is. Historically, fleet-sellers have requested some tailor-made condition reports for the cars they sell, but dealer-sellers have not followed suit. Also, the output from these reports was not standard, and buyers were not always pleased with the way in which information was presented.

In response, the auction house from which this paper's data originates developed a Standard Condition Report (SCR) designed to offer a standard set of inspections, and a standard way in which to present the information. The SCR is based on a detailed inspection that takes about twenty minutes per car. The inspections cover the vehicle's exterior condition, documenting all imperfections (including whether there is an additional layer of paint that implies some previous damage). The interior condition is also carefully documented, as is any visual damage to the chassis. The inspections *do not include* the mechanical condition of the car, except that the inspecting technician documents unusual

[11] The seller can also return the car to his lot. For the vehicles in our sample, 70 percent are consigned only once, 17 percent twice, 7 percent three times, 3 percent four times, 2 percent five times, and 1 percent more than five times.

engine sounds. The technician enters all the information through a computerized hand-held device that registers the information on a central computer, and creates a standardized report.

The SCR is then posted online in a standard one-page format. Aside from documenting a detailed summary of the inspection, two other summary statistics are generated. First, a "condition score" (CS) is calculated based on the input of the inspection.[12] The grading system runs from 1 through 5, with increments of 0.1, where $CS = 1.0$ is considered "rough," and $CS = 5.0$ is considered "clean." Second, the SCR calculates the expected number of labor hours needed for a body-shop technician to correct the reported damage, as well as the cost of the materials needed. Using a standard hourly labor rate, this translates into the cost of bringing the vehicle to a condition where exterior and interior damage are no longer noticeable. Hence, both the condition score and the estimated costs offer standardized measures of vehicle quality.

3 Data

3.1 Experimental Design

The purpose of the experiment was to measure the treatment effect of SCRs on the probability of sale and final price for cars that were consigned to the auction by used-car dealers. A subset of all dealer-consigned cars was inspected at one auction location over the course of nineteen weeks using the SCR inspection procedure. Inspected cars were randomly assigned to one of two conditions. In the treatment condition, the SCR of an inspected car was made available to buyers (and sellers). In the control condition, the SCR was withheld; only the auction house knew that these cars had been inspected and their corresponding SCRs.

Due to a limited number of certified vehicle inspectors, not all dealer-consigned cars were inspected. The number of inspected cars depended on the number of available inspectors during that week (between three and twelve). For an auction conducted on Wednesday of a given week, all cars that were checked in starting Friday morning of the prior week were candidates for inspection. On days with many inspectors, all cars that were checked in until mid-day Tuesday were inspected, whereas on days with few inspectors inspections were performed on cars that were checked in until some time on Monday. Specifically, out of approximately 1,500 dealer-consigned vehicles that were registered each week, between 150 and 600 cars were inspected per week (see Table 15). In total, 8,098 cars were inspected, 3,980 of which were in the control group (SCR not reported) and 4,118 were in the treatment group (SCR reported).

Cars were assigned to treatment and control groups during the check-in process. Cars whose VIN (vehicle identification number) ended in an even digit were assigned to the treatment group, while those with an odd digit were assigned to the control group. The first digits of a VIN number designate

[12] Genesove (1993) and Overby and Jap (2009) also investigated the role of condition reports.

manufacturer, country of origin, make, model, model-year, as well as some trim-level information, whereas the later digits are assigned sequentially as vehicles are produced. Hence, the last digit of the VIN is a good randomization device: whether the digit is even or odd is unrelated to the condition of the vehicle. Also, even and odd digits are equally represented in the population of produced cars. We thus expected an approximately even split between treatment and control groups. Consistent with this, the randomization procedure assigned 49.15 percent of cars to the control group and 50.85 percent to the treatment group.[13]

As we analyzed auction outcomes after the first nine weeks of the experiment, we found little evidence that cars with SCRs were more likely to sell or sold at higher prices (these findings are described in section 4). One possible explanation for the weak results could be that the information contained in SCRs had little content. Another explanation, however, could be that dealers did not know that SCRs were made available for a significant number of dealer-consigned cars. As discussed in the previous section, SCRs are available only online, not on the standard printout that dealers can obtain in advance on the website or on auction day at the facility. Hence, for the remainder of the experiment a weekly email was sent to all registered buyers informing them that they could find SCRs for some of the dealer-consigned cars on a particular website prior to the auction day.[14] As a result, our experiment covers two periods: weeks 21-30 (5,402 cars), during which dealers were not likely to have been aware of the existence of SCRs, and weeks 31-39 (2,696 cars), during which SCRs were publicized. This variation will prove useful in analyzing the data and shedding light on the impact of information disclosure.

3.2 Auction and Inspection Data

For each consigned car we observed the model; model-year; body type; engine and trim level (e.g., a Honda Accord, 1999, 4-door, V6, EX trim); as well as its mileage. More detailed information about the condition of the car came from the SCR as described in section 2.2. We used two key measures. The first measure is the condition score, a number between 1 (rough) and 5 (clean). The second measure is the estimated cost to fix the damage detailed in the SCR. This includes the auction house's estimates of both part and labor costs and is reported in dollars.

We observed a unique seller ID allowing us to identify whether different cars were consigned by the same seller. The data reports whether a car was sold during the auction, the final auction price, and a unique buyer ID allowing us to identify whether different cars were purchased by the same buyer. Finally, we had the average auction price for cars of the same car type that sold at any of the auction

[13]We cannot reject the hypothesis that our randomization procedure assigned an equal proportion of cars to treatment and control groups (at a 5 percent significance level).

[14]The emails stated that the company was ramping up its capabilities to offer SCRs, and therefore, technicians were assigned to inspect a subset of vehicles that were chosen randomly based on the availability of inspection technicians. It was made clear that these were not solicited or affected by the sellers.

house's locations nationwide during the prior week (henceforth "National Auction Price" or NAP). This allowed us to construct a useful normalization of price that was independent of the type of car. Summary statistics are reported in Table 16.

3.3 Randomization Check

We compared the treatment and control groups on a variety of observable characteristics. Specifically, if the randomization worked as intended, the distribution of condition scores, repair costs, mileage, vehicle age (model year), and national auction prices in the prior week should have been comparable across control and treatment groups. We used a Kolmogorov-Smirnov test for equality of distribution functions. The results are reported in Table 1.

Table 1: Kolmogorov-Smirnov test for equality of distribution functions

Variable	D	p-value
Condition score	0.0137	0.83
Repair costs	0.0301	0.05
Mileage	0.0172	0.58
Model year	0.0167	0.61
National Auction Price	0.0246	0.17

For four of the five measure we failed to reject the hypothesis that the distribution functions were the same. However, the test statistic for repair costs was just at the critical level, indicating that repair cost may have had a different distribution between control and treatment groups. We compared the means of repair costs across the two conditions. Repair costs for the control group were on average \$1,382, while for the treatment group they were \$1,316. We will account for this \$66 (less than 5 percent) difference when interpreting our auction price results.

In addition to comparing the treatment and control groups on a variety of observable characteristics, we will later explore our randomization when estimating treatment effects. We will analyze whether our estimates change as we add a large set of controls, namely fixed effects for seller ID, model year, vehicle segment, nameplate, and sale week, as well as measures of the condition of the car. If our randomization procedure worked then these controls should not substantially change our estimates, because the randomization should have ensured that cars of the same make, model year, segment, and approximate condition were randomly distributed between treatment and control groups. The results of this analysis are reported in section 7, where we explore the robustness of our findings.

4 Preliminary Findings

We organize the preliminary findings into three parts. First, in section 4.1 we report the aggregate findings of our experiment and show that more information increased the likelihood that cars sell, and that, conditional on selling, they sold for a slightly higher price. Second, in section 4.2 we show how the results vary by condition score, and whether the condition score was better or worse than expected. We then argue that standard information disclosure theory is inconsistent with these results.

4.1 Aggregate Findings

A car's expected revenue comprises the probability that it will sell (the reserve is met), and the price conditional on a sale. We consider each of these components separately.

Table 2 shows that during weeks 21-30, cars with and without a posted SCR were equally likely to sell; approximately 43 percent of cars sold in either condition. This suggests either that SCRs had no effect or that buyers were unaware of SCRs. During weeks 31-39, when the availability of SCRs was announced with a weekly email, cars with a posted SCR were 6.3 percentage points (or 16 percent) more likely to sell than cars without a posted SCR. This difference is highly statistically different from 0 (using a test of proportions with p-value < 0.01). One concern in evaluating the statistical significance of this difference is that 30 percent of cars in our sample were offered for sale more than once during the sample period. As a result, the error terms for cars that were offered multiple times could be correlated. To account for this potential correlation, we clustered the standard errors at the VIN level (for detailed results see section 7, where we explore the robustness of our results). The correction had a very small effect and does not alter our conclusions in Table 2 or any other table in section 4.

Table 2: Sales probability by experimental condition

	No posted SCR	Posted SCR	Difference	% Difference	z-statistic	p-value
Weeks 21-30	0.43 2,605 cars	0.436 2,797 cars	0.006	1.39%	0.43	0.66
Weeks 31-39	0.392 1,375 cars	0.455 1,321 cars	0.063	16.1%	3.31	0.001

Prices in the two experimental conditions were not significantly different, in either period. Table 3 shows these results.

A problem in concluding that transaction prices did not differ between experimental conditions is that the prices variance of sold cars is high. This is because sales include everything from 11-year-old small cars (e.g., Honda Civic) to current-year luxury cars (e.g., BMW 740). Ideally, we should specify prices relative to the typical price for cars of the same car type, i.e., of the same make, model, and

Table 3: Transaction prices by experimental condition

	No posted SCR	Posted SCR	Difference	% Difference	t-statistic	p-value
Weeks 21-30	\$8,742.9 1,121 cars	\$8,616.9 1,220 cars	-\$126.0	-1.4%	-0.51	0.61
Weeks 31-39	\$8,502.2 539 cars	\$8,738.9 601 cars	\$236.7	2.7%	0.68	0.50

model-year. To do this we used the average auction price for cars of the same car type that sold at any of the auction house's locations during the prior week, what we referred to earlier as the National Auction Price (NAP). We used this measure to construct a normalized price for each car in the sample, specifically, the price of the car divided by the NAP. This allowed us to reevaluate whether there were price differences between experimental conditions. Table 4 shows these results.

Table 4: Transaction prices/NAP by experimental condition

	No posted SCR	Posted SCR	Difference	% Difference	t-statistic	p-value
Weeks 21-30	1.064 1,106 cars	1.058 1,202 cars	-0.006	-0.5%	-0.56	0.58
Weeks 31-39	1.035 531 cars	1.055 590 cars	0.02	1.9%	1.61	0.11

After week 31, prices were higher by 1.9 percent for cars with a posted SCR relative to cars without a posted SCR. The difference, however, is only marginally significant (p-value of 0.11).

In interpreting these results one might be concerned that bidders could have responded to something other than the informational content of SCRs. Of particular concern is that the emails from week 31 onward just focused buyers' attention on cars with posted SCRs (i.e., a "salience" effect.) We will later show that bidder behavior was not consistent with such a salience effect. Instead, it seems that bidders were reacting to the information contained in SCRs, as opposed to the mere existence of SCRs (see section 7.3 for a detailed analysis of this issue).

Overall, an analysis of the probability of sale and prices conditional on sale suggests that most of the effect of SCRs on expected auction revenues comes from an increased probability of sale; transaction prices did increase, but only by a little.

4.2 Decomposing the Effects

We now investigate how the effect of posted SCRs differs by the condition of the vehicle. As before, we decomposed the auction revenue effect into a sales probability and price effect.

4.2.1 Transactions by Condition Scores

To analyze the effect of posted SCRs by condition of the vehicle, we split our sample into condition score terciles. The terciles contain cars of below-average, average, and above-average condition, respectively. Table 5 reports the effect of posted SCRs on sales probabilities by condition score tercile.[15] To assess the statistical significance of the sales probability differences, we restricted ourselves to weeks 31-39 and used a test of proportions. For cars of *below*-average and *above*-average condition we conclude that a posted SCR was associated with a higher sales probability. However, the effect of a posted SCR for cars of *average* condition was clearly too small to be considered different from 0.

Table 5: Sales probability by condition score (CS), weeks 31-39

Tercile of CS	# of Cars	No posted SCR	Posted SCR	Difference	% Difference	z-statistic	p-value
Worse than average	928	0.35	0.427	0.077	23.3%	2.40	0.016
Close to average	888	0.427	0.442	0.015	3.5%	0.45	0.65
Better than average	880	0.4	0.495	0.095	22.0%	2.79	0.005

We can also assess the price effect of posted SCRs by condition score tercile. Using a *t*-test, Table 6 compares prices by condition score during weeks 31-39. We conclude that a posted SCR is associated with a significantly higher auction price only for below-average condition cars.[16]

Table 6: Price/NAP by condition score (CS), weeks 31-39

Tercile of CS	# of Cars	No posted SCR	Posted SCR	Difference	% Difference	t-statistic	p-value
Worse than average	327	0.985	1.03	0.049	5.0%	1.98	0.05
Close to average	371	1.07	1.086	0.016	1.5%	0.76	0.45
Better than average	423	1.07	1.08	0.018	0.2%	0.1	0.92

The fact that cars with relatively low disclosed condition scores experienced a *positive effect* on sales (and on prices) seems to be inconsistent with standard information disclosure theories. For example, the Linkage Principle states that more information increases expected revenues *unconditional* on the actual quality level. However, *conditional on bad news*, typical disclosure models predict that bidders *lower* their willingness to pay, making this finding puzzling. We have not, however, established that relatively

[15]Note that the terciles have a slightly different number of cars due to ties in condition scores. We can also split condition scores in other ways with similar results. For example, we have combined cars by the first digit of their condition score, i.e., 1, 2, 3, 4, and 5. The results are similar to what we will show in this section, namely that the expected revenue is affected positively at the extremes of the condition score distribution.

[16]Notice that in Table 9 the number of vehicles in each tercile is quite different. The terciles are defined over all inspected vehicles, while the subsets of sold vehicles are not equally distributed across the three terciles.

low condition scores are in fact bad news. Whether a reported SCR is good or bad news will depend on the expectations that bidders have about the condition of the vehicle before the information is disclosed. Next, we investigate not just whether expected revenues increase for condition scores that are low, but whether expected revenues increase for condition scores that are low *relative to expectations*.

4.2.2 Transactions by Informational Content

Bidders have some information (regardless of whether an SCR is posted) that can predict of the condition score, namely mileage and age. As shown in Table 7, the average condition score varies substantially by vehicle age and mileage, as one would predict: cars that are older or that have higher mileage will, on average, have worse condition scores. This information allows buyers to estimate the condition of the car as a function of age and mileage.

Table 7: Average condition score (CS) by mileage category and vehicle age

Mileage Category	Average CS	Vehicle Age	Average CS
0-20,000	4	1	4.2
20,001-40,000	3.6	2	3.9
40,001-60,000	3.1	3	3.3
60,001-80,000	2.7	4	3.1
80,001-100,000	2.5	5	2.9
100,001-120,000	2.3	6	2.5
120,001-140,000	2	7	2.2
140,001-160,000	1.9	8	2.1
160,001-180,000	1.6	9	2
180,001-200,000	1.3	10	1.9
>200,001	1.4	11	1.8
		12	1.7

As a result, it is necessary to perform an empirical test that explicitly allows for condition score expectations that differ with vehicle age and mileage. To proceed, we first estimated the predicted condition score of each car in our sample based on the vehicle age and vehicle mileage. We made this prediction by regressing condition score on vehicle age year dummies, vehicle mileage, and vehicle mileage deciles. To account for potential interaction effects between vehicle age and mileage, we interacted the vehicle age year dummies with vehicle mileage. We took the difference between the actual condition score and the predicted condition score to construct a distance measure from the expected condition score. Finally, we split this distance measure into terciles, where the bottom tercile contains cars with worse-than-expected condition scores, the middle tercile contains cars with close-to-expected condition scores, and the top tercile contains cars with better-than-expected condition scores.

Table 8: Sales probability by difference of expected condition score (CS), weeks 31-39

Tercile of Difference from Expected CS	# of Cars	No posted SCR	Posted SCR	Difference	% Difference	z-statistic	p-value
Worse-than-expected	899	0.327	0.411	0.084	25.7%	2.61	0.009
Close-to-expected	899	0.429	0.418	-0.011	-2.6%	0.34	0.74
Better-than-expected	899	0.419	0.529	0.109	26.1%	3.28	0.001

As Table 8 shows, during weeks 31-39 there is no statistically significant effect of a posted SCR on the probability of sale for cars in the middle tercile where actual condition scores are close to expected condition scores. However, in both terciles where condition scores have informational content, the effect on the probability of sale is positive and significant.[17]

We replicated this analysis for prices. As Tables 9 shows, the difference in normalized average prices between cars with and without a posted SCR remain positive, but is statistically no longer different from zero at the bottom tercile compared to the findings in Table 6.

Table 9: Price/NAP by difference of expected condition score (CS), weeks 31-39

Tercile of Difference from Expected CS	# of Cars	No posted SCR	Posted SCR	Difference	% Difference	t-statistic	p-value
Worse than expected	324	0.978	0.999	0.022	2.2%	1.05	0.30
Close to expected	375	1.04	1.08	0.035	3.3%	1.58	0.11
Better than expected	422	1.07	1.08	0.006	0.6%	0.31	0.75

As before, the findings are inconsistent with standard information disclosure models. We observed that both good news and bad news caused more sales without reducing prices.

5 Information Disclosure Theory Revisited

This section offers a theoretical explanation for why both good news and bad news caused expected auction revenues to increase. We first review existing theory in more detail. We then outline a simple new theoretical framework that both rationalizes the empirical findings in section 4.2, and suggests additional empirical implications that are confirmed in section 6.

[17]Note that vehicles in the middle and top tercile sell much better even without SCRs being reported. This is consistent with the fact that buyers can cruise the lot before the auction begins, and thus identify characteristics of the vehicle that are informative, but for which we cannot control in our prediction regression.

5.1 The Linkage Principle and the Allocation Effect

The Linkage Principle derived in the seminal work of Milgrom and Weber (1982; henceforth, MR) provides the benchmark of how information disclosure affects auction outcomes. It shows that in a symmetric affiliated values auction setting, a seller who commits to disclose all information ex ante can increase expected revenues.[18] Information disclosure causes the assessments of the bidders to be more congruent, which results in lower "information rents." Still, given a fixed set of bidders, if the information is good news, then expected revenue will increase, while if the information is bad news then expected revenue will decrease.

Our empirical results in section 4 confirm that expected revenues increase. However, they increase even for cars with a condition score of 1.0, the lowest possible quality level, and moreover, controlling for informational content, both bad news and good news increase expected revenues. This empirical finding is striking because any rational expectations model in which bidders have monotonically increasing values in quality implies that disclosing bad news must result in lower values, and hence in lower bids than those with no information disclosure (Milgrom, 1981).

As Board (2009) observed, MR imposed two simultaneous assumptions. First, bidders are symmetric. Second, their valuations are monotonic in the information. As a result, the order of valuations coincides with the order of types, and information disclosure affects the expected price without changing the type who wins the auction. Board (2009) showed that when either of these assumptions are dropped, the Linkage Principle may fail.

As a simple example, imagine that the seller's item has quality q uniformly distributed over $[0, 1]$, and there are two different bidders. The first bidder (H) has a valuation equal to $v_H(q) = q$ and the second bidder (L) has valuation $v_L(q) = 0.25 + 0.5q$. Hence, the H bidder values relatively high quality ($q > 0.5$) more than an L bidder, while the reverse is true for relatively low quality ($q < 0.5$). If the seller discloses no information and uses a second-price auction, each bidder bids his expected value, both equal to 0.5, and revenue is 0.5. If, instead, the seller discloses the realization of q, then bidder 1 bids $b_1 = q$ while bidder 2 bids $b_2 = 0.25 + 0.5q$. Revenue is then $\min\{b_1, b_2\}$, which equals q if $q \leq 0.5$ and $0.25 + 0.5q$ if $q > 0.5$. Expected revenue is then equal to $\frac{7}{16}$, less than the expected revenue without information disclosure. This illustrates what Board (2009) labeled the *allocation effect*, where new information affects which bidder wins the good. Simply put, asymmetry implies a kind of horizontal differentiation across bidders.[19]

The potentially negative impact of the allocation effect is inconsistent with our data, because revenues increase after information is disclosed. Bidder heterogeneity, however, and the implied allocation

[18] To be precise, the information disclosed must be affiliated. That is, once it is revealed, the valuations of the bidders move closer to each other in a statistical sense. See MR.

[19] An earlier example showing the failure of the Linkage Principle with multi-unit auctions was derived by Perry and Reny (1999), yet the underlying forces share much in common.

effect may still be present in our setting. Board (2009) showed that with many bidders, revenues will increase when more information is disclosed. To see this, imagine that there are four bidders, two of type H and two of type L. With no information disclosed, everyone bids 0.5 and revenue is 0.5. If information is disclosed, then *two bids* are equal to $b_H = q$ while *two other bids* are equal to $b_L = 2 + \frac{q}{5}$. The price is then $\max\{b_L, b_H\}$, which equals q if $q \geq 0.5$ and $0.25 + 0.5q$ if $q < 0.5$. The expected revenue is now $\frac{9}{16}$, consistent with our finding that revenues are higher after information is disclosed.

Having more bidders does not, however, resolve the puzzle that even disclosing worse-than-expected quality information yields higher expected revenues than disclosing no information, because $\max\{b_L, b_H\} < 0.5$ for all $q < 0.5$. As implied from Milgrom (1981), if preferences are monotonically increasing in quality, then revenues obtained with no information disclosure can *never be lower* than revenues obtained following bad news.[20] Hence, through the lens of conventional bidding models our empirical findings still beg an explanation.

5.2 Information as a Matching Mechanism: Theory

Most auction models assume that one auction is being conducted at any given time and that the set of bidders at the auction is fixed.[21] Both assumptions are violated in our environment because multiple auctions are conducted simultaneously and bidders have to exclusively choose which of these to participate in. Perhaps, the disclosure of SCRs affects the decisions of bidders regarding *which items* to bid on, if at all. Here we explore this idea by constructing a simple two-type, two-good model to analyze a situation in which heterogeneous bidders choose which of two heterogeneous items to bid on.[22]

Discussions with industry participants suggest that used-car dealers are heterogeneous. Dealers sell to customers in their geographical vicinity, implying that local tastes will shape their value for different vehicles. For instance, high-income consumers will not be interested in a beaten-up (low-quality) vehicle, while low-income consumers cannot afford to be as picky. Dealers from low-income neighborhoods will then outbid their counterparts from high-income neighborhoods on low-quality cars. High-income consumers will pay more for cars in better condition than low-income consumers because

[20] If some preferences are decreasing in quality, then bad information could lead to higher prices. This violates the whole notion of calling q quality because we typically think of quality as a vertical dimension over which preferences are increasing and monotonic. The condition of a car clearly falls in this category of cases.

[21] Some models consider a random number of bidders (e.g., McAfee and McMillan, 1987), while others consider endogenous entry of bidders (e.g., Levin and Smith, 1994). These studies consider one auction, so that the endogenous choice of *which auctions* to participate in, which is the focus of our analysis, has not been considered. That said, Roberts and Sweeting (2011) show that selection effects may have subtle auction design implications even when one auction is considered.

[22] Developing and analyzing a more general formal model is beyond the scope of this paper as it would be a challenging stand-alone theoretical analysis.

their marginal value for appearance is greater. Hence, in reference to the example we use earlier, dealers in low-income neighborhoods seem similar to L type bidders, whereas dealers in high-income neighborhoods seem similar to H type bidders, resulting in horizontal differentiation across quality.

We also learned that auction bidders are quite experienced in assessing the condition of vehicles. Recall that the condition score reflects mostly the exterior and interior condition of the vehicle. By observing the vehicles up close, a relatively quick visual inspection can identify to a large degree whether the condition score ought to be low, high, or somewhere in between. As a consequence, once bidders show up at a lane and see a vehicle, they have a pretty good idea of its condition as measured by the condition score. That is, *conditional* on a bidder showing up at an auction, the information revealed by the SCR is not very discriminating.

These observations suggest that a formal analysis of our environment should include three basic assumptions. First, bidders are heterogeneous and horizontally differentiated with respect to condition scores (A1). Second, there are several goods selling at several *mutually exclusive,* simultaneous auctions (A2). Third, the disclosure of SCRs may help bidders find the vehicles they are interested in, but once they see a vehicle, the information content in the SCR is small (A3).[23] To proceed, we develop a simple, stylized model based on these assumptions as follows:

Preferences: Consider two types of bidders (A1), $\theta \in \{L, H\}$, with $v_H(q) = q$ and $v_L(q) = 0.25 + 0.5q$ as described above and depicted in Figure 1. The quality of vehicles q is random and uniformly distributed on the interval $[0, 1]$. A bidder i of type θ has a value $v_{\theta i}(q) = v_\theta(q) + \varepsilon_i$ where ε_i is a private shock that is independently distributed over $[-\bar{\varepsilon}, \bar{\varepsilon}]$, with $\bar{\varepsilon}$ being very small, the density being everywhere positive and well defined, and $E[\varepsilon_i] = 0$. Hence, the expected value of a type θ bidder from a vehicle of quality q is $E[v_{\theta i}(q)] = v_\theta(q)$. We assume that there are four bidders, *exactly two* of each type.[24]

Mechanism: There are two open ascending auctions on two lanes that sell vehicles simultaneously, and each bidder can only be present at one lane at a time (A2). Quality is independent across vehicles and lanes.

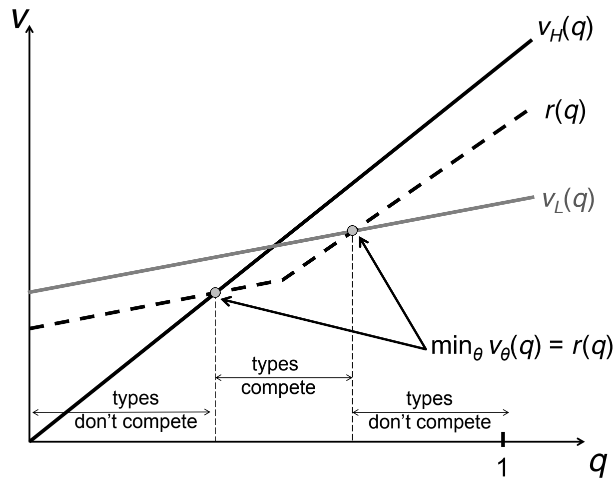
Information: Sellers can either disclose nothing, or they can disclose perfect, verifiable information about quality $q \in [0, 1]$. Once bidders arrive at a lane, they perfectly observe the quality q (A3),

[23] "Small" means that the condition score plays a more important role in matching bidders to cars, and a less important role in revealing information once a bidder arrives at an auction.

[24] This is for computational ease. The qualitative results will hold if there are at least two bidders of each type. Also, the particular choices for $v_\theta(q)$ are not important. What is important is that the value of one type (L) has a greater intercept and a smaller slope, so that the value functions cross at some interior quality level. Note that this is not the same as the single-crossing condition, which only requires a ranking of slopes.

but before choosing which lane to attend, bidders only know what the seller chooses to disclose.[25]

Figure 1: Valuations of two types of bidders



As before, horizontal differentiation in quality is captured by the fact that $v_L(q) > v_H(q)$ for $q < 0.5$, while $v_L(q) < v_H(q)$ for $q > 0.5$. With no disclosure, both types have an expected value of 0.5. Let $v^q_{\min} \equiv \min_\theta v_\theta(q)$, the lower of the two expected valuations, and $v^q_{\max} \equiv \max_\theta v_\theta(q)$, the higher of the two.

Timing proceeds as follows. Bidders observe any information the seller discloses and then choose which lane to participate in. An equilibrium will be characterized by a lane choice, followed by the standard dominant strategy of bidding up to one's valuation in an ascending auction. To make things simple, assume that there are two distinct vehicles, one with quality $q < 0.5$ and the other with quality $q' > 0.5$, and their assignment to one of two lanes is random.

5.2.1 No Disclosure

If the seller does not disclose the quality of the vehicles, we have (proofs appear in the appendix):

Claim 1: With no disclosure there are two equilibrium outcomes: a unique symmetric *random equilibrium* where each bidder chooses each lane with equal probability and a unique asymmetric *coordinated equilibrium* where exactly one bidder of each type is in each lane.

The two equilibria identified in Claim 1 have different outcomes. In the coordinated equilibrium with no disclosure, the allocation is efficient, but the expected winning bid of a quality q vehicle is always $ER[q \mid \text{ND coordinated}] = v^q_{\min}$.

[25] Alternatively, information can takes the form of a partition of the quality interval. For example, a score of $s \in \{1, 2, ..., 5\}$ could correspond to the vehicle's true quality being uniformly distributed in the interval $[\frac{s-1}{5}, \frac{s}{5}]$. The qualitative results and comparative statics that follow would persist.

The random equilibrium, which is the unique symmetric equilibrium, results in sixteen distinct outcomes with equal probabilities, and the expected price is the expected second-highest value of the bidders. There are three configurations of bidders at a lane that are of interest. First, if no more than one bidder shows up, then the price is 0. For any given lane this happens with probability $\frac{5}{16}$. Second, if more than one bidder shows up, but no more than one of them has value $v^q_{\max}$, then expected revenue is $v^q_{\min}$. This happens with probability $\frac{7}{16}$. Last, if more than one bidder shows up, and two of them have value $v^q_{\max}$, then expected revenue is $v^q_{\max}$.[26] This happens with probability $\frac{1}{4}$. Hence, expected revenues of a quality q vehicle in this equilibrium is $ER[q|\text{ND random}] = \frac{7}{16}v^q_{\min} + \frac{1}{4}v^q_{\max} < v^q_{\max}$.

5.2.2 Full Disclosure

If the seller discloses the quality of the vehicles he puts up for sale in each lane, we have:

Claim 2: Given two vehicles with qualities $q < 0.5$ and $q' > 0.5$ auctioned in two lanes with full disclosure, the unique equilibrium has perfect sorting: both L types choose the q-lane and both H types choose the q'-lane.

This is a consequence of optimal sorting. Each type will select into the lane where a comparative advantage exists, and information disclosure acts as a matching mechanism. Given Claim 2, it is easy to see that with disclosure, the expected revenue of each vehicle is $ER[q \mid \text{D}] = v^q_{\max}$.

5.2.3 Comparing Information Policies

This corollary follows from the analysis above:

Corollary: Information disclosure increases expected revenues for any given quality level. The impact is larger as quality is farther away from 0.5, the value at which the two types' valuations cross. Furthermore, with information disclosure the variance of winning bids is lower than in the random equilibrium with no information disclosure.

The intuition for this result is similar to that of the allocation effect identified by Board (2009). If heterogeneous bidders are at a lane and a relatively high-quality vehicle comes through, then the H type wins, while the opposite happens for a low-quality vehicle. What differs in our setting is that the ex ante arrival of information on quality causes bidders to *endogenously choose lanes* where they can win. As a consequence, the composition of bidders at lanes is rearranged to create assortative

[26] Throughout the analysis we ignore the ε_i's, which play a tie-breaking role for identical types by adding some natural idiosyncrasies. The qualitative conclusions are valid as long as $\bar{\varepsilon}$ is not too large. If we don't ignore the ε_i's, then we need to calculate the second order statistic of ε_i to correctly determine the expected price in this situation. However, as $\bar{\varepsilon} \to 0$ the expected price goes to $v^q_{\max}$.

matching, which guarantees that two high-valuation bidders will be present, intensifying competition for any quality level.

The reason that the increase in revenues is "U-shaped" follows from the fact that $v^q_{\max} - v^q_{\min}$ increases as q moves away from the point at which $v_H(q) = v_L(q)$, as illustrated in Figure 1. If the equilibrium play is random, then for any level of q, the seller's revenue will sometimes be $v^q_{\max}$, sometimes be $v^q_{\min}$, and sometimes be 0. With information disclosure, however, the seller always receives $v^q_{\max}$, increasing the expected price and reducing the price variance (to zero as $\bar{\varepsilon} \to 0$). Assortative matching, therefore, has less of an impact for quality levels at which the two types have similar valuations.

5.2.4 The Effect of Reserve Prices

The use of reserve prices by sellers is obviously missing in the analysis above. As described in section 2, about half the vehicles do not sell on any given auction day because their reserve price is not met, and in many of these cases the seller keeps the vehicle at the site (see footnote 11). This results in an "outside option" that is surely not zero.[27] Reserve prices must therefore be considered to correctly predict the effect of information disclosure on auction outcomes.

Imagine that the seller has a small opportunity costs $k > 0$ of keeping the car at the auction house for the next auction, and a discount factor $\delta < 1$ applies for each period of delay between auctions. Consider the random equilibrium above with no disclosure. The seller expects one of three outcomes: a price of zero (with probability $\frac{5}{16}$), a price of $v^q_{\min}$ (with probability $\frac{7}{16}$), and a price of $v^q_{\max}$ (with probability $\frac{1}{4}$). Given these beliefs, the seller will prefer to reject a bid b and wait for $v^q_{\max}$ if,

$$\begin{aligned} b &< \delta(-k + \frac{1}{4}v^q_{\max}) + \frac{3}{4}\delta^2(-k + \frac{1}{4}v^q_{\max}) + \left(\frac{3}{4}\right)^2 \delta^3(-k + \frac{1}{4}v^q_{\max}) + \cdots \\ &= \frac{\delta}{1 - \frac{3}{4}\delta}(-k + \frac{1}{4}v^q_{\max}) \equiv r(q). \qquad (1) \end{aligned}$$

For small opportunity costs (δ close to 1 and k small), the value of $r(q)$ will be somewhat below the "upper envelope" of $v^q_{\max}$, as depicted by the dashed-line in Figure 1.[28]

Our earlier observation that assortative matching has less impact when the two types have similar values implies that there is an important difference between mid-range quality, where $v^q_{\max} - v^q_{\min}$ is small, and between either very low or very high quality levels, where $v^q_{\max} - v^q_{\min}$ is large. As depicted in Figure 1, there are two values of quality at which $r(q) = v^q_{\min}$. In the interval between these values, the two different types are similar enough so that a $v^q_{\min}$ bid is not rejected. The H and L types are

[27] Two other alternatives are available to dealers whose cars did not sell. They have the option of returning the car to their own lot, where there is some chance it can sell, instead of waiting several days at the auction site. Another way to sell a car is using wholesale buyers who visit dealer lots to buy cars that the dealers have a hard time selling and then relocate those cars to other dealers.

[28] As $\delta \to 1$ (low opportunity cost of delay), the limit of $r(q)$ equals $-4k + v^q_{\max}$, and as $k \to 0$ (low opportunity cost of leaving the car at the auction) it equals $v^q_{\max}$. The order of limits is inconsequential.

"competitive enough" to make the option of waiting for $v^q_{\max}$ not worthwhile. In contrast, when quality is outside of this interval, $v^q_{\max} - v^q_{\min}$ is large and the option value of waiting for $v^q_{\max}$ makes rejecting $v^q_{\min}$ bids worthwhile.[29]

With disclosure, however, the probability of obtaining $v^q_{\max}$ goes up to 1, implying that a larger impact on the probability of sale will result when the quality is farther away from the middle. Furthermore, with relatively low opportunity costs, the curve $r(q)$ will be close to $v^q_{\max}$, and *conditional on selling,* the prices with or without information disclosure will be similar. That is, most of the effect will be on the probability of sale.

We conclude that our simple model of information as a matching mechanism is consistent with the results from our experiment.

6 Information as a Matching Mechanism: Evidence

Here we test our theory of information as a matching mechanism. We begin by searching for evidence in the experimental data for a key premise of our model, that dealers horizontally differentiate with respect to condition scores. Next, we test whether information disclosure changes auction outcomes in the way hypothesized in section 5.2.4. Finally, we provide evidence suggesting that information disclosure influences which vehicles bidders chooses to bid on.

6.1 Bidders are Horizontally Differentiated in Condition Scores

To show that bidders are horizontally differentiated with respect to condition scores, we split each dealer's purchases into "early" and "late" car purchases: early purchases comprise the first 50 percent of cars purchased by the dealer during our sample period; late purchases comprise the remaining cars that the dealer bought. We tested whether the condition scores of cars purchased early by each dealer predicts the condition scores of cars purchased late by the corresponding dealer.

We began by calculating for each dealer the average condition score of cars purchased early and late. If dealers specialize in cars of specific condition scores, we would expect the average condition scores between the two samples to be positively correlated. Indeed, for dealers who purchased more than two cars during the sample period, we found that the correlation coefficient is 0.45 (p-value < 0.01). Another way of analyzing specialization is to calculate a transition matrix between the condition scores chosen for early and late purchases. Specifically, for each dealer we measured the average condition score of cars purchased early and late. We split these average condition scores into quintiles and calculated what percentage of dealers who were in a specific quintile for early purchases were in the same quintile for

[29]The variation due to the private noise determined by $\bar{\varepsilon}$ will effect the reserve price strategy, but will still result in some bound below the upper envelope of $v^q_{\max}$. The qualitative comparative static results will still hold.

late purchases. Using 407 dealers who purchased more than two cars during the sample period, Table 10 illustrates the transition matrix.

Table 10: Early vs. late purchase transition matrix

	condition score Quintile	Late purchases 1	2	3	4	5	Total
	1	34	14	9	7	2	66
		51.52%	21.21%	13.64%	10.61%	3.03%	100%
	2	28	21	17	10	10	86
		32.56%	24.42%	19.77%	11.63%	11.63%	100%
Early	3	13	21	24	15	12	85
purchases		15.29%	24.71%	28.24%	17.65%	14.12%	100%
	4	19	7	15	21	30	92
		20.65%	7.61%	16.30%	22.83%	32.61%	100%
	5	8	9	17	15	29	78
		10.26%	11.54%	21.79%	19.23%	37.18%	100%
	Total	102	72	82	68	83	407
		25.06%	17.69%	20.15%	16.71%	20.39%	100%

Clearly, buyers who chose cars of particular condition scores during early car purchases tended to choose cars of similar condition scores during late car purchases as well. These findings are consistent with the assumption of our simple example, namely that bidders are heterogeneous and horizontally differentiated with respect to condition scores.

6.2 How Information Disclosure Changes Auction Outcomes

Our matching theory predicts that information disclosure increases the probability of sale for cars in the bottom (worse-than-expected) and top (better-than-expected) terciles, and not for cars in the middle (close-to-expected) tercile. This is consistent with our earlier results: Table 8 shows that our experimental data for weeks 31-39 follow this prediction. Moreover, because early during the experiment the wide availability of SCRs was not publicized, we should not find the hypothesized pattern during weeks 21-30. Indeed, as Table 11 shows, there was no statistically significant effect of a posted SCR on the probability of sale for cars in any of the terciles.

Overall, our results are consistent with our model's main prediction: there will be a larger positive impact on the probability of sale when the score is farther away from the expected condition score. Furthermore, consistent with the low opportunity costs of keeping the car at the auction site, we found that most of the effect was on the likelihood of sale with a small effect on prices. This latter point can be seen in Table 9 for weeks 31-39 and in Table 12 for weeks 21-30.

Table 11: Sales probability by difference of expected condition score (CS), weeks 21-30

Tercile of Difference from Expected CS	# of Cars	No posted SCR	Posted SCR	Difference	% Difference	z-statistic	p-value
Worse than expected	1,802	0.383	0.375	-0.08	-0.2%	-0.36	0.72
Close to expected	1,800	0.429	0.452	0.02	4.6%	0.99	0.32
Better than expected	1,800	0.477	0.483	0.005	1.3%	0.23	0.82

Table 12: Price/NAP by difference of expected condition score (CS), weeks 21-30

Tercile of Difference from Expected CS	# of Cars	No posted SCR	Posted SCR	Difference	% Difference	t-statistic	p-value
Worse than expected	680	0.99	0.98	-0.006	-0.6%	-0.35	0.73
Close to expected	781	1.09	1.08	-0.019	-1.7%	-0.88	0.37
Better than expected	847	1.1	1.1	0.004	0.36%	0.24	0.81

6.3 How Information Disclosure Changes Bidder Behavior

Next, we considered the choice of bidders regarding which scores of vehicles to bid on. Ideally, we would have observed that after information was disclosed, we had less variance in the condition score of vehicles that any given bidder chose to bid on. Unfortunately, we did not know on which cars bidders chose to bid, but only the cars that bidders successfully won. Using a variance test on the vehicles that a bidder wins is not informative for a simple reason: given the endogenous choice of the reserve price, both with and without information disclosure the right "type" of bidder should win most of the time.[30]

Instead, we indirectly tested to see if bidders responded to the disclosed information. The auction registration process assigned vehicles to lanes prior to the SCRs being generated. During weeks 21-30 bidders knew where vehicles were but had less information about them. Hence, the benefit of switching from one lane to another in search of better matched vehicles was limited. After week 30, however, bidders had more information about the vehicles. Using this information, the benefit of switching lanes in pursuit of a better matched vehicle was higher. We expected, therefore, that for any given number of vehicles that a bidder bought, he would have visited more lanes after week 30.

We regress the number of vehicles purchased by each dealer per week on the number of lanes in which the dealer purchased the cars. We allowed this relationship to differ for weeks 21 to 30 and 31 to 39, respectively. To ensure that the estimation was from within-dealer variation in the number of cars purchased over time, we estimated all specifications with buyer fixed effects. The results are in column 1 of Table 13.

As hypothesized, buyers on average used more lanes after week 30: up to week 30, for every additional

[30]The variance in quality of the cars bought by each bidder with and without SCRs was indeed the same.

Table 13: Number of lanes used by dealers per week†

	All Cars	SCR Cars	Non-SCR Cars
Number of cars	.47**	.42**	.49**
	(.05)	(.075)	(.076)
Week 31-39	-.21**	-.31*	-.17+
	(.067)	(.12)	(.1)
Week 31-39 * Number of cars	.17**	.25*	.13
	(.055)	(.098)	(.082)
Buyer Fixed Effects (837)	yes	yes	yes
Constant	.58**	.64**	.55**
	(.062)	(.097)	(.096)
Observations	2690	1401	1289
R-squared	0.779	0.796	0.843

* significant at 5%; ** significant at 1%; + significant at 10% level. Robust SEs in parentheses.

† An observation is a dealer-week conditional on the dealer having made any purchases during a week. If a dealer makes any purchases during a week, on average a dealer purchases 1.47 cars per week.

car purchased, dealers purchased cars on 0.47 additional lanes. Starting in week 31, for every additional car purchased, dealers purchased these on 0.64 (0.47+0.17) additional lanes. Notice, however, that this relationship should have held only for cars with an SCR. This is because even after week 31, dealers had no additional information about cars without an SCR. In columns 2 and 3 of Table 13, we thus split our sample into cars with an SCR and cars without an SCR. As shown, there was only a difference in how many lanes were used before and after weeks 30 for cars with an SCR. The interaction between the dummy for weeks 31-39 and number of cars was only significant for cars with an SCR but not for cars without an SCR.

Hence, in lieu of actually observing which cars bidders chose to bid on, we concluded that they used the disclosed information to switch lanes more often and target vehicles that were a better match for their potential customers.

7 Robustness

This section presents further evidence about the reliability and interpretation of our results. We first analyzed whether the statistical inferences in Tables 2 through 9 are robust to a standard error correction. We also revisited our randomization procedure by checking whether our findings are robust to the inclusion of fixed effects. Second, we use online bidding to test our interpretation of how weeks 21-30 and 31-39 differed during the experiment. Finally, to the extent possible, we ruled out that bidders were responding to something other than the informational content of SCRs.

7.1 Standard Error Correction and Randomization Check

Recall from footnote 11 that 30 percent of cars in our sample were offered for sale more than once during the sample period. In this subsection we account for the potential correlation in the error terms for cars that were offered multiple times. We do so by estimating linear probability models with robust clustered standard errors at the VIN level. This accommodates arbitrary correlation between the errors of observations of the same car.

We also revisited whether our randomization procedure yielded a random assignment to treatment and control groups. We analyzed whether our basic results change after adding a large set of controls, namely seller fixed effects (267), model year fixed effects (13), vehicle segment fixed effects (21), nameplate fixed effects (38), sale week fixed effects (9), condition score tercile (3), and some (non-SCR) measures that represented the condition of the car, namely its mileage and whether it was offered under a green, yellow, or red light, as well as a blue light.[31] If our randomization procedure worked, these controls should not substantially change our estimates, because the randomization should have ensured that cars of the same make, model year, segment, and approximate condition were randomly assigned between treatment and control groups.

First consider the aggregate finding that during weeks 31-39, cars with posted SCRs had a significantly higher probability of sale than cars without a posted SCR (second row of Table 2). Column 1 of Table 21 in the appendix shows that clustered standard errors don't change this inference. Column 2 contains the treatment effect on the probability of sale controlling for the large set of controls we listed above. The point estimate of the treatment effect drops from 6.3 percentage points to 4.6 percentage points. However, we can't reject the hypothesis that the treatment effect is unchanged by the inclusion of the extensive set of fixed effects. Columns 3 and 4 of Table 21 show that our inference about the effect of a posted SCR on prices during weeks 31-39 (second row of Table 4) also remains unchanged. Clustering standard errors and controlling for fixed effects did not alter our conclusion that average prices seem not to have significantly increased due to SCRs.

[31] The seller of every car sold at the auction has to offer their car under some lights. A green light means that the seller declares that the car has no known mechanical problems. A yellow light means that the seller declares that the car has no known mechanical problems other than those that are listed (e.g., "rough engine"). A red light means that the seller sells the car "as is" with no assurance of its mechanical condition. The auction company will arbitrate disputes that may arise for cars that were offered under a green and yellow light if the buyer finds undisclosed mechanical problems. A blue light means that the title of the car is not at the auction site.

We repeated these tests for the results that were decomposed by conditions scores (Tables 5 and 6). Columns 1 and 2 of Table 22 contain the effect of posted SCRs on the probability of sale by condition score. The relevant comparisons to the effects listed in Table 5 under the "Difference column" are the first three coefficients in the table. Clustering standard errors (column 1) does not change our inference. Adding controls changes the estimated coefficients very little. Similarly, columns 3 and 4 of Table 22 don't affect the interpretation of our price results.

Finally, we repeated this analysis for the results that were decomposed by the difference from the expected conditions score (Tables 8 and 9). The results are reported in columns 1-4 of Table 23. Neither the coefficient estimates nor the standard errors are affected significantly by the standard error correction and the addition of controls.

In summary, the conclusions of the key specifications in the paper are unaffected by clustering standard errors and by adding a large set of controls—there is no evidence that our procedure yielded a non-random assignment to treatment and control groups.

7.2 Online Transactions

We argued that SCRs had no effect on auction outcomes during weeks 21-30 because dealers were not aware that they had been posted. We confirmed this by exploring the behavior of dealers who must have been aware that SCRs were posted during weeks 21-30. Showing that these dealers behaved no differently before and after week 31 would present evidence in support of our argument.

To do this we made use of the auction house's online bidding feature. To access online bidding, dealers must use the web portal where SCRs are posted. Moreover, this is the only source of information that puts online dealers on some equal footing with the on-site lane bidders. We considered three measures of online behavior as a function of whether an SCR was posted or not: (1) the percentage of vehicles that received an online bid, (2) the percentage of sold vehicles bought by an online bidder, and (3) the average number of online bidders. The results reported here are based on Tables 17, 18, and 19, all which appear in the Appendix.

Over all weeks (21-39), 3.45 percent of cars with a posted SCR received an online bid, compared to 2.54 percent without a posted SCR. This 36 percent difference is statistically significant (using a test of proportions, p-value 0.02). The key comparison is whether a similar difference already existed in weeks 21-30 or whether it was driven by dealer behavior in weeks 31-39. We found that the effect of an SCR on the percentage of vehicles that received an online bid was statistically no different in weeks 21-30 as it is in weeks 31-39 (p-value 0.75).

Similarly, over all weeks, the winning bids of 4.7 percent of cars with a posted SCR were placed online, compared to 3.07 percent without a posted SCR. This 53 percent difference is statistically significant (p-value 0.01). The effect of an SCR on the percentage of vehicles with an online winning bid was statistically no different in weeks 21-30 from weeks 31-39 (p-value 0.44).

Finally, we found that over all weeks, more online bidders participated in auctions for cars with a posted SCR (4.74 per 100 auctions) than for cars without a posted SCR (3.66 per 100 auctions). Similar to the previous two measures, the SCR effect was statistically no different in weeks 21-30 compared to weeks 31-39 (p-value 0.71).

Given that online dealers knew about SCRs from the beginning of the experiment (week 21), and given that the effect of a posted SCR on their behavior was similar between weeks 21-30 and 31-39, we concluded that the effect of SCRs we observed offline during weeks 31-39 was most likely tied to dealers learning about SCRs.

7.3 Salience and Substitution

It is important to rule out that bidders were responding to something other than the informational content of SCRs. Of particular concern is that emails from week 31 onward just focused buyers' attention on cars with posted SCRs (a "salience" effect.) This would increase the number of bidders for cars with posted SCRs and decrease the number of bidders for cars without posted SCRs. As a consequence, reserve prices were more likely to be met for cars with posted SCRs, increasing their probability of sale, and less likely to be met for cars without posted SCRs, decreasing their probability of sale. This substitution would be consistent with the time-pattern of our data in Table 2. The probability of a sale was 43 percent in weeks 21-30 for both conditions. In weeks 31-39 the *average* probability of sale was just below 43 percent but cars without a posted SCR sold 39 percent of the time while cars with a posted SCR sold 45.5 percent of the time.

To address this concern, consider the results in Table 8. If SCRs caused a pure salience effect, bidders should have responded to SCRs *regardless* of their informational content. The table offers strong evidence that bidders responded only to the informational content of the SCRs and not to their mere presence. In particular, for the middle tercile, where SCRs had little informational content, there was no significant effect of SCRs on sale probabilities. In contrast, for the worse-than- and better-than-expected terciles, where SCRs had stronger informational content, there was a positive effect. We conclude that salience was not driving our results.

We are left to explain why the probability of sale for cars without a posted SCR dropped in weeks 31-39. Recall that the experiment was conducted between May and September 2008, a period of significant decline in the stock market and the housing market. Arguably, these events may have affected sales probabilities. We therefore tested whether the decline in the sales probability of cars without a posted SCR from weeks 21-30 to weeks 31-39 reflected a general market trend. To estimate a secular trend we needed data that were not part of the market in which the experiment was conducted, so we chose cars that were offered for sale by fleet-sellers. For these cars there was no change in available information due to the experiment.[32]

The probability of sale for fleet-seller-consigned cars was 67.25 percent in weeks 21-30 (13,491 cars) and 59.83 percent in weeks 31-39 (12,864 cars), a drop of more than 7 percentage points. This suggests that demand for cars at the auction site decreased over the period of the experiment. Adding fleet-seller-consigned cars to our sample allowed us to use a difference-in-differences linear probability regression to estimate the change over time in the probability of sale for cars with and without a posted SCR relative to fleet-seller-consigned cars.[33]

The results are in column 1 of Table 14. The constant is the probability of sale for fleet-seller-consigned cars during weeks 21-30. The coefficient of *Week 31-39* is the change in the probability of sale for fleet-seller-consigned cars relative to weeks 21-30 and measures the secular trend. The variables of interest are the interaction between *Week 31-39* and the two dealer-consigned car conditions. To account for correlation in the errors when a car was offered for auction more than once during the sample period, we cluster the standard errors at the VIN level.

The coefficient on *Week 31-39 * Dealer-consigned car, no posted SCR* is 0.031 (p-value 0.19). We cannot therefore reject the hypothesis that the change between weeks 21-30 and weeks 31-39 in the probability of sale for dealer-consigned cars without a posted SCR was the same as for fleet-seller-consigned cars. In contrast, the coefficient on *Week 31-39 * Dealer-consigned car, posted SCR* is 0.089 and is significantly different from 0 (p-value < 0.01). The interpretation of these results is as follows: under the maintained assumption that the demand conditions of fleet-seller-consigned cars changed similarly to the those for dealer-consigned cars, we found no evidence that the emails sent out starting in week 31 led dealers to substitute cars without posted SCRs with cars with posted SCRs. Instead, it seems that the probability of sale for cars without posted SCRs was unchanged (relative to fleet-seller-consigned cars), while the probability of sale for cars with posted SCRs increased.

A concern may be that the type of cars sold by fleet-sellers were not comparable to cars sold by dealers, making fleet-seller-consigned cars unsuitable for estimating the secular trend. We can (partially) address this concern by re-estimating the specification in column 1 of Table 14 with model-year, vehicle segment, nameplate and sale-week fixed effects, and some (non-SCR) measures that represented the

[32] More than 98 percent of fleet-seller-consigned cars receive some type of inspection by the auction house. The inspection is generally not as thorough as the inspection that underlies the SCR in our experiment. The exact nature of fleet-seller-inspection depends on the requirements of the fleet-seller and thus varies by fleet-seller.

[33] The maintained assumption in using this difference-in-differences approach is that fleet-seller-consigned cars and dealer consigned cars are subject to the same secular trend. While we cannot test whether this was the case during the treatment period, we can test for equality of pre-treatment trends between fleet-seller- and dealer-consigned cars. Using data from the beginning of the year to one week before the experiment started (nineteen weeks), we used a linear probability model that estimated a linear time trend in the probability of sale for cars, separately for fleet-seller- and dealer-consigned cars. The results are in Table 20. We cannot reject the hypothesis that the secular trend in probability of sale was the same for fleet-seller- and dealer-consigned cars.

Table 14: Linear probability model: diff-in-diff specification†

Dependent Variable: Sold	(1)	(2)
Dealer-consigned car, no posted SCR	-.24**	-.27**
	(.012)	(.015)
Dealer-consigned car, posted SCR	-.23**	-.27**
	(.012)	(.015)
Week 31-39	-.07**	
	(.0066)	
Week 31-39 * Dealer-consigned car, no posted SCR	.031	.029
	(.019)	(.02)
Week 31-39 * Dealer-consigned car, posted SCR	.089**	.087**
	(.02)	(.019)
Mileage on Car		1.6e-07
		(1.0e-07)
Green light		.14**
		(.0081)
Yellow light		-.011
		(.01)
Blue light		-.11**
		(.0096)
Sale Week Fixed Effects	no	yes
Model Year Fixed Effects	no	yes
Vehicle Segment Fixed Effects	no	yes
Nameplate Fixed Effects	no	yes
Constant	.67**	.66**
	(.0049)	(.2)
Observations	35287	35287
R-squared	0.034	0.119

* significant at 5%; ** significant at 1%; + significant at 10% level. SEs (robust and clustered at the VIN) in parentheses.

† Notice that our specification does not distinguish between fleet-seller consigned cars with and without inspections. This is because the inspections are not comparable to the inspections that yield SCRs in our experiment. In addition, more than 98% of fleet-seller consigned cars have some form of inspection.

car's condition, namely, mileage and whether it was offered under a green, yellow, or red light and a blue light. This identifies the secular trend and the result of inspections within cars of the same make, model-year, segment, and approximate condition. As can be seen in column 2 of Table 14, there was very little change in the estimates.

A remaining concern may be that there was substitution between fleet-seller-consigned cars and dealer-consigned cars with a posted SCR. If so, controlling for the secular trend by using the change in probability of sale of fleet-seller-consigned cars would no longer be valid. To address this concern we constructed a sample of buyers who only purchased fleet-seller-consigned cars during weeks 21-30. This category comprised 616 dealers, a large fraction of the 1,670 dealers who purchased at least one car (fleet-seller- or dealer-consigned) during our experimental period. If there was substitution between fleet-seller- and dealer-consigned cars with a posted SCR, we should find that these 616 dealers—if they purchased *any* dealer consigned cars during weeks 31-39—were more likely to buy cars with a posted SCR than without a posted SCR. We found no evidence of such behavior: dealers who purchased only fleet-seller-consigned cars during weeks 21-30 purchased forty-eight dealer-consigned cars with a posted SCR and fifty-three dealer-consigned cars without a posted cars during weeks 31-39. We concluded that substitution is unlikely to explain why SCRs increased expected auction revenues.

8 Concluding Remarks

It is well established that information disclosure can help market participants better evaluate the value of goods and services they are interested in, often resulting in more efficient outcomes and less distortionary information rents. For example, Lewis (2010) showed that by voluntarily disclosing private information on eBay Motors, sellers could effectively offer protection to buyers from adverse selection. This revealing insight helped explain the prevalence of many online transactions that otherwise may seem puzzling due to concerns over potential "lemons."

We have demonstrated that in addition to these effects, information disclosure can play an important role in providing information that helps buyers choose *which market* to participate in. This simple, yet novel insight has broader applications beyond our market for used automobiles. If heterogeneous participants can sort into markets for heterogeneous goods, then better ex ante information will help them sort into markets for which they have the most value, and in turn, *effective competition* will intensify in all markets. Turning back to eBay's huge marketplace, sellers who reveal more information give buyers the chance to self-select into those auctions that they are most interested in.

Our stylized model is tailored to the environment we analyzed, and was successfully used both to rationalize our preliminary results and to generate hypotheses that could be tested in the particular auctions that we studied. Having exclusive, simultaneous auctions for similar yet differentiated goods is close in spirit to online auction sites such as eBay, but other markets will have different institutional details. Developing a general model of information disclosure in markets is beyond the scope of this paper, yet the intuitive driving forces behind our results seem both fundamental and more general. For example, a firm looking to hire people for similar, yet distinct positions may gain from providing more information on its positions, even if the information for some positions may make them seem unattractive relative to others. If a firm posts job vacancies for two positions that share some similarities, each position will receive a more refined and better-matched pool of applicants if more information that distinguishes the two positions in terms of requirements, skills, and job descriptions is disclosed.

The implications of information disclosure as a matching mechanism may also apply to government procurement. Typically, governments engage in both parallel and sequential procurement of many similar yet distinct projects. For example, several construction, road, or defense acquisition projects may be let out to bid simultaneously, and yet many more are anticipated to materialize within weeks or months. Though these are often thought of as sequential and not exclusive simultaneous auctions, bidders (contractors) with capacity constraints may not be able to bid on later auctions if they win earlier ones. If the procurement authority releases information not only on current but also on future tenders, heterogeneous contractors may be able to better select which of the coming auctions to participate in.[34] This in turn can increase effective competition both within and between projects, which would be a consequence of information disclosure as a matching mechanism.

[34] Previous theoretical studies have shown that providing information for sequential auctions can increase expected revenues by having bidders use early auctions to learn about their competitors' willingness to pay, and then decide whether and how to bid in future auctions. See Budish (2008) and Zeithammer (2006).

References

Board, Simon. (2009) "Revealing Information in Auctions: The Allocation Effect," *Economic Theory*, 38(1):125-135.

Budish, Eric (2010) "Sequencing and Information Revelation in Auctions for Imperfect Substitutes: Understanding eBay's Market Design," mimeo.

Cho, Sung-Jin, Harry J. Paarsch and John Rust (2010) "Is the 'Linkage Principle' Valid?: Evidence from the Field," mimeo.

Cutler, David M., Robert S. Huckman and Mary Beth Landrum. (2004) "The Role Of Information In Medical Markets: An Analysis Of Publicly Reported Outcomes In Cardiac Surgery," *American Economic Review Papers and Proceedings*, **94(2)**:342-346.

De Silva, Dakshina G., Timothy Dunne, Anuruddha Kankanamge and Georgia Kosmopoulou (2008) "The impact of public information on bidding in highway procurement auctions," *European Economic Review* **52(1)**:150–181.

Genesove, David (1993) "Adverse selection in the wholesale used car market," *Journal of Political Economy* **101(4)**:644-665.

Goeree, Jacob. K., Offerman, Theo. (2002) "Efficiency in Auctions with Private and Common Values: An Experimental Study," *American Economic Review* **92(3)**:625-643.

Hishleifer, Jack (1971) "The Private and Social Value of Information and the Reward to Inventive Activity." *American Economic Review*, **61(1)**:561–574.

Jin, Ginger Z. and Phillip Leslie (2003) "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics*, **118(2)**:409-51.

Kagel John H. and Dan Levin (1986) "The Winner's Curse and Public Information in Common-Value Auctions," *American Economic Review*, **76(5)**:894-920.

Kagel, John H., Ron M. Harstad and Dan Levin (1987) "Information impact and allocation rules in auctions with affiliated private values: A laboratory study." *Econometrica* **55(6)**:1275–1304.

Levin, Dan and Janus L. Smith, (1994) "Equilibrium in auctions with entry," *American Economic Review* **84(3)**:585-99.

Levin Dan, John H. Kagel and Jean-Francois Richard (1996) "Revenue Effects and Information Processing in English Common Value Auctions," *American Economic Review*, **86(3)**:442-460.

Lewis, Gregory (2010) "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors," *American Economic Review*, forthcoming.

R. Preston McAfee and John McMillan (1987) "Auctions with a stochastic number of bidders," *Journal of Economic Theory* 43(1):1-19.

Milgrom, Paul R. (1981) "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12(2):380-391.

Milgrom, Paul R. and Robert J. Weber (1982) "A theory of auctions and competitive bidding," *Econometrica* **50 (5)**:1089–1122.

Ottaviani, Marco, and Andrea Prat (2001) "The Value of Public Information in Monopoly," *Econometrica* **69(6)**:1673–83.

Overby, Eric and Sandy Jap (2009) "Electronic and Physical Market Channels: A Multi-Year Investigation in a Market for Products of Uncertain Quality," *Management Science*, **55(6)**:940-957.

Perry, Motty and Phillip .J. Reny (1999) "On the failure of the Linkage Principle in multi-unit auctions," *Econometrica* **67(4)**:895–900

Porter, Robert H. (1995) "The role of information in US offshore oil and gas lease auction," *Econometrica* **63(1)**:1–27.

Roberts, James W. and Andrew Sweeting (2011) "Competition versus Auction Design," mimeo.

Zeithammer, Robert (2006) "Forward-Looking Bidding in Online Auctions," *Journal of Marketing Research,* **43(3):**462-476.

Appendix

Claim 1: With no disclosure there are two equilibrium outcomes: a unique symmetric *random equilibrium* where each bidder chooses each lane with equal probability and a unique asymmetric *coordinated equilibrium* where exactly one bidder of each type is in each lane.

Proof: Consider pure strategies. First, it is easy to see that if one lane has no bidders, then any bidder in the other lane would prefer to switch lanes and win at a price of zero. Second, imagine that only one bidder is choosing lane 1. Each of the two identical bidders in lane 2 would have a strict preference to switch lanes *ex ante,* because in lane 1 each can win half the time and obtain some rents, while staying in lane 2 they will either lose to the third bidder, or they will win but compete away most of the rents (with the exception of the difference in their ε_i's). This last argument also rules out an equilibrium where each lane has two bidders of the same type, because every bidder will have an incentive to switch lanes. The only other pure-strategy configuration involves one L-type and one H-type in each lane, each winning half the time. In this configuration, when winning a vehicle with quality q, the winner obtains expected rents equal to $\max_\theta\{v_\theta^q\} + \varepsilon_i - \min_\theta\{v_\theta^q\}$. Any bidder who switches from this configuration will compete with his own type, implying that he would obtain expected rents strictly less than $\overline{\varepsilon} < \max_\theta\{v_\theta^q\} + \varepsilon_i - \min_\theta\{v_\theta^q\}$. Hence, this is the unique pure-strategy equilibrium. Consider mixed strategies. It is easy to see that randomly choosing a lane with equal probability is an equilibrium, because a bidder who believes that the other bidders are using this strategy is indifferent between the two lanes. No other mixed-strategy profile can be an equilibrium, because if some bidder of type θ chooses a lane with probability greater than $\frac{1}{2}$ then the best reply of the other bidder of the same type would be to choose the other lane with probability 1 to increase the probability of winning with positive rents. ■

Claim 2: Given two vehicles with qualities $q < 0.5$ and $q' > 0.5$ auctioned in two lanes with full disclosure, the unique equilibrium has perfect sorting: both L types choose the q-lane and both H types choose the q'-lane.

Proof: We show that in any other configuration, at least one bidder has an incentive to switch lanes. First, it is easy to see that random assignment is not an equilibrium. If all the other bidders choose lanes randomly, an H type bidder has a strict incentive to choose the q' lane, because his probability of winning that vehicle is higher, and conditional on winning, he is left with higher rents. (A symmetric argument applies to a L type choosing the q lane.) Second, with pure strategies it is easy to see that if one lane has no bidders, anyone from the other lane would have preferred to switch lanes. Third, imagine that there is only one bidder in lane q and three in lane q'. If the sole bidder in lane q is an H-type, then each of the L-types in lane q' has an incentive to switch lanes, because they lose in the q' lane and they would win in the q lane and obtain rents. (A symmetric argument holds for a sole L-type in the q' lane.) If the sole bidder in lane q is an L-type, then the L-type in lane q' has an incentives to switch lanes. If he stays in lane q' then he will definitely lose against the two H-types. If he switches, there is a positive probability that his idiosyncratic noise ε_i is greater than that of the other L-type in lane q, in which case the switching bidder would win and obtain a small rent.[35] (A symmetric argument holds for a sole H-type in the q' lane.) Finally, assume that each lane has two bidders, one of each type. In

[35]If there is no idiosyncratic noise, this bidder would be indifferent between losing to the H-types and switching lanes only to see the price of the g vehicle rise to v_L^q, leaving him with no rents.

this case the H-type in lane q and the L-type in lane q' both lose, while if they switched there would be a positive probability that each of their idiosyncratic noise ε_i is enough to make them win and obtain a small rent. To complete the analysis, observe that if the two types perfectly sort as stated in Claim 2 above, no one has an incentive to switch. Each has a positive probability of winning and obtaining a small rent, while by switching each is guaranteed to lose. ■

Table 15: Dealer-consigned and inspected cars by week†

Sale Week	Dealer-Consigned Total	With SCR Not reported	With SCR Reported
21	1,442	237	223
22	1,709	195	186
23	1,438	324	330
24	1,606	281	365
25	1,249	303	344
26	1,408	229	250
27	1,170	290	305
28	1,462	245	245
29	1,440	267	281
30	1,621	231	269
31	1,533	233	247
32	1,590	214	215
33	1,329	237	154
34	1,555	225	185
35	1,526	150	140
36	1,474	73	85
37	1,418	90	107
38	1,554	71	84
39	1,639	82	104
Total	28,163	3,977	4,119

Weeks are of 2008.

Table 16: Summary Statistics

Variable	N	mean	p50	sd	min	max
Model Year	8098	2003.5	2004	2.7	1997	2009
Mileage	8098	75958.6	71315.5	44359.1	0	508112
condition score	8098	2.42	2	1.31	1	5
Repair Costs	8098	1347.9	1024	1236.7	0	16110.8
Sold	8098	0.43	0	0.50	0	1
Sales Price	3481	8660.8	7300	5929.9	500	59000
National Auction Price	3429	8397.2	6975	5810.8	200	62000
Sales Price/National Auction Price	3429	1.06	1.03	0.24	0.24	5.6

* The number of observations for the "National Auction Price" and "Sales Price/National Auction Price" is lower than for "Sales Price" because the "National Auction Price" is missing for a few cars in our data.

Table 17: Percentage of dealer-consigned cars which received an online bid

	No posted SCR	Posted SCR	Difference	% Difference	z-statistic	p-value
All weeks	2.54% 3,980 cars	3.45 % 4,118 cars	0.91%	35.8%	2.40	0.016
Weeks 21-30	2.69% 2,605 cars	3.50% 2,797 cars	0.81%	30.2%	1.73	0.084
Weeks 31-39	2.25% 1,375 cars	3.33% 1,321 cars	1.08%	47.7%	1.70	0.089

Table 18: Percentage of sold dealer-consigned car where winning bid was placed online

	No posted SCR	Posted SCR	Difference	% Difference	z-statistic	p-value
All weeks	3.07% 1,660 cars	4.72 % 1,821 cars	1.65%	53.6%	2.50	0.01
Weeks 21-30	3.21% 1,121 cars	4.51% 1,220 cars	1.29%	40.3%	1.62	0.10
Weeks 31-39	2.78% 539 cars	5.15% 601 cars	2.37%	85.3%	2.03	0.04

Table 19: Expected number of online bidders per 100 auctions

	No posted SCR	Posted SCR	Difference	% Difference	t-statistic	p-value
All weeks	3.66 3,980 cars	4.74 4,118 cars	1.08	29.8%	2.22	0.026
Weeks 21-30	3.77 2,602 cars	4.72 2,798 cars	0.95	25.3%	1.58	0.11
Weeks 31-39	3.42 1,375 cars	4.77 1,321 cars	1.35	39.5%	1.60	0.11

Table 20: Pre-promotion trends: Sales probability during weeks 1-19

	Sold
Time Trend	-.0045** (.0005)
Fleet-Seller	.33** (.0084)
Fleet-Seller*Time Trend	-.00096 (.00073)
Constant	.48** (.0057)
Observations	57513
R-squared	0.105

* significant at 5%; ** significant at 1%; + significant at 10% level. Robust SEs in parentheses.

Table 21: Randomization check on aggregate results: Sales probability and Transaction Prices for weeks 31-39

	Sales Probability		**Transaction Prices**	
	Base Result	Fixed Effects	Base Result	Fixed Effects
Posted SCR	.063** (.019)	.046* (.021)	.02 (.012)	.0097 (.013)
CS close to expected		.035 (.028)		.08** (.018)
CS better than expected		.11** (.033)		.085** (.021)
Mileage on Car		6.8e-07 (4.4e-07)		3.1e-07 (3.3e-07)
Green light		.087+ (.047)		.17** (.046)
Yellow light		-.039 (.033)		-.033 (.027)
Blue light		-.12+ (.069)		-.0093 (.037)
Seller Fixed Effects	no	yes	no	yes
Model Year Fixed Effects	no	yes	no	yes
Vehicle Segment Fixed Effects	no	yes	no	yes
Nameplate Fixed Effects	no	yes	no	yes
Sale Week Fixed Effects	no	yes	no	yes
Observations	2696	2696	1121	1121
R-squared	0.004	0.273	0.002	0.433

* significant at 5%; ** significant at 1%; + significant at 10% level. Robust and clustered (by VIN) SEs in parentheses.

Table 22: Randomization check on results by CS: Sales probability and Transaction Prices for weeks 31-39

	Sales Probability		**Transaction Prices**	
	Base Result	Fixed Effects	Base Result	Fixed Effects
Posted SCR * CS Worse than average	.077* (.037)	.09* (.037)	.049* (.025)	.054* (.024)
Posted SCR * CS Close to average	.015 (.036)	.026 (.036)	.016 (.021)	.0034 (.019)
Posted SCR * CS Better than average	.094* (.036)	.083* (.036)	.0018 (.018)	-.00052 (.016)
CS Close to average	.077* (.037)	.1** (.038)	.086** (.023)	.13** (.022)
CS Better than average	.052 (.036)	.14** (.04)	.059** (.021)	.14** (.022)
Mileage on Car		-2.2e-07 (4.2e-07)		4.9e-07 (3.0e-07)
Green light		.11* (.042)		.19** (.036)
Yellow light		-.029 (.031)		-.041+ (.022)
Blue light		-.11 (.066)		.015 (.037)
Model Year Fixed Effects	no	yes	no	yes
Vehicle Segment Fixed Effects	no	yes	no	yes
Nameplate Fixed Effects	no	yes	no	yes
Sale Week Fixed Effects	no	yes	no	yes
Constant	.35** (.026)	.27 (.2)	.98** (.018)	.84** (.13)
Observations	2696	2696	1121	1121
R-squared	0.008	0.074	0.022	0.236

* significant at 5%; ** significant at 1%; + significant at 10% level. Robust and clustered (by VIN) SEs in parentheses.

Table 23: Randomization check on results by expected CS: Sales probability and Transaction Prices for weeks 31-39

	Sales Probability		**Transaction Prices**	
	Base Result	Fixed Effects	Base Result	Fixed Effects
Posted SCR * CS Worse than expected	.084* (.036)	.084* (.036)	.022 (.021)	.029 (.02)
Posted SCR * CS Close to expected	-.011 (.036)	.00045 (.036)	.035 (.022)	.019 (.022)
Posted SCR * CS Better than expected	.11** (.038)	.11** (.037)	.0065 (.02)	.0022 (.018)
CS Close to expected	.1** (.036)	.1** (.036)	.067** (.021)	.061** (.021)
CS Better than expected	.092* (.036)	.11** (.037)	.094** (.02)	.1** (.02)
Mileage on Car		-5.3e-07 (4.2e-07)		1.9e-07 (2.9e-07)
Green light		.11* (.042)		.19** (.036)
Yellow light		-.033 (.031)		-.042+ (.022)
Blue light		-.11 (.067)		.012 (.037)
Model Year Fixed Effects	no	yes	no	yes
Vehicle Segment Fixed Effects	no	yes	no	yes
Nameplate Fixed Effects	no	yes	no	yes
Sale Week Fixed Effects	no	yes	no	yes
Constant	.33** (.025)	.25 (.21)	.98** (.015)	.87** (.13)
Observations	2696	2696	1121	1121
R-squared	0.015	0.076	0.034	0.218

* significant at 5%; ** significant at 1%; + significant at 10% level. Robust and clustered (by VIN) SEs in parentheses .

Determining Consumers' Discount Rates with Field Studies

Song Yao Carl F. Mela Jeongwen Chiang Yuxin Chen[1]

June 15, 2011

[1]Song Yao is an Assistant Professor of Marketing at the Kellogg School of Management, Northwestern University (email: s-yao@kellogg.northwestern.edu). Carl F. Mela is the T. Austin Finch Foundation Professor of Business Administration at the Fuqua School of Business, Duke University (email: mela@duke.edu). Jeongwen Chiang is a Professor of Marketing at China Europe International Business School (email: jwchiang@ceibs.edu). Yuxin Chen is the Polk Brothers Professor in Retailing and Professor of Marketing at the Kellogg School of Management, Northwestern University (email: yuxin-chen@kellogg.northwestern.edu). The authors would like to thank seminar participants at Northwestern University, 2010 China India Consumers Insights Conference, 2011 UTD Forms Marketing Conference, 2011 Marketing Science Conference, as well as Eric Anderson, Praveen Kopalle, Harikesh Nair, Aviv Nevo and Florian Zettelmeyer for their feedback and Hong Ke and Jason Roos for their invaluable research assistance. The authors are listed in reverse alphabetical order.

Abstract: Determining Consumers' Discount Rates With Field Studies

Because utility/profits, state transitions and discount rates are confounded in dynamic models, discount rates are typically fixed to estimate the other two factors. Yet these rate choices, if mis-specified, generate poor forecasts and policy prescriptions.

Using a field study wherein cellphone users transitioned from a linear to three-part-tariff pricing plan, we estimate a dynamic structural model of minute usage and obtain discount factors that would normally be unidentifiable. The identification rests upon imputing the utility under linear pricing plans without dynamic structure; then using these utilities to identify discount rates when consumers were switched to a three-part tariff where dynamics became material.

We find that the estimated segment-level weekly discount factors (0.86 and 0.91) are much lower than the value typically assumed in empirical research (0.995). When using a standard 0.995 discount factor, we find the price coefficients are underestimated by 23%. Moreover, the predicted intertemporal substitution pattern and demand elasticities are biased, leading to a 27% deterioration in model fit; and suboptimal pricing recommendations that would lower potential revenue gains by 74-88%.

Keywords: dynamic structural model, identification, forward-looking consumers, heterogeneous discount rates, nonlinear pricing.

1 Introduction

Individuals often face situations where they must choose between engaging in consumption in the present or waiting to consume at a future time. A rich stream of recent literature has adopted dynamic structural models to study intertemporal consumption, yielding deep insights into consumer behavior, such as Rust (1987), Erdem and Keane (1996), Hendel and Nevo (2006), and Sun (2005).

Albeit the increasingly ubiquitous use of structural models to study dynamic consumption behavior due to their appealing theoretical foundation, the identification of these models are problematic (Rust, 1994). In particular, to identify consumer utility functions, it is often necessary to i) assume or fix the discount factor at a given value (normally between 0.90 and 0.9999), and ii) assume that this rate is common across individuals. For example, an annual discount factor of 0.95 is often justified via the argument that this value is consistent with an annual interest rate of about 5%. While this rate might be suitable for analyzing the decisions of firms who are subject to capital constraints in the financial market, it is not clear whether this interest rate applies across consumption contexts where individuals have different degrees of access to capital and credit rates. Moreover, it would be desirable to relax the assumption of homogeneous discount rates. Those with low discount rates are more likely to defer consumption. As a result, targeting policies predicated on low discount rates, such as low introductory rates or trial promotions, might be misplaced if the future is not so material to less forward looking consumers.

Because the use of dynamic structural models in marketing is becoming more common and because the assumed discount rate affects inferences regarding agent behavior, optimal policy and forecast outcomes, we develop a dynamic structural model in order to identify and measure heterogeneous consumer discount rates using field data. More specifically, we estimate a dynamic structural model using customer cellphone minute usage data during a field study that involves switching pricing plans to consumers. In our data, customers were initially under a linear "pay-per-minute" plan. Later the cellphone service provider implemented a field experiment and these customers were switched to a nonlinear three-part tariff plan.[1] The switch induced intertemporal substitution trade-offs in the later period of the data as a consumer's decision of minute usage early in the month had consequences for the rates she faced later in the month. In comparison, under the preceding linear pricing schedules, the consumer paid a constant marginal price for usage. Consequently, there was no intertemporal substitution of minute usage and customers made consumption decision statically. Relying on this field experiment, we first obtain customers' utilities of phone usage from the data of the linear plan. Conditional on these identified utilities, we then estimate heterogeneous discount factors of customers with the three-part tariff data.

As an ancillary benefit of the field study, we are able to explore the potential for hyperbolic discounting. Although there are studies show the existence of hyperbolic discounting (see the survey by Angeletos et al. (2001)), we find that no strong evidence of hyperbolic discounting in the focal context of monthly cellphone usage. Our result is consistent with the studies by Chevalier and Goolsbee (2005) and Dubé et al. (2010b), which also do not find strong support for hyperbolic discounting.

Our results indicate that customers demonstrate considerable heterogeneity in their preferences and discounting patterns. In particular, the customers have very high discount rates. The customers only has a weekly discount factor of around 0.86 to 0.91, much lower than the 0.995 weekly discount factor commonly assumed in the empirical dynamic model literature (e.g., Erdem and Keane (1996) and Sun (2005)). This traditional discount rate implies an equally priced minute at the beginning of the month to be worth the same as 1.02 minutes at the end of the month; instead we compute that customers value a minute now more closely to 1.5 to 1.9 minutes at the end of the month (depending on the consumer segment). Furthermore, setting the discount factor to 0.995 leads to an underestimation of price coefficients

[1] A three-part tariff plan contains three components: an access fee, a certain amount of allowance minutes, and a marginal price if the customer's usage exceeds the allowance within a billing cycle. As a result, the customer may be subject to extra fees when the consumption exceed the allowance (overage) and overpays if the consumption falls below the allowance (underage). Due to the existence of the allowance and the high marginal price, a customer needs to decide how to allocate her consumption across time within a billing cycle, intending to avoid overage and underage so as to maximize her total utility.

of up to 23%. The intuition behind this result is that setting the discount rate too high implies customers would excessively tradeoff current within-allowance minute consumption with their future consumption, where they may have to pay the marginal price. Given that they do not actually do this in the data, it results in the model making an effort to lower cross-period substitution in demand; this is achieved by estimating a lower price sensitivity. Correspondingly, the estimated demand elasticities and predicted intertemporal usage patterns under the 0.995 discount factor are also biased, leading to worse fit. In the case of usage patterns, the traditional discount rate assumes individuals are more patient than they really are, understating the tendency to consume minutes early.

The model also enables us to investigate the impacts of the firm's pricing strategy on its revenue and the attendant implications of discount rates for pricing strategy. We conclude that roughly 10% of the customers in our context have a greater baseline consumption rate and lower price sensitivity than the remaining customers, making them good candidates for customized pricing. Accordingly, we calculate the firm's revenue and customers welfare under alternative pricing strategies. We find that the firm's revenue under the alternative pricing plan may potentially increase by nearly 2% on those targeted customers, with little impact on their welfare as measured by the utilities of phone usage. In contrast, alternative strategies developed from a model with commonly used discount rates lead to sub-optimal outcomes that lower potential revenue gains by 36-78%, depending on the market segment considered.

The remainder of the paper is organized as follows. In section 2, we overview the relevant literature to differentiate our paper from past research. Next, to better illustrate our model and identification strategy, we introduce the unique aspects of our data. We then detail the model and estimation. Subsequently, we present and discuss the results and corresponding managerial implications. We conclude with some future research directions.

2 Past Literature

Given that discount rates are not typically identified, several approaches have emerged to contend with the problem, including i) assuming a fixed value for the discount rate, ii) functional identification via structural assumptions and/or estimation via exclusion restrictions, and iii) experimental approaches. Table 1 overviews a sample of these approaches and their resulting discount values converted to their weekly equivalents. Table 1 makes it apparent that discount rates vary considerably across studies. The mean weekly discount factor is 0.979 with a standard deviation of 0.034. The corresponding weekly discount rates average 2.25% with a large standard deviation of 3.76%. In short, there is no clear consensus regarding the value of discount factors, partially due to the fact that discount rates are typically not identified. This large variation is problematic in practice because, as we shall show, the optimal policy for the firm or consumer can vary substantially with the imputed or articulated discount rate.

[Insert Table 1 about here.]

First, most studies assume or fix the discount factors to certain values, typically between 0.995 to 1.0. For the purpose of identification, it is also a common practice to assume the discount factor is the same across individuals which might be, in some instances, a strong assumption (Frederick et al., 2002).

A second approach to the identification of discount rates includes the imposition of structure on the model such as assuming the distribution of the model errors, individuals knowing the state transition probability, and no unobserved heterogeneity (e.g., Hotz and Miller, 1993).[2] However, these structures for the purpose of identification may be difficult to sub-

[2] Besides dynamic structural models, it should be noted that there are studies estimating implied discount factors, using data of consumers trading off immediate return/cost and future flow of return/cost, e.g., current capital cost of a more energy efficient durable product and future operating cost of that product. The literature include Hausman (1979), Dubin and McFadden (1984), Warner and Pleeter (2001), Harrison et al. (2002), Allcott and Wozny (2011) and others. For a more thorough discussion, see Frederick et al. (2002). However, to identify the discount factor, it is still necessary to impose some structures, which are sometimes difficult to verify. For example, to quantify the future operating cost of a durable appliance, it is

stantiate in some contexts (e.g., homogeneous consumers, full information of state transition probability for new products or markets, etc.).

A third identification strategy is to rely on the exclusion restriction condition (Magnac and Thesmar, 2002). Exclusion restrictions involve specifying a set of exogenous variables that do not affect current utility but do affect state transitions. Accordingly, variation in these exogeneous variables affects future utilities through their impact on the state transition but do not have an effect on current utility. By exploring how choices are made in light of changes in future utility when current utility remains fixed, the utility and the discount factor can be simultaneously identified. Some studies implement exclusion restriction to identify not only the exponential discount factor but also the hyperbolic discount factor (e.g., Chung et al. 2010, Fang and Wang 2010). However, such exclusion restrictions are often unavailable in field data or are difficult to validate. Furthermore, though discount factors may vary across individuals (Frederick et al., 2002), it is not clear whether the identification of heterogeneous discount factors under the exclusion restriction is feasible.

To alleviate these concerns, recent work by Dubé et al. (2010b) use experimental conjoint analysis data to identify dynamic model in the context of durable goods adoption. In particular, Dubé et al. (2010b) manipulate consumers' beliefs about state transitions by informing them alternative future market situations in the experiments. As a result, they are able to identify utility and discount factors. This approach is most similar to ours in that it uses data rather than invoking assumptions to infer discount rates. Though an important step forward, it is often difficult to replicate dynamic choices in lab settings owing to demand artifacts and contracted durations; it would therefore be desirable to supplement this research using a field context with choices made in practice and over extended periods.[3] Moreover, field data enables one to explore the potential revenue and utility consequences

necessary to assume the consumer has perfect foresight about future prices of electricity. As a result, there is no uncertainty about the future state transtion.

[3]For example, Dubé et al. (2010b) consider annual budget tradeoffs in a lab experiment that lasts one session.

of mis-specifying discount rates and also conduct appropriate policy simulations involving changes in marketing strategy in the context of dynamic choice.

Therefore we advance the research in dynamic structural models by identifying heterogeneous discount factors using field experiment data. Our identification strategy is to first identify consumers' heterogeneous utilities and the distribution of random consumption shocks using data that have no dynamics involved. Then we further recover their discount factors when the dynamic structure was exogeneously imposed.

Our contributions are fourfold. First, we identify and measure discount rates using field data. This is useful because the estimates are informative for setting discount rates in cases where exogenous variation in temporal decision making does not exist. Second, we consider the role of heterogeneity and the potential for hyperbolic discounting in the context of a field setting. Third, we explore the potential for biased parameter estimates as a result of mis-specifying discount rates as well as the potentially suboptimal marketing decision making. Fourth, our research also advances the empirical literature on the non-linear pricing of telephony or Internet services (e.g., Narayanan et al. 2007, Lambrecht 2006, Lambrecht et al. 2007, Iyengar et al. 2007). Most previous empirical studies are based on aggregate usage data, limiting their ability to investigate customers' intertemporal substitution in consumptions. Since our data are at the disaggregate level, we are able to evaluate the tradeoff of consumptions across time and the corresponding managerial implications for the firm's pricing strategy.

3 Data

3.1 Consumer Usage Data and Carrier Tariff Structure

The data for our analysis are supplied by a major mobile phone service provider in China, covering the period from September 2004 to January 2005. The data provider accounted for more than 70% of the market share of Chinese mobile phone service market during that time. Initially, this firm used only linear pricing schedules, i.e., customers were billed on a pay-per-minute basis. In November 2004, on an experimental basis, the firm offered three-part tariff plans to a randomly selected set of customers. The firm divided these customers into multiple groups based on their past usage volumes. The firm then offered each group a respective three-part tariff plan. A customer could choose the three-part tariff plan or remain on the original pay-per-minute plan.

3.1.1 Tariff Structure

Table 2 depicts the pricing structure of the most popular three-part tariff plans, covering 90% of the customer base. Customer who enroll in one of the listed plans are allowed a fixed number of free calling minutes in a given calendar month by paying the monthly access fee.

[Insert Table 2 about here.]

When a given customer places or receives a call, the minutes of the phone call are deducted from the allowance and the customer does not need to pay for that usage. However, when the monthly allowance is exhausted, the customer is billed the marginal price for each minute of usage beyond the allowance. There is no "roll-over" for these plans, i.e., unused allowance minutes can not be carried over to next month. At the beginning of next month, the customer's allowance of free minutes is replenished after paying the new month's access fee. The customers have different linear rates before the switch. The mean linear rate is 0.27 with a standard deviation of 0.09.

3.1.2 Usage Data

For the first four months (from September 2004 to December 2004), we observe each customer's aggregate monthly minute usage and expenditures. However, in the last month (January 2005), we observe call level customer records, including the starting time, duration, and expense of each phone call. The data also include some demographic information, including the age, gender, and zip code of each customer.

Table 3 summarizes the customers' average usage levels (normalized by their allowance level) and demographic information. The average usage under both the linear and three-part tariff plans are close. However usage variation is considerable, suggesting heterogeneity in usage behavior is material.

[Insert Table 3 about here.]

3.2 Overage and Underage

Underage occurs when customers do not use all the allowance at the end of a month. In this case, customers are overpaying in the sense that they have been charged for minutes they do not use. In comparison, overage occurs when usage exceeds the allowance. In this case, customers again overpay inasmuch as a plan with more allowance minutes normally has a lower average price per allowance minute (Iyengar et al. (2007)). As a result, customers who strategically manage the minutes should evidence less underage or overage.

In Figure 1, we plot the histogram of the ratios of minutes used to minutes allowed for the last month of data. The average ratio is close to 1 (0.96) under the three-part tariff, suggesting that customers on average tend to avoid overage or underage. Yet this average behavior belies a large standard deviation (0.35). Hence, we next consider whether and how users manage their minutes over the month to comport with the allowance; to the extent this behavior changes as the allowance becomes more salient, evidence is afforded for the strategic use of minutes.

[Insert Figure 1 about here.]

3.3 Strategic Minute Usage within a Month

We consider some model free evidence that customers are strategic in their usage of allowance minutes. This evidence is predicated on the notion that minute consumption changes as the distance between minutes used and the allowance becomes small; in particular, consumers start to conserve minutes as the number of free minutes dwindles and the overage potential increases.

Dividing the last month of the data into five 6-day periods, $t = 1, ..., 5$,[4] we compute the ratio of cumulative minutes used to the allowance for each customer at the end of each six-day period. Figure 2 portrays a scatter plot of this ratio and its lag value for each period t, $t = 2, ..., 5$. The line in this figure depicts a nonparametric function relating the ratio and its lag and the gray band indicates the 95% confidence interval for this function.[5] A key insight from this figure is that this function is concave when the cumulative usage is within quota (the ratio in period $t-1$ is less than 1). In contrast, when usage exceeds quota (the ratio in period $t-1$ is greater than 1), the line is almost linear. The concavity of the line pre-quota suggests that people decelerate usage as they approach the quota, that is, they start to ration their minutes to avoid overage. Moreover, those who are far from the quota appear to accelerate usage to avoid underage. In comparison, customers who have already exceeded the quota do not decelerate (or accelerate) their usage. Instead, they follow some relatively stable usage rates, as might be expected were they to no longer face an intertemporal tradeoff in usage. Misra and Nair (2009) and Chung et al. (2010) use similar methods to investigate the effect of quota on salesperson's allocation of efforts across time. They found analogous patterns of dynamic effort allocation in salesforce due to the existence of quota.

[Insert Figure 2 about here.]

To further elaborate upon these insights arising from Figure 2, we consider how usage acceleration (deceleration) changes as individuals approach their allowance/quota. This acceleration can be summarized by the statistic (Usage during period t)/(Usage during period $t-1$). This ratio is analogous to the slope of the line in Figure 2. When the ratio is one, consumers are neither decelerating or accelerating use. When the ratio is greater than one, usage is accelerating. When the ratio is less than one, usage is decelerating. We compute

[4] We use t to index periods within a month and τ to index months.

[5] We consider both spline and local regression methods. The results are similar. The figure presented shows the results from spline method.

this ratio for each person in each period and then, in Figure 3, present a histogram of this minute acceleration measure across persons and periods conditioned on users distance to quota. For example, the upper leftmost histogram shows the distribution of usage acceleration observations conditioned upon customer usage at time $t-1$ being less than 20% of their allowance. Figure 3 indicates that customers usage decelerates as they approach their allowance. Further, when the customer reaches an overage situation (where there is no longer an intertemporal tradeoff in usage), the slopes average around 1, indicating a stable usage rate. These observations are consistent with Figure 2. Overall, we conclude that there exists some model free evidence of strategic behavior on the part of consumers.

[Insert Figure 3 about here.]

4 Model

4.1 Utility under the Linear Pricing Plan

In this section, we first specify the consumer utility for consumption under a linear pricing plan and derive the optimal level of consumption. We then extend the analysis to the case of the three-part tariff plan.

Similar to Lambrecht et al. (2007), we begin by assuming that customer i, in market segment g, derives utility from phone usages and the consumption of a composite outside product (numeraire). To be specific,

$$\begin{aligned} u_{it}(x_{it}, z_{it}) &= (\frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g}) + z_{it}, && (1) \\ s.t. \quad & z_{it} = y_i - p_{i0}x_{it}, && (2) \\ & d_{it}, b_g > 0 \end{aligned}$$

where $t = 1, 2, ..., T$ are the periods within a month; x_{it} is the minutes of phone usage during period t; p_{i0} is the linear price rate of customer i before switching to the three-part tariff; z_{it} is the consumption of the numeraire; y_i is the income; d_{it}/b_g is the main effect of minute usage; b_g is the price sensitivity; and d_{it} represents the baseline consumption level (cf. equation 5).[6]

Following Narayanan et al. (2007) and Lambrecht et al. (2007), we further allow baseline consumption, d_{it}, to be affected by time-variant consumer characteristics, D_{it}, and a random shock ν_{it},

$$d_{it} = \exp(D_{it}^{'}\alpha_g) + \nu_{it} \tag{3}$$

where α_g is a vector of parameters and $\nu_{it} \sim N(0, \zeta_g^2)$ is exogeneously i.i.d. across customers and periods.[7] Sources of the shock may include (1) technical problems with the customer's phoneset or coverage which limit the phone usage; (2) unexpected events that require extra communications with others, and so on. Though the random shocks are unknown to the researchers, the customer observes the shocks *before* deciding her usage levels accordingly. Given all customers in the dataset are experienced users, we assume the distribution of the shocks is known to the customer.[8]

Substituting the budget constraint into equation 1, the utility function can be rewritten as

$$u_{it}(x_{it}, z(x_{it})) = \frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g} + y_i - p_{i0}x_{it} \tag{4}$$

Customer i then chooses the optimal levels of phone usage x_{it} and numeraire consumption z_{it} so as to maximize her total utility subject to the budget constraint. Solving the maximization

[6]Note that the budget constraint normalizes the benefits of consuming one unit of the numeraire to 1. The purpose of this normalization is twofold. First, it transforms the utility up to a monetary scale, which makes any welfare interpretations more meaningful. Second, since we do not observe customer churning or variations in plan choices in the data, the identification of the benefits of consuming the numeraire is infeasible. Such a normalization treatment for the purpose of identification is similar to Narayanan et al. (2007) and Ascarza et al. (2009).

[7]The exponential function ensures that, on average $d_{it} > 0$. One related concern with the use of a normal distribution assumption for the random shocks is that the baseline demand, d_{it}, may become negative. One approach is to consider a truncated normal distribution. However this imposes a considerable computational challenge. Hence we instead assume that the magnitude of ν_{it} (standard deviation) is small compared to $\exp(D_{it}^{'}\alpha_g)$ so a normal distribution is a good approximation of a truncated normal distribution. This assumption is confirmed to be sensible based on the estimation results.

[8]Though this assumption is not material for the static model because the error is revealed prior to the usage decision, the assumption becomes important under the context of the three-part tariff when future shocks become relevant to current period consumptions.

problem of equation 4 yields the optimal consumption

$$x_{it}^{*} = \begin{cases} 0, & \text{if } d_{it} - b_g p_{i0} < 0 \\ d_{it} - b_g p_{i0}, & \text{if } d_{it} - b_g p_{i0} \geq 0 \end{cases} \tag{5}$$

The foregoing equation clarifies the interpretation of (1) b_g as the price sensitivity and (2) d_{it} as the baseline consumption level under the linear pricing plan as it represents a fixed shift in the demand curve as well as the minute consumption level when $p_{i0} = 0$ (Lambrecht et al., 2007). Summing optimal period consumptions within the same month τ yields the optimal total minutes consumed within a month as

$$q_{i\tau} = \sum_{t'=1}^{T} x_{it'}^{*}$$

Finally, we presume customers are heterogeneous across segments but homogeneous within a segment.[9] Accordingly, the conditional probability of customer i belonging to segment g is

$$f_{ig} = \exp(\lambda_{0g} + D_i'\lambda_g) / \sum_{g'} \exp(\lambda_{0g'} + D_i'\lambda_{g'}) \tag{6}$$

where D_i are time-invariant customer characteristics.

4.2 Utility under the Three-part Tariff Plan

The three-part tariff plan can be described as the triple $\{F, A, p\}$, where F is the fixed access fee, A is the allowance amount, and p is the marginal price after the customer exhausts the allowance.

[9] Although we can estimate a model with heterogeneity at the individual level, the scarcity of observations per customer renders a very noisy identification of the preferences and computational difficulties. Hence we use a latent class structure to capture preference heterogeneity in the spirit of Kamakura and Russell (1989).

4.2.1 Period Utility and Budget Constraint

At period t during a given month, customer i who belongs to segment g has a utility level

$$u_{it}(x_{it}, z(x_{it}); s_{it}, \nu_{it}) = \frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g} + z_{it}(x_{it}), \tag{7}$$

$$s.t.\ z_{it}(x_{it}) = y_i - C(x_{it})$$

$$C(x_{it}) = (\sum_{k=1}^{t-1} x_{ik} + x_{it} - A)pI_{\sum_{k=1}^{t-1} x_{ik}+x_{it}>A}$$

$$d_{it} = \exp(D_{it}^{'}\alpha_g) + \nu_{it}$$

where s_{it} is a vector containing state variables at period t that include (1) cumulative usage up to period $t-1$, $\sum_{k=1}^{t-1} x_{ik}$, and (2) period t (or the distance to the terminal period). Among these state variables, the cumulative usage is endogenous and the period t is exogenous. $I_{\sum_{k=1}^{t-1} x_{ik}+x_{it}>A}$ is an indicator, which takes the value of 1 if $\sum_{k=1}^{t-1} x_{ik} + x_{it} > A$ and 0 otherwise. Note that the fixed access fee F does not enter the period budget constraint since it is essentially a sunk cost. It does not affect the optimal decision at period t as long as the choice is not a corner solution.

Substituting the budget constraint to equation 7, we may rewrite the period utility as

$$u_{it}(x_{ijt}, y_i - C(x_{it}); s_{it}, \nu_{it}) = (\frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g}) + y_i - C(x_{it}) \tag{8}$$

Similarly to the period utility under the linear pricing plan, we assume that d_{it} is affected by the random shock ν_{it} and $\nu_{it} \sim N(0, \zeta_g^2)$. ν_{it} is observed by customer i at the beginning of period t, before making the decision of minute consumption.

4.2.2 Total Discounted Utility

As a customer's current minute consumption may affect her future marginal price, the customer aims to maximize her total discounted utility by optimizing her consumption over time. In particular, the total discounted utility at period $t \leq T-1$ can be presented as

$$U_{it}(u_{it}, u_{i(t+1)}, ..., u_{iT}) \equiv u_{it} + \sum_{k=1}^{T-t} \delta_g^k u_{i(t+k)}$$

where $\delta_g \in [0,1]$, representing the discount factor.

We model the customer's minute usage decision as the dynamic optimization problem of a Markov Decision Process (MDP) such that the strategy of minute usage of period t only depends on the then-current state vector s_{it} (Rust (1994)) and the random shock ν_{it}. To facilitate the exposition, we first define $\sigma_{it} = \sigma_{it}(s_{it}, \nu_{it})$ as the strategy of customer i at period t, depending on the state variables s_{it} and random shock. We also define $\Sigma_{it} \equiv (\sigma_{it}, \sigma_{i(t+1)}, ..., \sigma_{iT})$ as a strategy profile for this MDP from period t onwards; this profile includes a set of decision rules that dictate current and future consumptions. Also denote $V_{it}(s_{it}; \Sigma_{it})$ as the expected continuation utility at period t conditioned on s_{it} and Σ_{it}. Because of the finite horizon of this MDP, V_{it} can be defined recursively as follows:

The expected utility of the terminal period T for a given s_{iT} is

$$V_{iT}(s_{iT}; \Sigma_{iT}) \equiv \mathbf{E}u_{iT}(\sigma_{iT}, y_i - C(\sigma_{iT}); s_{iT}, \nu_{iT}) \tag{9}$$

$$\text{where } C(\sigma_{iT}) = (\sum_{k=1}^{T-1} x_{ik} + \sigma_{iT} - A)pI_{\sum_{k=1}^{T-1} x_{ik} + \sigma_{iT} > A} \tag{10}$$

where the expectation is taken over the random shock ν_{iT}.

Then the continuation utility function V_{it} at period $t < T$ can be written recursively as

$$V_{it}(s_{it}; \Sigma_{it}) = \mathbf{E}u_{it}(\sigma_{it}, y_i - C(\sigma_{it}); s_{it}, \nu_{it}) + \delta_g[V_{i(t+1)}(s_{i(t+1)}; \Sigma_{i(t+1)})|s_{it}, \sigma_{it}] \tag{11}$$

$$\text{where } C(\sigma_{it}) = (\sum_{k=1}^{t-1} x_{ik} + \sigma_{it} - A)pI_{\sum_{k=1}^{t-1} x_{ik} + \sigma_{it} > A} \tag{12}$$

where the expectation is taken over the random shock ν_{it}. Further, given s_{it}, ν_{it} and $\sigma_{it}(s_{it}, \nu_{it})$, the state transition $\pi(s_{i(t+1)}|s_{it}, \sigma_{it})$ is deterministic such that $\sum_{k=1}^{t} x_{ik} = \sum_{k=1}^{t-1} x_{ik} + \sigma_{it}$, and the customer is one period closer to the terminal period T.

We further recursively define the optimal strategy profile $\Sigma_{it}^* \equiv (\sigma_{it}^*, \sigma_{i(t+1)}^*, ..., \sigma_{iT}^*)$, starting with the terminal period:

$$\sigma_{iT}^* = \arg\max_{x_{iT}} u_{iT}(x_{iT}, y_i - C(x_{iT}); s_{iT}, \nu_{iT}) \tag{13}$$

and optimal strategy of period $t < T$ is defined recursively as

$$\sigma_{it}^* = \arg\max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \delta_g[V_{i(t+1)}(s_{i(t+1)}; \Sigma_{i(t+1)}^*)|s_{it}, x_{it}] \tag{14}$$

4.2.3 Hyperbolic Discounting

A complication embedded in the dynamic behavior of minute usage is potential time inconsistency among customers. As shown in the literature (cf. Angeletos et al. (2001)), customers may demonstrate time inconsistency in their inter-temporal preferences such that they have a keener preference for short-term return than long-term return (short-term impatience for receiving the return). Accordingly, we also consider an alternative specification to accommodate hyperbolic discounting, which captures the potential time-inconsistency in preference (Phelps and Pollak (1968); Laibson (1997); O'Donoghue and Rabin (1999)).

To be specific, the preference at period $t < T$ is presented by

$$U_{it}(u_{it}, u_{i(t+1)}, ..., u_{iT}) \equiv u_{it} + \beta_g \sum_{k=1}^{T-t} \delta_g^k u_{i(t+k)}$$

where $\beta_g \in [0,1], \delta_g \in [0,1]$. δ_g is the standard exponential discount factor that captures long-term, time-consistent discounting. β_g is the present-bias factor which represents short-term impatience. The commonly used exponential discounting specification is a special case where $\beta_g = 1$ (O'Donoghue and Rabin (1999)).

In comparison to Equation 14, the optimal strategy of period $t < T$ becomes

$$\sigma_{it}^* = \arg\max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \beta_g \delta_g [V_{i(t+1)}(s_{i(t+1)}; \Sigma_{i(t+1)}^*)|s_{it}, x_{it}] \tag{15}$$

5 Estimation and Identification

5.1 Minutes Usage under the Linear Pricing Plans

For a given month τ under the linear pricing plans, we observe customer i's characteristics D_{it} $(t = 1, 2, ..., T)$. In our specific application, D_{it} is time-variant demographic information, including (1) age and (2) the customer's tenure with the firm.

We also observe individual customer's monthly aggregate usage $q_{i\tau} = \sum_t x_{it}^*$.[10] As shown in Appendix, while there is no closed form for the distribution of $q_{i\tau}$, the distribution can be approximated by a truncated normal distribution. As a result, we can write down the likelihood function of customer i for the minutes usage under linear pricing plans.

$$L_{i \cdot linear}^g = \prod_\tau \widetilde{f}(q_{i\tau}|\Omega_g) \tag{16}$$

where $\widetilde{f}(\cdot)$ is the approximation density function of $q_{i\tau}$ detailed in the Appendix; $\Omega_g \equiv \{\alpha_g, b_g, \zeta_g\}$, i.e., the utility parameters and the distribution of random shocks.

5.2 Minutes Usage in Terminal Period T under the Three-part Tariff

In terminal period T, the consumption becomes a static decision given the allowance will be reset next month. Hence we may solve the optimal minute consumption strategy σ_{iT}^* such

[10]Note that for the linear pricing plans, we only observe $q_{i\tau}$ but not the individual x_{it}^*'s.

that

$$\sigma^*_{iT} = \begin{cases} d_{iT} - b_g p, & \text{if } \sum_{t=1}^{T-1} x_{it} + d_{iT} - b_g p > A \\ d_{iT}, & \text{if } \sum_{t=1}^{T-1} x_{it} + d_{iT} < A \\ A - \sum_{t=1}^{T-1} x_{it}, & \text{if } \sum_{t=1}^{T-1} x_{it} + d_{iT} - b_g p \leq A \leq \sum_{t=1}^{T-1} x_{it} + d_{iT} \end{cases} \tag{17}$$

The first component of equation 17 accounts for the situation under which the customer faces a positive marginal price after her cumulative usage exceeds the allowance. The second component represents the situation when the customer's cumulative usage is less than the allowance and the marginal price is zero. The third component represents the situation when the cumulative usage under the optimal σ^*_{iT} exceeds the allowance at a zero marginal price but falls below the allowance with a positive marginal price. We follow Lambrecht et al. (2007) and set the optimal usage under such a situation at a mass point $\sigma^*_{iT} = A - \sum_{t=1}^{T-1} x_{it}$.

According to equation 17, the density of each observed consumption level x_{iT} conditioned on σ^*_{iT} can be written as

$$f_T(x_{iT}|\sigma^*_{iT}, \Omega_g) = \begin{cases} f(x_{iT} = d_{it} - b_g p) & \text{if } \sum_{t=1}^{T} x_{it} > A \\ f(x_{iT} = d_{iT}) & \text{if } \sum_{t=1}^{T} x_{it} < A \\ \Pr(\sum_{t=1}^{T-1} x_{it} + d_{iT} - b_g p \leq A \leq \sum_{t=1}^{T-1} x_{it} + d_{iT}) & \text{if } \sum_{t=1}^{T} x_{it} = A \end{cases} \tag{18}$$

and the likelihood for the customer i in the terminal period T is

$$L^g_{iT} = f_T(x_{iT}|x^*_{iT}, \Omega_g) \tag{19}$$

5.3 Minutes Usage in Period $t < T$ under the Three-part Tariff

For period $t < T$ of January 2005, we observe each customers period minute consumption x_{it}. However, since there is no closed form solution to the optimal strategy profile Σ^*_{it}, a likelihood function based on observed x_{it} and Σ^*_{it} becomes infeasible. Instead, we implement a numerical approximation method to establish a simulated likelihood function for estimation. This approximation method contains two steps: (1) using Monte Carlo integration to simulate the value function V_{it} at a subset of state points and interpolating V_{it} at the remaining state points using regression; (2) simulating the density for each observed x_{it} using V_{it} from the previous step. We elaborate each step below.

5.3.1 Simulating and Interpolating $V_{it}(s_{it}; \Sigma_{it}^*)$

Using backward recursion and simulation, it is possible to numerically evaluate the value function under the optimal strategy Σ_{it}^* specified in equations 14, 15, and 17. To be specific, starting with the terminal period T

1. Conditioned on Ω_g, make $nr = 100$ draws from the distribution of random shocks ν_{iT}.

2. Make $ns = 250$ draws of state points $\sum_{k=1}^{T-1} x_{ik}$, i.e., the cumulative minute usage at the beginning of period T.

3. At each of the ns state points, calculate nr optimal minute consumption levels $x_{iT}^*(s_{iT}, \nu_{iT})$ using equation 17, one for each random shock draw ν_{iT}.

4. For each state point that we draw, the continuation value function $V_{iT}(s_{iT})$ can be approximated by $\widetilde{V}_{iT}(s_{iT}) = \frac{1}{nr} \sum_{\nu_{iT}} u_{iT}(x_{iT}^*; s_{iT}, \nu_{iT})$.

5. For state points that are not drawn, based on the value functions obtained from step 4, we use a spline interpolation to approximate their values.

Then for period $t < T$, we have the following backward recursion steps:

6. Conditioned on Ω_g, make $nr = 100$ draws from the distribution of random shocks ν_{it}.

7. Make $ns = 250$ draws of state points $\sum_{k=1}^{t-1} x_{ik}$, i.e., the cumulative minute usage at the beginning of period t.

8. At each of the ns state points, conditioned on Ω_g, δ_g, β_g and the $\widetilde{V}_{i(t+1)}$, calculate nr optimal minute consumption levels $x^*_{it}(s_{it}, \nu_{it})$ using the following equations (the first for exponential discounting and the second for hyperbolic discounting), one for each random shock draw ν_{it}.

$$x^*_{it}(s_{it}, \nu_{it}) = \arg\max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \delta_g \widetilde{V}_{i(t+1)}(s_{i(t+1)}|s_{it}, x_{it})$$

or

$$x^*_{it}(s_{it}, \nu_{it}) = \arg\max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \beta_g \delta_g \widetilde{V}_{i(t+1)}(s_{i(t+1)}|s_{it}, x_{it})$$

Note that $s_{i(t+1)}$ is deterministic given s_{it} and x_{it}.

9. For each state point that we draw, the continuation value function $V_{it}(s_{it})$ can be approximated by.

$$\widetilde{V}_{it}(s_{it}) = \frac{1}{nr} \sum_{\nu_{it}} [u_{it}(x^*_{it}; s_{it}, \nu_{it}) + \delta_g \widetilde{V}_{i(t+1)}(s_{i(t+1)}|s_{it}, x^*_{it})]$$

or

$$\widetilde{V}_{it}(s_{it}) = \frac{1}{nr} \sum_{\nu_{it}} [u_{it}(x^*_{it}; s_{it}, \nu_{it}) + \beta_g \delta_g \widetilde{V}_{i(t+1)}(s_{i(t+1)}|s_{it}, x^*_{it})]$$

10. For state points that are not drawn, based on the continuation functions obtained from step 9, we use a spline interpolation to approximate their values.

5.3.2 Simulating the Density of Observed x_{it}, $\widetilde{f}_{it}(x_{it}|s_{it}, \Omega_g, \delta_g)$ or $\widetilde{f}_{it}(x_{it}|s_{it}, \Omega_g, \delta_g, \beta_g)$

For each x_{it} observed in the data and its corresponding state point s_{it}, we use the following steps to simulate its density:

1. First draw $nr_{density} = 100$ random shocks ν_{it};

2. For each random draw of ν_{it} and the observed s_{it}, calculate the optimal minute consumption by solving the following equations (the first for exponential discounting and

the second for hyperbolic discounting).[11]

$$x^*_{it}(s_{it}, \nu_{it}) = \arg\max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \delta_g \widetilde{V}_{i(t+1)}(s_{i(t+1)}|s_{it}, x_{it})$$

or

$$x^*_{it}(s_{it}, \nu_{it}) = \arg\max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \beta_g \delta_g \widetilde{V}_{i(t+1)}(s_{i(t+1)}|s_{it}, x_{it})$$

3. Using the calculated $nr_{density} = 100$ optimal $x^*_{it}(s_{it}, \nu_{it})$'s, simulate $\widetilde{f}_{it}(\cdot)$, the density of the observed x_{it}, using Gaussian kernel density estimator.

With the simulated densities for all observed x_{it}, we are able to write a likelihood function for customer i such that

$$L_i^g = \prod_t \widetilde{f}_{it}(x_{it}|s_{it}, \Omega_g, \delta_g)$$

or

$$L_i^g = \prod_t \widetilde{f}_{it}(x_{it}|s_{it}, \Omega_g, \delta_g, \beta_g)$$

5.4 Heterogeneity

We use a finite mixture approach to capture heterogeneity because i) this approach invokes minimal structure on the distribution of preferences and ii) the limited number of observations per subject suggests person-specific effects would be weakly identified. The prior probability of customer i belonging to segment g is $f_{ig} = \exp(\lambda_{0g} + D_i'\lambda_g)/\sum_{g'} \exp(\lambda_{0g'} + D_i'\lambda_{g'})$, where D_i is time-invariant demographic information, including gender and rural residency

[11] In step 7 and step 8 of section 5.3.1, one may choose to include the observed states in the set of $ns = 250$ draws of state points. Then there would be no need to recompute the optimal x^*_{it}'s in this current step. However, observed states may be sparse in some areas of the state space. As a result, the interpolated $\widetilde{V}_{i\cdot}$'s may be inaccurate in those areas. So we choose not to use the observed states in section 5.3.1. Instead, we draw all of the 250 state points randomly so as to cover the state space as well as possible and then supplement these draws with the observed states.

status. Consequently, the unconditional likelihood of the whole data is

$$L = \prod_i \sum_g [f_{ig} L^g_{i \cdot linear} L^g_{iT} L^g_i]$$

We use MLE to estimate the parameters.

5.5 Identification

We provide an informal discussion of the identification of parameters. Since the identification of segment parameters λ_g follows classical argument (cf. McHugh, 1956), we will focus on the remaining parameters. Besides λ_g's, the parameters that construct our model can be categorized into two sets. The first set of parameters appear under both the linear pricing plan and the three-part tariff plan, including $\Omega_g \equiv \{\alpha_g, b_g, \zeta_g\}$, i.e., the utility parameters and the distribution of random shocks. The second set of parameters only affect the demand under the three-part tariff plan, including the discount factors δ_g and β_g. In essence, the parameters $\Omega_g \equiv \{\alpha_g, b_g, \zeta_g\}$ are identified from choices under the linear plan and the terminal period of the three-part tariff, where there are no dynamics involved. Conditioned on Ω_g, we then recover the discount factors δ_g and β_g.

5.5.1 The Identification of Ω_g

The consumption decisions under the linear plans and the terminal period of the three-part tariff have no dynamics involved. Besides individual consumption across time, we further observe the following information under the linear plan and the terminal period of the three-part tariff.

- Different linear prices across individual customers.
- Depending on whether a customer has exhausted her allowance at the beginning of the terminal period, there are variations in marginal prices across individuals.

- Variation in demographic characteristics across individuals and time.

The variations in consumption across and within individuals over time, conditioned on the variations in prices and demographics, enable us to identify α_g and b_g. Together, α_g, b_g, prices and demographics determine the mean levels of consumptions of each individual over time. The observed deviations from such mean levels across individuals and time identify ζ_g, the standard deviation of the random shocks ν_{it}'s.

5.5.2 The Identification of Discount Factors

Myopic customers do not tradeoff consumption over time, meaning they are inclined to ignore potential overage charges later in the month and, as a result, consume many minutes earlier in the month. In contrast, fully forward-looking customers (no discounting of future utilities) consider overage and lower early minute usage accordingly. Hence, there is a difference in the distribution of minutes over time between the two types, with forward-looking consumers shifting a greater proportion of consumption to later periods. These difference in consumption patterns over time enable identification of the discount rate, conditioned on knowing the utility of consumption.

The fact that such intertemporal tradeoffs might be inconsistent between contiguous periods and discontiguous periods distinguish hyperbolic discount factor from exponential discount factor.

More formally, conditioned on the already identified Ω_g, the state variables, marginal prices and demographics at each period $t < T$, we can compute the static consumption levels if the customers were myopic (i.e., $\delta_g = 0$). Instead, if the data demonstrate inconsistency from those static consumption levels, the discount factors can be identified.

More important, when there are at least three periods of data within a billing cycle, exponential discount factor and hyperbolic discount factor can be separately identified (Fang and Silverman, 2009). For example, consider three periods of consumption data. For the case wherein there is only *exponential* discounting, during the first period, observed consumptions would be consistent with a pattern of discounting the second period's utility with a factor of δ_g and discounting the third period with a factor of δ_g^2. In comparison, if there is hyperbolic discounting, the observed consumption should be consistent with discounting the second period by a factor of $\beta_g\delta_g$ and discounting the third period by a factor of $\beta_g\delta_g^2$. Unless $\beta_g = 1$, the aforementioned two data generating processes and hence the observed data are distinguished. Consequently, δ_g and β_g are separately identified.

6 Results

To conserve space, we report results for the 83 customers (15% of the observations) who select the three-part tariff plan with a monthly access fee of 168RMB (about $20.30) an allowance of 800 minutes, and a marginal price of 0.4RMB. As a robustness check, extending the analysis to a second group of 284 customers (access fee 98RMB, allowance 450 minutes) indicates the results change little. The key difference stems from an increase in the standard errors on the order of 10%; this increase arises from computational considerations that require a smaller numbers of simulations than the focal group.[12]

6.1 Segmentation

We consider different potential degrees of unobserved heterogeneity based on the number of segments. Table 4 compares the BIC for alternative specifications of segment numbers and indicates that a two-segment model provides the best fit. Accordingly, our subsequent analyses are predicated upon the 2-segment specification.

[Insert Table 4 about here.]

Based on the magnitude of the segment parameters, the implied segment sizes are 88.8% and 11.2%, respectively. We find that urban male customers are more likely to be in segment

[12] As the second group is three times larger, we reduce the number of draws; instead of $nr = nr_{density} = 100,\ ns = 250$, we use $nr = nr_{density} = 50,\ ns = 100$.

2. Given the nature of the Chinese economy and society, such a cohort of customers are more likely to be white-collar or business people with a relatively higher income level.

6.2 Parameter Estimates

Table 5 reports the estimates of the two-segment model.

[Insert Table 5 about here.]

We find that forward-looking behavior exists for both segments. Segment 2 customers, whom we conjecture to be more likely business oriented, have a higher exponential discount factor, which implies that they are more patient than segment 1 customers (0.91 vs. 0.86).[13] Of particular interest, discount factors in both segments are much smaller than those typically used in empirical studies (mean=0.979, see Table 1). Placing this result in perspective, a weekly discount factor of 0.995 implies that a consumer values a one-minute phone call at the beginning of the month to be worth about 1.02 minutes at the end of the month (under the assumption of a constant pricing rate). In contrast, our estimates indicate a minute phone call now is worth closer to 1.6 to 2.1 minutes at the end of the month (depending on the consumer segment).

Like Dubé et al. (2010b), we do not find strong support for the existence of hyperbolic discounting in our context. For both segments, the estimates of hyperbolic discount factors are approaching 1. Since the exponential discounting model is nested within the hyperbolic discounting model with $\beta_g = 1$, we also implement the nested Likelihood Ratio test for the two specifications. We can not reject the null hypothesis that the hyperbolic discount factors are not different from 1. Such a result is consistent with Chevalier and Goolsbee (2005) and Dubé et al. (2010b). Although time-inconsistent preferences and hence hyperbolic discounting exist (Angeletos et al., 2001), it may not be universal in all contexts.

[13] We reparametrize the discount factors as $\delta_g = \exp(\pi_g)/(1 + \exp(\pi_g))$ during estimation. The same treatment applies to β_g.

Between the two segments, segment 2 (11.2%) has a higher average consumption level, lower variance in usage, and is less price sensitive. This is also consistent with our intuition about the nature of more business oriented usage. Such a pattern makes segment 2 customers potentially better candidates for targeting with a higher price level. We will further explore the implication on firm pricing strategy in next section.

7 Managerial Implications

7.1 Usage Prediction and Intertemporal Substitution Pattern

7.1.1 Biased Price Effects

To asses the potential bias in model estimates arising from specifying the commonly employed discount factor of 0.995 rather than using the estimated heterogeneous discount factors, we re-estimate the model by fixing the discount factor to 0.995. While there is little impact on most estimates, we find the price coefficients to be underestimated (have smaller absolute magnitudes) by 23% in segment 1 and 15% in segment 2. The price coefficients of segment 1 and segment 2 become 2.31 and 2.25 (vs. 2.99 and 2.64 in Table 5), with the standard deviations as 0.03 and 0.04, respectively.[14] Setting a higher discount factor such as 0.995 implies customers excessively substitute future consumption for current within allowance consumption. Given future over-allowance consumption is costly, the model compensates by lowering price sensitivity to generate the same level of overall utility for the future consumption occasion. Consequently, the smaller price coefficients imply that the overage charge has less impact on future utility; as a result there is no need for the customer to make the tradeoff between consumptions across time.

[14] The estimates of remaining parameters can be obtained from the authors.

7.1.2 Biased Forecasts

To ascertain how well the model fits the data and resulting intertemporal substitution pattern, we calculate the mean absolute percentage error (MAPE) and mean percentage error (MPE) across segments and time under both the estimated discount rates and under the assumed weekly discount factor of 0.995. The MAPE measures a model's overall accuracy of fitting the data while the MPE indicates bias in model predictions. Table 6 and Table 7 depict the results.

[Insert Table 6 about here.]

According to Table 6, the fit under the 0.995 discount factor is universally worse than the fit under the estimated coefficients across segments and across time. To develop a better sense of why the higher discount factor performs more poorly, we next turn to the MPE.

[Insert Table 7 about here.]

Based on Table 7, there is no obvious forecasting bias from using the higher discount rate when summing across all periods, yet aggregating across time obscures the patterns in intertemporal substitution. When setting the discount factor at 0.995, the demand in the first 4 periods is under-estimated; and the demand in the last period is over-estimated. This bias occurs because that customers are more impatient than what $\delta_g = 0.995$ implies. As a result, in the early periods, when the allowance has not been exhausted, impatient customers consume more than predicted under 0.995 discount factor. Further, customers are more price sensitive than the 0.995 discount factor case implies (recall that the price coefficients are underestimated under the 0.995 discount factor). As a result, customers in overage (roughly coincident with the last period) evidence lower consumption than predicted under the 0.995 discount factor.

7.2 Elasticities

To ascertain how customers' minutes usage varies under alternative allowance and the marginal price levels, we compute their respective monthly minutes demand changes across segments for both the estimated discount factors and 0.995. Table 8 presents the results.

[Insert Table 8 about here.]

The elasticities in Table 8 suggest that the 0.995 discount factor leads to an underestimation of the effect of allowances and price on usage (that is, users are not as price sensitive as it implies when the discount factor is set to 0.995). Were consumers to actually have a discount factor of 0.995, they would be more forward looking than they were under the actual discount rates we estimate. As a result, the more forward looking consumers implied by 0.995 should conserve minutes so as not to pay overage in later periods. Because they do not actually conserve minutes, the model with a 0.995 discount factor needs to rationalize the observed overage. It does so by estimating a relative lower sensitivity to price and allowance; a lower price and allowance sensitivity means that consumers do not mind paying overage as much and have lower elasticities.

7.3 Alternative Pricing Schedule

Based on our communication with the data provider, their process of picking the three-part tariffs in this field experiment is ad hoc. There was no optimization consideration during the design of the experiment. As a result, the focal three-part tariff is not likely to be optimal for the firm in terms of maximizing its revenue. To access the potential for revenue improvement, we create a grid of alternative allowance and price levels for each segments. For each combination of allowance and price, we calculate the percentage of revenue change.[15] Table 9 reports the results.

[Insert Table 9 about here.]

Table 9 includes the current plan (allowance=800 minutes, price=RMB0.40). Surrounding the current plan, each column from left to right represents a 2-cent change in the marginal price for minutes in overage and each row from top to bottom stands for a 20-minute change in the allowance. As customers are heterogeneous, especially in their price sensitivity, the optimal price structure differs across the two segments. For both segments, a lower allowance enhances the possibility of overage; and a moderately decreased price tends to increase the consumption level under the overage situation. For segment 1, the customers are more impatient. Hence the optimal allowance level of segment 1 is relatively higher than that of segment 2 since the former are already more likely to be overage (780 vs. 760). Also, segment 1 are more price sensitive. So the corresponding optimal marginal price for segment 1 should be lower than that of segment 2 (0.36 vs. 0.38). The revenue of the firm would increase by 0.42% and 1.90% for segment 1 and segment 2, respectively. To the extent that similar exercises can be implemented across all groups of customers, the revenue increase would be considerable.

As shown earlier, under the discount factor of 0.995, the model may lead to biased estimates of coefficients and elasticities. To see whether such biases may lead to inaccurate policy recommendations, we re-create the same grid but calculate the revenue changes using the estimates under the 0.995 discount factor. Table 10 reports the results. As indicated in the Table, with the 0.995 discount factor, the model generates notably different pricing

[15] Note that we do not model plan choice since there are no plan choice data available. To ensure that the presented price/allowance changes do not lead to customer leaving the company or opt to a different plan, we calculate customer welfare for each point on the grid as measured by customers' total discounted utility. We then compare it with the welfare level under the original plan. None of the welfare changes is significantly different from zero, hence we do not believe that the recommended policies will result in substantial plan switching. Further, the company had a significant market share and there was no cellphone number portability in China until October, 2010 (ChinaTechNews.com, 2010). Consequently, we conclude (1) that the new price structures in the Table are unlikely causing large customer churning and plan switching, and (2) that the effect of competitive response is likely to be modest.

plan recommendations for both segments. Since the assumed discount factor is much higher, to enhance customers' likelihood of overage, the allowance levels would be much lower. As a result, under the 0.995 discount factor, the optimal allowance levels for segment 1 and segment 2 become 700 and 720 (instead of 780 and 760), respectively. Further, since the price sensitivities are underestimated, the optimal price would be higher. This effect manifests for segment 1, the optimal price changes from 0.36 to 0.38. In short, the firm sets its allowance too low and its marginal price somewhat high, thereby overcharging its customers when using the standard practice of setting discount rates.

[Insert Table 10 about here.]

The predicted revenue gains are also quite different between the two scenarios, ($\delta_1 = 0.86, \delta_2 = 0.91$) vs. ($\delta_1 = \delta_2 = 0.995$). To illustrate the difference, for each grid point of each segment, we calculate the predicted revenue difference between the two scenarios. As shown in the Figure 4, the percentage differences can be substantial. By implementing the pricing plans as suggested by the model with $\delta_1 = \delta_2 = 0.995$, the firm would foregone potential revenue gains. Take segment 1 as an example, the firm's revenue improvement would be 1.90% with an allowance of 760 and a price of 0.38 (where $\delta_1 = 0.86, \delta_2 = 0.91$). Instead, if the firm adopted the plan with an allowance of 720 and a price of 0.38 (where $\delta_1 = \delta_2 = 0.995$), the revenue improvement would only be 0.55%, which would reduce the potential gains by 74% relative to the optimal pricing level. As for segment 2, the corresponding loss would be even higher, at 88%. It is interesting to note that the potential bias is more substantial for the business oriented segment with its relatively lower level of price sensitivity.

[Insert Figure 4 about here.]

8 Conclusion

Owing to the ability to capture the trade-off between long-term and short-term goals, the application of dynamic structural models to consumer and firm decision making has become increasingly widespread. However, dynamic structural models face a fundamental identification problem, namely, the preference, the state transition, and the discount factor are confounded and become difficult to identify simultaneously. Should the rate be misspecified, inferences about agent behavior might be misleading and the implied policies for improving agent welfare might be suboptimal.

To address this problem, several solutions have been proposed. The most common approach has been to assume a fixed discount factor that is consistent with the market interest rate and is common across individuals. Given that consumers' discounting behavior may be inconsistent from the market interest rate and may vary across individuals (Frederick et al., 2002), it would be desirable to relax these assumptions.

A second identification strategy is to invoke the exclusion restriction condition (Magnac and Thesmar, 2002), where a set of exogenous variables affecting future state transition but not current utility. The variation in these exogenous variables then helps to separately identify the utility and discount factor. Though a promising approach, exogenous variables may not exist in some contexts and when they do, the exogeneity assumption may be hard to validate.

A third approach, pioneered by Dubé et al. (2010b), is to use experiment data that enable researchers to disentangle discount effects from changes in utility; and our research extends work in this vein. We augment this analysis by considering field studies, an approach that considers decisions made over a longer duration than regular lab settings to measure their trade-offs and involves monetary incentives on the scale of the choices.

Accordingly, we advance the literature on identifying heterogeneous discount factors by using field data to measure them. Specifically, we estimate a dynamic structural model using consumers' cellphone usage data. The data contain observations of consumers' cellphone consumptions under both static setting and dynamic setting. Using the static data, we first identify consumers' heterogeneous utilities and the distribution of random consumption shocks. Conditioned on the identified utilities and random shocks, we then recover the heterogeneous discount factors using the dynamic data.

Findings suggest that discount rates in practice (0.86 and 0.91 across segments) are well below those commonly assumed in the literature (0.995). As a consequence, price effects are underestimated in our application. Moreover, the higher rate leads to a mistaken presumption that more minutes would be saved for later use, leading to a 27% increase in the mean absolute percentage error in model fit. The attendant consequences for pricing policy are notable, leading to pricing recommendations that are generally too high and would lower potential revenue gains by 74-88%.

The inherent complexity of dynamic structural models often requires simplifications that correspondingly represent future research opportunities. Our model is no exception. First, we note that risk aversion may play a role in consumers' dynamic decision making. While the quadratic (concave) form of utility we estimate under linear pricing does not explicitly capture risk aversion (there is no uncertainty so there can be no risk), this form of utility does have risk implications for forward looking consumers. Given the quadratic utility, an increase in the forecast demand variation will strictly lower utility over the case where future demand is less volatile. Hence, riskier decisions have lower utility in our model. That said, it would be desirable to allow the utility function to accommodate differences between satiation (as in the linear pricing case) and risk aversion (as in the three-part tariff); our analysis, like all those that proceed it (e.g., Erdem and Keane (1996)), do not make this distinction. Future research should therefore consider to collect data that enable researchers to disentangle these two effects.

Second, our study focuses on a specific consumption context with a small focal group of customers over a specific duration. Therefore, the results may not generalize to other contexts involving different consumers, decisions or decision durations. Hence, more research is necessary to generalize the degree to which discount rates are an inherent trait or the degree to which they are context dependent. For example, it would be fruitful to consider how discount rates might vary in practice when one considers different contexts of intertemporal consumption; do consumers invoke the same level of patience when making choices over years as they do when making decisions over days?

Third, we do not explicitly consider learning, but rather control for it by incorporating a consumption effect for new users. While most users in our data are experienced, estimates from Table 5 indicate the few users who are new in segment 2 evidence lower consumption rates. Clearly, uncertainty plays a crucial role for new users in their plan choices and consumption levels (Lambrecht, 2006; Lambrecht et al., 2007; Iyengar et al., 2007). Accordingly, a more formal characterization of learning that also considers consumers' forward-looking behavior would provide more insights in contexts with a greater number of novice consumers.

Fourth, two potential sources of selection bias exist in our field study. The first arises from the firm's choices of customers to participate in the plans. Per our discussion with the firm's managers, customer selection was randomized so this form of selection bias is not germane. The second selection bias arises from the customer's decision of whether to adopt the three-part tariff plan that was offered. If the decision to adopt the plan is correlated with usage model error, then our estimates will be biased. This correlation could arise from unobserved person specific factors common to both choices, or omitted person time effects. Via inclusion of unobserved heterogeneity, we control for the former. Given plan choice is not a time varying decision over the duration of our data, the latter source of omitted factors are also not likely to be material. Regardless, one limit of our data is that we have no valid instruments to model plan choice, nor do we observe which agents declined the plan. Accordingly, to the extent selection does manifest, our analyses should be considered to be conditional on plan adoption. As a result, another area of interest is to extend our research into the domain of plan choice.

In sum, this paper is the first (to our knowledge) to provide field-study based evidence regarding the nature of discount rates that obviate the need for structural assumptions or exclusion restrictions to identify discount rates. Consistent with Dubé et al. (2010b) and Ishihara (2010), we find evidence that discount rates are substantially lower than those used in practice and that this difference is material from a policy perspective. Given the widespread use of dynamic models in marketing and economics, we hope our analysis will spark future work to pin down how such intertemporal trade-offs are made in practice.

References

Ackerberg, Daniel A. 2003. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review* **44**(3) 1007–1040.

Allcott, Hunt, Nathan Wozny. 2011. Gasoline prices, fuel economy, and the energy paradox. *Working Paper* .

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, Stephen Weinberg. 2001. The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *The Journal of Economic Perspectives* **15**(3) pp. 47–68.

Ascarza, Eva, Anja Lambrecht, Naufel Vilcassim. 2009. When talk is "free": An analysis of subscriber behavior under two- and three-part tariffs. *Working Paper* .

Chevalier, Judith, Austan Goolsbee. 2005. Are durable goods consumers forward-looking? evidence from college textbooks. *NBER Working Paper 11421* (4).

ChinaTechNews.com. 2010. China to start trial of mobile phone number portability in october 2010 .

Chung, Doug, Thomas Steenburgh, K. Sudhir. 2010. Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. *Working Paper* .

Dubé, Jean-Pierre, G unter J. Hitsch, Pradeep Chintagunta. 2010a. Tipping and concentration in markets with indirect network effects. *Marketing Science* **29**(2) 216–249.

Dubé, Jean-Pierre, G unter J. Hitsch, Pranav Jindal. 2010b. Estimating durable goods adoption decisions from stated preference data. *Working Paper* .

Dubin, Jeffrey A., Daniel L. McFadden. 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* **52**(2) pp. 345–362.

Erdem, Tulin, Michael P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* **15**(1) 1–20.

Fang, Hanming, Dan Silverman. 2009. Time-inconsistency and welfare program participation: Evidence from the NLSY. *International Economic Review* **50**(4) 1043–1077.

Fang, Hanming, Yang Wang. 2010. Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *NBER Working Paper* .

Frederick, Shane, George Loewenstein, Ted O'Donoghue. 2002. Time discounting and time preference: A critical review. *Journal of Economic Literature* **40**(2) pp. 351–401.

Gordon, Brett R. 2009. A dynamic model of consumer replacement cycles in the pc processor industry. *Marketing Science* **28**(5) 846–867.

Härdle, Wolfgang, Oliver Linton. 1994. Applied nonparametric methods. Robert Engle, Daniel McFadden, eds., *Handbook of Econometrics*, vol. 4, chap. 38. Amsterdam: Elsevier Science, 2295–2339.

Harrison, Glenn W., Morten I. Lau, Melonie B. Williams. 2002. Estimating individual discount rates in denmark: A field experiment. *The American Economic Review* **92**(5) pp. 1606–1617.

Hartmann, Wesley R. 2006. Intertemporal effects of consumption and their implications for demand elasticity estimates. *Quantitative Marketing and Economics* **4**(4) 325–349.

Hartmann, Wesley R., Harikesh S. Nair. 2010. Retail competition and the dynamics of demand for tied goods. *Marketing Science* **29**(2) 366–386.

Hausman, Jerry A. 1979. Individual discount rates and the purchase and utilization of energy-using durables. *The Bell Journal of Economics* **10**(1) pp. 33–54.

Hendel, Igal, Aviv Nevo. 2006. Measuring the implications of sales and consumer inventory behavior. *Econometrica* **74**(6) 1637–1673.

Hotz, V. Joseph, Robert A. Miller. 1993. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies* **60**(3) 497–529.

Ishihara, Masakazu. 2010. Dynamic demand for new and used durable goods without physical depreciation: The case of japanese video games. *Working Paper* .

Iyengar, Raghuram, Asim Ansari, Sunil Gupta. 2007. A model of consumer learning for service quality and usage. *Journal of Marketing Research* **44**(4) 529–544.

Kamakura, Wagner A., Gary J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* **26**(4) 379–390.

Kim, Jun, Paulo Albuquerque, Bart J. Bronnenberg. 2009. Online demand under limited consumer search. *Working Paper* .

Laibson, David. 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* **112**(2) 443–477.

Lambrecht, Anja, Katja Seim, Bernd Skiera. 2007. Does uncertainty matter? consumer behavior under three-part tariffs. *Marketing Science* **26**(5) 698–710.

Lambrecht, Bernd Skiera, Anja. 2006. Paying too much and being happy about it: Existence, causes and consequences of tariff-choice biases. *Journal of Marketing Research* **43**(2) 212–223.

Magnac, Thierry, David Thesmar. 2002. Identifying dynamic discrete decision processes. *Econometrica* **70**(2) 801–816.

McHugh, Richard. 1956. Efficient estimation and local identification in latent class analysis. *Psychometrika* **21** 331–347, 10.1007/BF02296300.

Misra, Sanjog, Harikesh Nair. 2009. A structural model of sales-force compensation dynamics: Estimation and field implementation. *Working Paper* .

Narayanan, Sridhar, Pradeep Chintagunta, Eugenio Miravete. 2007. The role of self selection, usage uncertainty and learning in the demand for local telephone service. *Quantitative Marketing and Economics* **5**(1) 1–34.

O'Donoghue, Ted, Matthew Rabin. 1999. Doing it now or later. *The American Economic Review* **89**(1) 103–124.

Phelps, Edmund S., Robert A. Pollak. 1968. On second-best national saving and game-equilibrium growth. *The Review of Economic Studies* **35**(2) 185–199.

Rust, John. 1987. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* **55**(5) 999–1033.

Rust, John. 1994. Structural estimation of markov decision processes. Robert F. Engle, Daniel L. McFadden, eds., *Handbook of Econometrics*, vol. IV. Amsterdam: Elsevier Science.

Sun, Baohong. 2005. Promotion effect on endogenous consumption. *Marketing Science* **24**(3) 430–443.

Warner, John T., Saul Pleeter. 2001. The personal discount rate: Evidence from military downsizing programs. *The American Economic Review* **91**(1) pp. 33–53.

Table 1: Example Discount Rates in Empirical Studies

Study	Choice Domain	Approach[1]	Discount factor	Period	Weekly Discount Factor	Implied Weekly Interest Rate (%)
Rust (1987)	Bus engine replacement	F	0.9999	month	$\to 1$	0.002
Hotz and Miller (1993)	Children/sterilization	E	0.65 (0.68)	year	0.992	0.83
Erdem and Keane (1996)	Laundry detergent	F	0.995	week	0.995	0.50
Ackerberg (2003)	Yogurt	E	0.98 (0.02)	week	0.98	2.04
Hartmann (2006)	Golf	F	0.99	day	0.932	7.29
Gordon (2009)	Personal computers	F	0.98	month	0.995	0.47
Kim et al. (2009)	Camcorders	F	1.0	minutes	1.0	0
Dubé et al. (2010a)	Video game consoles	F	0.9	month	0.976	2.49
Dubé et al. (2010b)	HD DVD players	P	0.7	year	0.993	0.69
Hartmann and Nair (2010)	Razors and blades	F	0.998	week	0.998	0.20
Chung et al. (2010)[2]	Salesforce compensation	E	0.95	month	0.988	1.20
Fang and Wang (2010)[2]	Mammography exams	E	0.72 (0.09) 0.80 (0.03)	two years	0.997 0.998	0.32 0.21
Ishihara (2010)	Video games	E	0.885 (0.006)	week	0.885	12.99
Mean					0.979	2.25
Std. Dev					0.034	3.76

Notes:
1. F in the "Approach" column indicates an assumed fixed value for the discount factor. The study labeled P estimates the discount rate using experimental data. E indicates the discount parameter is estimated by functional restrictions and/or the use of exclusion restrictions. The standard errors of the estimates are reported in the parentheses.
2. Chung et al. (2010) and Fang and Wang (2010) also consider hyperbolic discounting. We only report their results of exponential discount factors. Fang and Wang (2010) use two specifications in their estimation, hence we report two discount factors. Chung et al. (2010) obtain the discount rate via grid search so there is no sampling error to report. Note that the grid search approach yields estimated parameter distributions that are conditionally marginal with respect to the discount rate, which can lead to inefficient estimates.

Table 2: Three-part Tariff Plans

Access Fee (CNY)	Allowance (Minutes)	Marginal Price (CNY)	Number of Enrollees	Percentage
98	450	0.40	284	50.35
128	600	0.40	111	19.68
168	800	0.40	83	14.72
218	1100	0.36	50	5.92
288	1500	0.36	21	2.86
388	2500	0.30	15	2.31

Table 3: Summary Statistics

	Mean	Std. Dev.	Min.	Max.
Monthly Usage under Linear Plan/Allowance Level	0.92	0.46	0.02	2.22
Monthly Usage under Three-part Tariff/Allowance Level	0.96	0.35	0.01	1.63
Female	0.16	-	0	1
Rural Residency	0.41	-	0	1
Age (years)	36.18	7.45	19	58
New Customer (enrolled less than 12 months)	0.16	0.36	0	1

Table 4: Alternative Numbers of Latent Segments

	BIC
1 Segment	3411.11
2 Segments	**3381.32**
3 Segments	3405.01
4 Segments	3430.21

Note: Bold fonts indicate the best fit

Table 5: Estimates of Model Parameters

	Segment 1 (88.8%) Rural, Female	Segment 2 (11.2%) Urban, Male
	Estimate (S.E.)	Estimate (S.E.)
Satiation		
Constant	**4.99 (0.21)**	**5.29 (0.55)**
Price (cent)	**2.99 (0.02)**	**2.64 (0.03)**
New Customer	-0.06 (0.05)	**-0.30 (0.05)**
Age	0.03 (0.11)	0.29 (0.38)
Age^2	-0.02 (0.02)	-0.04 (0.06)
Std. Dev. of Shocks	**8.99 (1.91)**	**8.75 (2.31)**
Segment Parameters		
Constant	**1.83 (0.31)**	__
Rural Residency	**0.39 (0.08)**	__
Female	**0.51 (0.12)**	__
Discount Factors	**0.86 (0.04)**	**0.91 (0.05)**

Note: Bold fonts indicate the estimates being significant at 95% level.

Table 6: Mean Absolute Percentage Error (MAPE) Comparison

		Mean Absolute Percentage Error		
		Segment 1	Segment 2	Aggregate
Estimated Discount Factors	First 4 Periods	0.09	0.10	0.09
	Final Period	0.10	0.09	0.10
	Monthly Aggregate	0.10	0.10	0.10
0.995	First 4 Periods	0.13	0.12	0.13
	Final Period	0.13	0.15	0.14
	Monthly Aggregate	0.13	0.13	0.13

Table 7: Mean Percentage Error (MPE) Comparison

		Mean Percentage Error		
		Segment 1	Segment 2	Aggregate
Estimated Discount Factors	First 4 Periods	0.02	-0.02	0.01
	Final Period	$\to 0$	0.01	$\to 0$
	Monthly Aggregate	0.01	$\to 0$	0.01
0.995	First 4 Periods	-0.06	-0.05	-0.06
	Final Period	0.08	0.05	0.08
	Monthly Aggregate	0.02	$\to 0$	0.02

Table 8: Demand Elasticities

	Demand Elasticity w.r.t. Price		
	Segment 1	Segment 2	Aggregate
Estimated Discount Factors	0.10 (0.08, 0.11)	0.08 (0.07, 0.09)	0.09 (0.08, 0.11)
0.995	0.08 (0.07, 0.08)	0.05 (0.04, 0.07)	0.08 (0.06, 0.08)
	Demand Elasticity w.r.t. Allowance		
	Segment 1	**Segment 2**	**Aggregate**
Estimated Discount Factors	0.29 (0.27, 0.31)	0.08 (0.06, 0.10)	0.28 (0.27, 0.30)
0.995	0.25 (0.24, 0.26)	0.03 (0.02, 0.03)	0.24 (0.24, 0.25)

Note 1: 95% confidence intervals are in parentheses.

Note 2: Bold fonts indicate the biases are significant ($p < 0.05$).

Table 9: Revenue Percentage Change under Alternative Pricing Schedules ($\delta_1 = 0.86, \delta_2 = 0.91$)

Segment 1	Marginal Price						
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	-0.42	-0.34	-0.26	-0.30	-0.37	-0.39	-0.44
700	-0.01	0.02	0.05	0.09	0.07	0.01	0.002
720	0.16	0.20	0.24	0.29	0.27	0.22	0.14
740	0.23	0.24	0.25	0.21	0.17	0.16	0.12
760	0.33	0.37	0.34	0.32	0.30	0.29	0.25
780	0.37	**0.42**	0.40	0.39	0.37	0.34	0.31
800	0.04	0.05	0.02	0	-0.26	-0.56	-0.79
820	-0.23	-0.22	-0.26	-0.28	-0.85	-1.16	-1.37
840	-0.51	-0.46	-0.48	-0.53	-1.43	-1.74	-2.62
Segment 2	Marginal Price						
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	-0.72	-0.70	-0.67	-0.63	-0.60	-0.51	-0.55
700	-0.35	-0.22	-0.10	-0.07	-0.05	-0.09	-0.13
720	0.08	0.13	0.50	0.46	0.42	0.32	0.08
740	1.02	1.07	1.14	1.10	0.76	0.38	0.11
760	1.56	1.70	**1.90**	1.37	1.19	0.54	0.41
780	1.01	1.04	1.07	1.09	1.17	1.03	0.82
800	-0.11	-0.05	-0.02	0	0.21	0.48	0.29
820	-0.17	-0.14	-0.12	-0.10	-0.15	-0.55	-0.86
840	-0.25	-0.23	-0.21	-0.20	-0.32	-0.63	-0.91

Table 10: Revenue Percentage Change under Alternative Pricing Schedules ($\delta_1 = \delta_2 = 0.995$)

Segment 1	Marginal Price						
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	0.86	0.89	0.93	0.92	0.92	0.91	0.90
700	1.05	1.10	**1.13**	1.12	1.10	1.09	1.01
720	0.68	0.70	0.71	0.73	0.72	0.71	0.71
740	0.50	0.52	0.53	0.55	0.52	0.50	0.49
760	0.32	0.34	0.38	0.37	0.36	0.35	0.33
780	0.15	0.18	0.19	0.20	0.20	0.19	0.19
800	-0.02	-0.01	-0.002	0	-0.002	-0.01	-0.02
820	-0.25	-0.25	-0.25	-0.26	-0.27	-0.28	-0.30
840	-0.49	-0.49	-0.50	-0.51	-0.52	-0.54	-0.55
Segment 2	Marginal Price						
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	0.10	0.13	0.15	0.18	0.10	0.02	-0.04
700	0.16	0.19	0.22	0.24	0.17	0.09	0.01
720	0.28	0.32	**0.55**	0.38	0.30	0.28	0.15
740	0.22	0.25	0.31	0.28	0.24	0.16	0.08
760	0.04	0.17	0.21	0.22	0.14	0.06	0.01
780	-0.02	0.01	0.04	0.06	0.08	0.10	0.05
800	-0.07	-0.05	-0.02	0	0.02	0.04	0.01
820	-0.16	-0.14	-0.11	-0.08	-0.10	-0.17	-0.21
840	-0.25	-0.23	-0.21	-0.20	-0.19	-0.28	-0.37

Figure 1: Histogram of Total Usage vs Allowance

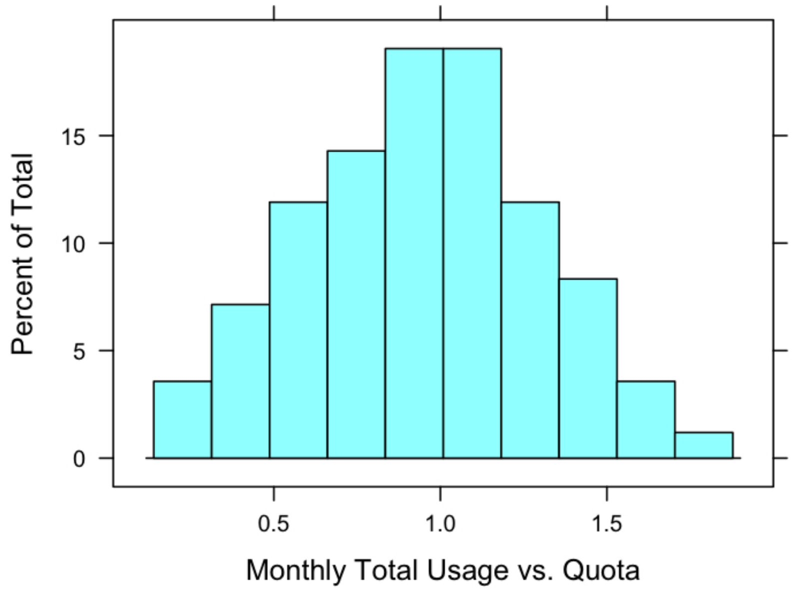


Figure 2: The Effect of Allowance on Minute Usage over Time

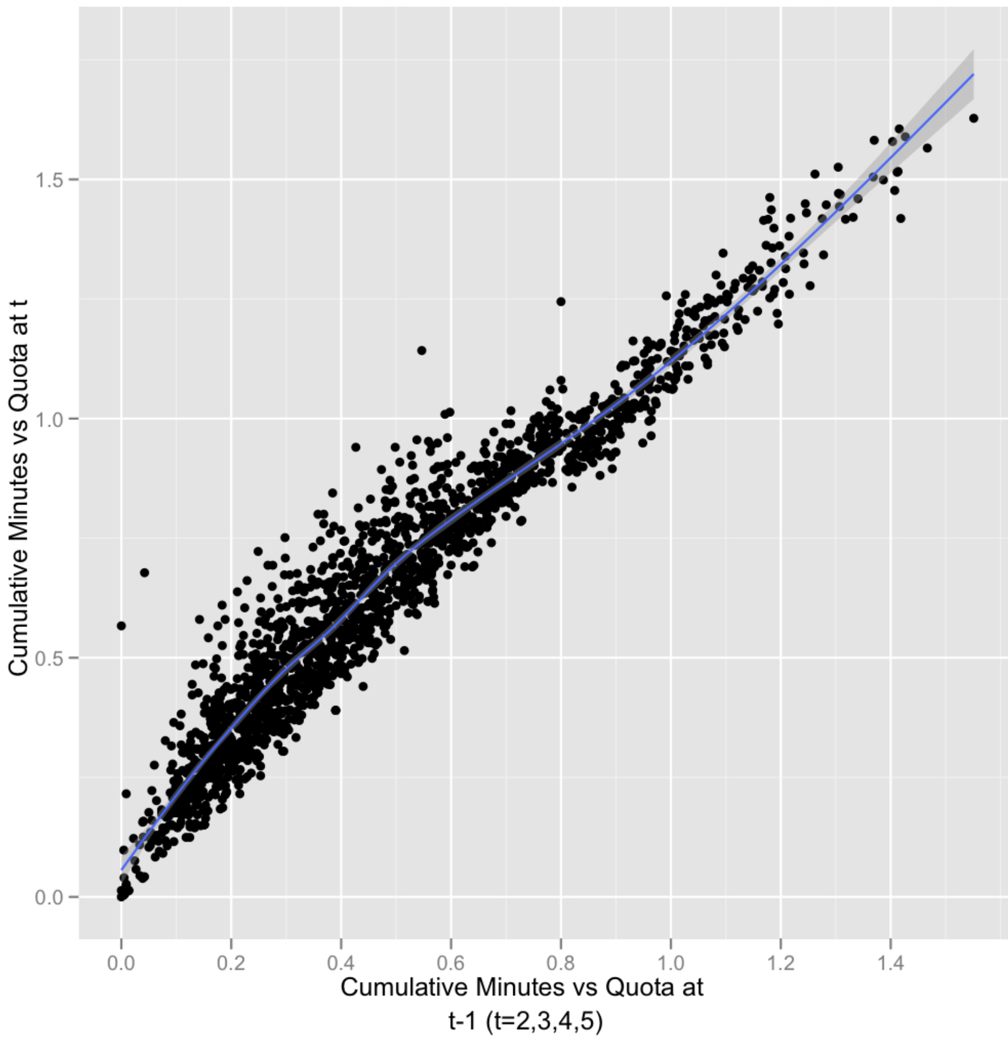


Figure 3: Usage Change over Time

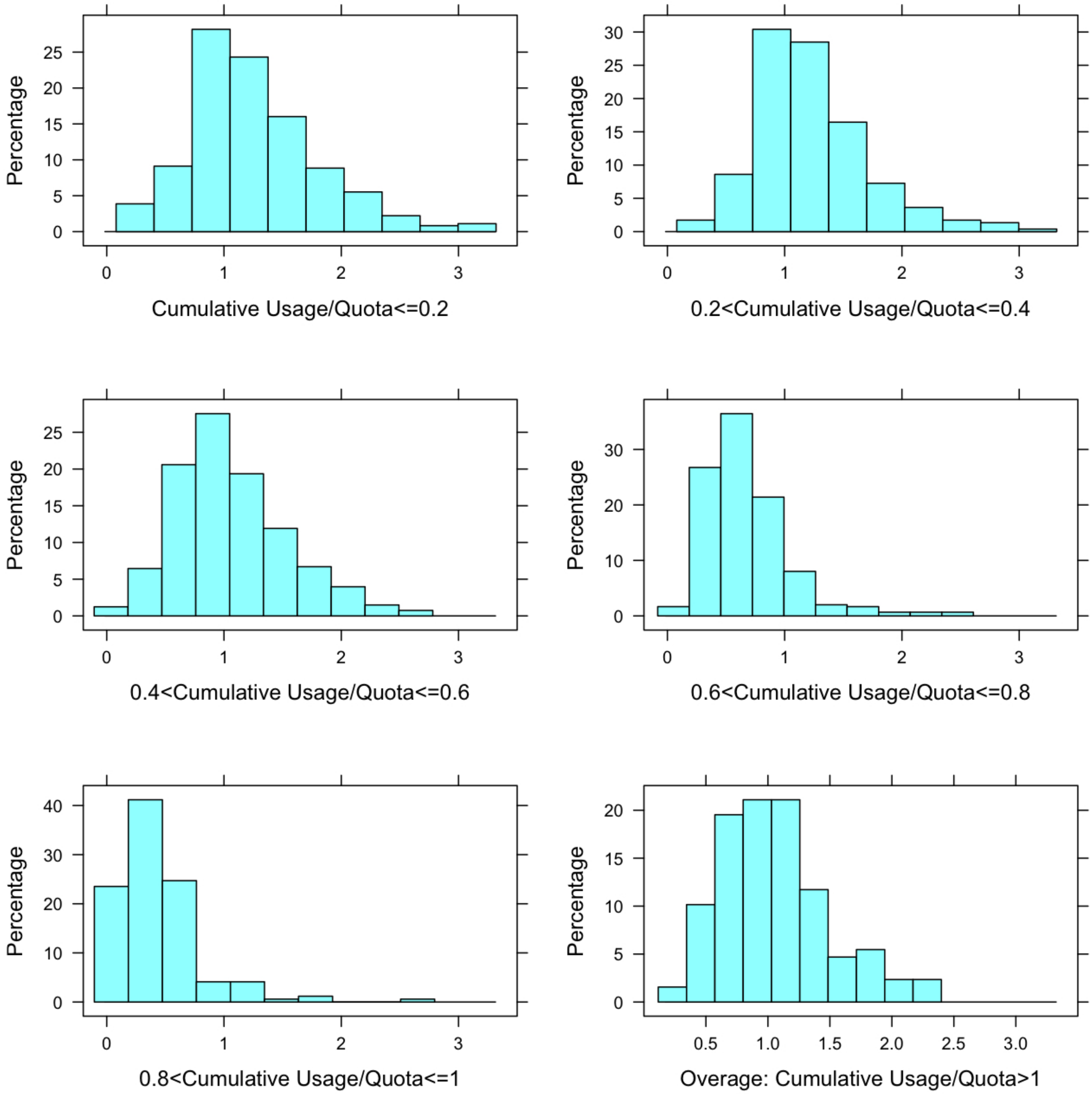
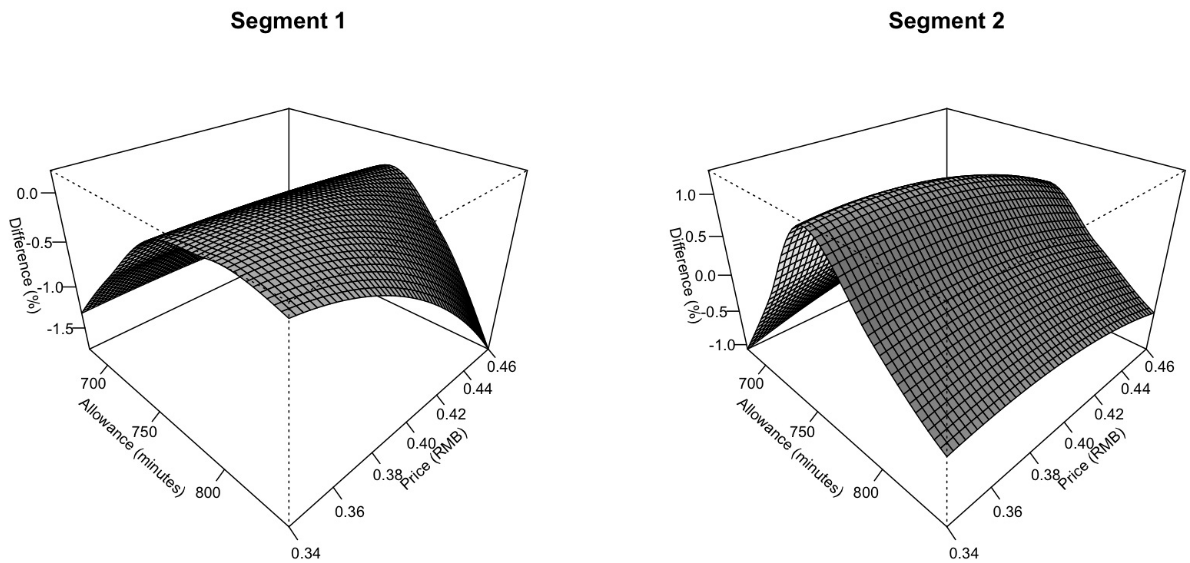


Figure 4: Revenue Prediction Differences: $(\delta_1 = 0.86, \delta_2 = 0.91)$ vs. $(\delta_1 = \delta_2 = 0.995)$



Appendix

A The Distribution of Monthly Minutes Usage $q_{i\tau}$ under Linear Pricing Plan

Since we only observe the monthly minute consumption $q_{i\tau} = \sum_t x^*_{it}$ but not each respective x^*_{it}, we have to find the likelihood of $q_{i\tau}$.

Suppose customer i is a member of preference segment g. As discussed in equation 5, the optimal minutes usage at period t, x^*_{it}, may take two values, 0 (if $d_{it} - b_g p_{i0} \leq 0$) and $d_{it} - b_g p_{i0}$ (if $d_{it} - bp_{i0} > 0$). Since $d_{it} = \exp(D'_{it}\alpha_g) + \nu_{it}$ and $\nu_{it} \sim N(0, \zeta_g^2)$, x^*_{it} follows a normal distribution $N(\exp(D'_{it}\alpha_g) - b_g p_{i0}, \zeta_g^2)$ that is truncated at zero. Thus the density of x^*_{it} is

$$f(x^*_{it}) = \frac{1}{\zeta_g}\phi(\frac{x^*_{it} - \mu_{it}}{\zeta_g})/[1 - \Phi(-\frac{\mu_{it}}{\zeta_g})] \tag{A1}$$
$$\text{where } \mu_{it} = \exp(D'_{it}\alpha_g) - b_g p_{i0}$$

The monthly minute consumption $q_{i\tau} = \sum_t x^*_{it}$ can then be written as the summation of a series truncated normal random variables with the same truncation at zero. Although there is no closed form for the distribution of $q_{i\tau}$, if the occurrence of zero minute consumption for any period is nearly zero ($\Pr(x^*_{it} > 0) \to 1$, $\forall t$), $q_{i\tau}$ can be approximated well by a normal density function that has the mean as $\sum_t \mu_{it}$ and the variance as $T\zeta_g^2$.[16] Intuitively, although x^*_{it} is a truncated normal r.v., if $\Pr(x^*_{it} > 0)$ is nearly one, the truncation becomes moot and the distribution of x^*_{it} can be approximated well by a normal distribution. The summation of a series of i.i.d. normal r.v.'s is also normally distributed.

[16] The analytical proof of the validity of this approximation and Monte Carlo simulation results can be obtained from the authors.

We implement a Monte Carlo simulation using the approximation mentioned above. The parameters are recovered with reasonable accuracy. As a robustness check to this approximation, we also use a Kernel estimator to compute the density in the Monte Carlo simulation (Härdle and Linton (1994)). The results are similar to the ones using the approximation but the Kernel estimation is much more computationally demanding.

Barriers to Entry in the Airline Industry: An Analysis of the Wendel H. Ford Aviation Act*

Connan Snider† Jonathan W. Williams‡
UCLA University of Georgia

February 2011

Abstract

In the airline industry, passengers pay higher fares at airports where a single carrier controls a high fraction of traffic. The economics of the industry suggests there is an inherent tradeoff between product quality and carrier size and concentration, making the welfare implications of these premia ambiguous. In this paper, we investigate the success of Congressional mandates aimed at increasing competition at highly concentrated major US airports. The mandates required airports above certain concentration thresholds to take concrete steps to ease and encourage new entry and expansion by smaller airlines, primarily by increasing access to airport facilities. We exploit a sharp discontinuity in the law's implementation to identify the effects of the law. In so doing we are able to shed light on the nature high fares at these concentrated airports. We find a statistically and economically significant decrease in fares resulting from an airport's coverage by the legislation. More specifically, we find that in markets where one (two) of the market's endpoints was (were) covered, fares dropped by 10% (20%). Moreover, most of this decrease has come from decreases in dominant carriers' fares. We also find that approximately half of this decline in fares is driven by the entry of low-cost carriers into new markets. We find little evidence that the fare declines have been accompanied by decreases in quality measures, with the exception of congestion related delays, suggesting the legislation has been welfare improving for consumers.

Keywords: Regression Discontinuity, Treatment Effect, Airline Industry, Barriers to Entry, Hub Premium. Airport Facilities.

JEL Codes: C14, L50, L13, L93.

*This paper has benefited from discussions with Federico Ciliberto, Jan Brueckner, Alon Eizenburg, Alfred Galichon, Yehua Li, Nick Rupp, and Steven Stern. Seminar participants at Clemson University, London Business School, NYU, and the Sauder School at the University of British Columbia provided valuable comments. We are grateful to Sharon Glasgow and Andrea Toney of the FAA as well as Liying Gu, Deborah McElroy, and A.J. Muldoon of the ACI-NA for assistance in collecting the data for this project and providing valuable industry insight. All remaining errors are our own.

†Department of Economics, UCLA, snider@econ.ucla.edu, ph: (310) 794-7104.

‡Department of Economics, University of Georgia, jonwms@uga.edu, ph: (706) 542-3689.

1 Introduction

One of the most enduring features of the post-deregulation U.S. airline industry has been the hub premium, the premium over average fares that large carriers command in markets from airports where they provide a large share of service. Though this phenomenon has been widely documented (e.g. Borenstein (1989)), its causes and consequences are still in question. To the extent that higher fares are a result of the exercise of market power, they are detrimental to consumer welfare and efficiency. On the other hand, there is substantial evidence that consumers value the large route network and high frequencies that dominant carriers often provide (e.g. Berry (1990)). To the extent that high prices derive from these quality factors, they benefit consumers. Because airport facilities, obviously a necessary input for the provision of air service, are scarce, increased concentration and market power go hand in hand with the airline scope and scale that consumers value and drive down costs. The relative contribution, and the optimality of the balance, of these factors is an empirical question.

In 2000, the U.S. congress enacted the Wendel H. Ford Aviation Investment and Reform Act for the 21st Century (AIR-21). A primary directive of the bill was to require airports, above a given level of concentration, to take concrete steps to ensure that new entrants had ample access to airport facilities.[1] Airport compliance requires filing a *Competition Plan* with the Federal Aviation Administration (FAA), detailing the steps taken. The FAA then reviews the plan and releases federal funding contingent on a satisfactory plan.

In this paper we empirically evaluate the impact of AIR-21 on prices, quality, and market structure in order to investigate the importance of access to airport facilities as barriers to entry in the airline industry. The nature of the implementation of AIR-21 is useful for solving identification problems that are common in industrial organization studies of competition and market structure and are present in our context. The problem is that elements of market structure (e.g. concentration, low cost presence, etc.) are determined simultaneously with the level of competition and usually depend on common, market-specific unobserved factors (e.g. demand elasticities or network economies associated with airport geography). We use the design of AIR-21 to formulate a differences in differences and regression discontinuity solution to these problems.

We first argue the AIR-21 mandates were enforced and effectively reduced barriers to entry at covered airports. This generates rarely available variation, with a plausibly known direction,

[1] The law applied to airports in which the top 2 airlines accounted for over 50% of total enplanements at the airport.

over time in barriers to entry within markets. This allows us to control for time invariant, market specific factors using standard panel techniques. Second, having contemporaneous treatment and control groups allows us to use differences in differences to address aggregate and market specific variation in these factors over time.

There are still likely to be selection problems associated with using the full sample for identification. Berry and Jia (2009), observing lower fares and diminished profit margins between the end of the 1990's and the middle 2000's, estimate discrete choice demand systems separately for 1997 and 2005 and conclude that increased passenger price sensitivity combined with increased penetration of low cost carriers were responsible for the change. Since airport concentration itself is likely highly correlated with product quality, the time varying relative valuations of quality found by Berry and Jia (2009) likely interact with our determinant of treatment. This causes differing average trends for treated and untreated markets, invalidating the simple diff-in-diff approach. We solve this and any similar such problem by arguing that, while there is likely a selection problem associated with highly concentrated airports, there is no such problem *locally* around the 50% two carrier concentration level specified by AIR-21. This allows us to develop a regression discontinuity estimator for the local average treatment effects associated with AIR-21. Essentially, we assume the distribution of unobservables for randomly selected market just below the cutoff is identical to a randomly selected market just above the cutoff.

The design of AIR-21 also helps us dismiss concerns about manipulation of the forcing variables. Airport coverage is determined by traffic data from two years prior to coverage, making coverage dependent on the past actions of the carriers which are not subject to manipulation. Given the complexity of airline pricing decisions it also seems unlikely that carriers would adjust fare setting behavior to manipulate enplanements at the airport level. Nevertheless, we design an informal test of manipulation. The test is based on the observation that, for a given two firm airport concentration level, airports with higher one firm concentration levels would be more likely to see manipulation since a single carrier has more control over the coverage variable. This test shows no evidence of manipulation.

Ultimately, we implement two RD estimators. The first takes Black (1999)'s boundary dummy approach, using observations from progressively smaller windows around the treatment cutoffs. The second is a novel, true RD estimator. Since airline markets necessarily involve both an origin and a destination airport, there are naturally two predictor variables and four treatment/control groups we have to consider when defining our treatment effects: Both the origin and destination are treated, just the origin is treated, just the destination is treated, and neither are treated. Our approach is to essentially estimate the surface for each of these "quadrants" and look for breaks at the boundaries of those quadrants. To the best of our knowledge, this is the first implementation of such a multi-dimensional design.

Our relatively clean identification strategy represents an contribution to the extant literature on airline market structure and the importance of barriers to entry more generally. A typical structural study of entry and market structure in concentrated industries (e.g. Bresnahan and Reiss (1991), Berry (1992) Mazzeo (2002), Seim (2004), Ciliberto and Tamer(2009)) looks at firm *choices* and uses a revealed preference approach to infer entry barriers. This approach necessarily requires the economist to rely on many restrictions of the empirical model derived from economic theory. Our approach, on the other hand, uses a known source of exogenous variation in entry barriers to investigate their effects on market *outcomes* and requires little in the way of theoretical structure. The minimal structural requirements is useful for an industry as complex as the airline industry and our focus on outcomes makes our results directly relevant for policy.

To preview results, we find AIR-21 had substantial, and evidently positive, impact on competition and fares in the airline industry. We find that markets for which one of the endpoint airports were subject to AIR-21 have seen price declines of 10% on average. Markets for which both endpoints were subject to the mandates have seen price declines of around 20% on average. These price declines were associated with no economically and statistically significant changes in measures of quality, with one exception. We find that the on-time performance of carriers at covered airports decreased. This is not particularly surprising, as we identify increased "low cost" penetration as a driving force behind the declines in fares, suggesting that increased competition at covered airports has resulted in additional congestion related delays. In addition, we find that the magnitude of the decline in fares is larger for carriers with a large presence at an airport than for other carriers. This suggests that AIR-21 was successful at reducing the hub premia identified by Borenstein (1989). All of our results indicate AIR-21 has been strongly welfare improving for passengers.

The remainder of the paper is organized as follows. In Section 2, we provide some background on the airline industry and discuss AIR 21 in detail. The data are described in Section 3 and we document some basic patterns in the data over the policy period. In Section 4, we discuss our identification strategy and the results of our analysis. Section 5 concludes and discusses possible extensions of our research.

2 The Aviation Investment and Reform Act for the 21st Century

The Government Accounting Office (GAO) and Transportation Research Board (TRB) released a series of reports, see GAO (1989, 1990, 2001) and TRB (1999), bringing attention to the limited amount of competition at many major US airports. These reports identified two types of barriers to entry in the airline industry, operating and marketing, that have the potential to limit competition and result in higher fares.

Marketing barriers include loyalty programs intended to tie consumers to an airline; frequent flyer programs, corporate incentive agreements, and travel agent commission overrides. A lack of data has limited the study of these type of barriers, Lederman (2007, 2008) and Goolsbee and Syverson (2008) as notable exceptions. Lederman (2007, 2008) finds evidence that improvements in loyalty programs enhance demand and can explain a modest portion of the "hub premium". Goolsbee and Syverson (2008) show that national carriers respond to the "threat of entry" by Southwest Airlines, a low-cost carrier, by lowering fares with the intention of strengthening consumer loyalties prior to entry of Southwest.

Operating barriers include limited access to boarding gates, ticket counters, baggage handling and storage facilities, and take-off and landing slots. Ciliberto and Williams (2010) were the first to directly link these operating barriers to the "hub premium". Using unique data on carrier-specific access to boarding gates, Ciliberto and Williams (2010) show that long-term exclusive-use leasing agreements for boarding gates are a major driver of the "hub-premium". In this paper, we employ a unique identification strategy to examine the success of AIR-21 in reducing these operating barriers and encouraging competition at major US airports. In the sections to follow, we discuss the details of AIR-21's design and implementation

2.1 Legislation and Airport Coverage

In response to governmental, public and academic concern with the existence of institutional barriers to entry in the airline industry, President Clinton signed into law AIR-21 on April 5, 2000. Section 155 of AIR-21 begins:

> *"The Congress makes the following findings:*
>
> *(1) Major airports must be available on a reasonable basis to all air carriers wishing to serve those airports.*
>
> *(2) 15 large hub airports today are each dominated by one air carrier, with each such*

carrier controlling more than 50 percent of the traffic at the hub.

(3) The General Accounting Office has found that such levels of concentration lead to higher air fares.

(4) The United States Government must take every step necessary to reduce those levels of concentration.

(5) Consistent with air safety, spending at these airports must be directed at providing opportunities for carriers wishing to serve such facilities on a commercially viable basis."

Together (1), (4), and (5) demonstrate Congress' clear intentions to reduce concentration by encouraging additional entry at concentrated airports. In order to encourage airports' cooperation in opening up airports to *"all air carriers wishing to serve those airports"*, Congress made federal sources of funding contingent on compliance:

"Beginning in fiscal year 2001, no passenger facility fee may be approved for a covered airport under section 40117 and no grant may be made under this subchapter for a covered airport unless the airport has submitted to the secretary a written competition plan in accordance with this subsection."

Passenger Facility Fees (commonly called PFCs) and Airport Improvement Program (AIP) grants are the primary sources of federal funding for the industry and make up a significant portion of capital (including maintenance) budgets for major airports.[2] PFCs were first authorized by Congress in 1990 and are tied to projects to preserve and enhance safety, reduce noise pollution, and provide opportunities for enhanced competition between carriers. The PFC ceiling, the maximum fee allowed by law, was increased from $1 to $4.50 between 1990 to 2001. This ceiling has not been increased since AIR-21 and is not indexed for inflation. AIP grants are part of a federal program to help cover costs for approved capital projects aimed at increasing safety and capacity as well as reducing environmental concerns.

A 2009 Airport Council International - North America (ACI-NA) study found that over 40% of airports' capital funding is drawn from PFCs (21.7%) and AIP grants (22.2%).[3] PFCs alone have funded $50 billion dollars worth of airport capital investments since 1990, including the addition and maintenance of passenger boarding gates and runways necessary to accommodate additional

[2] PFCs are charged by airlines at the time a ticket is purchased and are then transferred directly to the appropriate airports.

[3] A copy of the presentation describing this report is available from the authors upon request.

entry. An additional 30% of airports' revenues come from bonds which are often backed with future PFCs revenues. This substantial and stable revenue base allows airports to significantly lower the cost of borrowing and enjoy investment grade ratings. While the quasi-public status of many airports make it difficult to know their exact objectives, the strong dependence of airports' revenues on the federal government's control over the right to charge PFCs and distribute AIP grant funding would seem to imply strong incentives for compliance. All airports covered by AIR-21 are forced to file a Competition Plan with the FAA and the DOT, in turn, must certify the Plan as acceptable in order for funding to be released. [4]

Congress also made it clear that competition "plans" were to be implemented:

> *"The Secretary shall review any plan submitted...to ensure that it meets the requirements of this section, and shall review its implementation from time-to-time to ensure that each covered airport successfully implements its plan.....The Secretary shall ensure that gates and other facilities are made available at costs that are fair and reasonable to air carriers at covered airports...where a "majority-in-interest clause" of a contract or other agreement or arrangement inhibits the ability of the local airport authority to provide or build new gates or other facilities."*

In conversations with those at the FAA assigned to approve and ensure implementation of the competition plans, we learned that approval was not a certainty for any plan. In many cases, the plans were significantly revised after discussions between the FAA, DOT, and airport authorities to ensure the plans meet the goals of the legislation. After filing of the initial competition plan, airports were required to complete two updates (approximately 18 months apart) that demonstrate significant progress towards implementation of the competition plan. There are no mandatory steps after the second update for covered airports, unless the airport denies a carrier access to airport facilities or significantly amends an existing leasing agreement or enacts a new master-leasing agreement.

Section 155 continues:

> *"A competition plan under this subsection shall include information on the availability of airport gates and related facilities, leasing and sub-leasing arrangements, gate-use*

[4] The 44 airports required by AIR-21 to file a competition plan include: airport: ABQ, ANC, ATL, AUS, BNA, BUR, BWI, CLE, CLT, CVG, DAL, DCA, DEN, DFW, DTW, EWR, HOU, IAD, IAH, JAX, LAS, MDW, MEM, MIA, MKE, MSP, OAK, OGG, ONT, ORD, PBI, PHL, PHX, PIT, PVD, RNO, SAT, SDF, SFO, SJC, SJU, SLC, SMF, and STL. The majority (43) of the airports were immediately "covered" by the retroactive nature of the legislation. The only airport to be covered later was LAS in 2005.

requirements, patterns of air service, gate-assignment policy, financial constraints, airport controls over air- and ground-side capacity, whether the airport intends to build or acquire gates that would be used as common facilities, and airfare levels (as compiled by the Department of Transportation) compared to other large airports."

The typical competition plan ranges in length from 75 to 100 pages and contain a vast amount of information about the airports operations. Ciliberto and Williams (2009) use this information to demonstrate that Congress' focus on equal access to sunk airport facilities is not completely misguided. Using cross-sectional variation in gate allocations and leasing terms, Ciliberto and Williams (2009) are able to explain an economically significant fraction of the hub premium, with this fraction being larger at congested airports. In this paper, we focus on measuring any reduction in the hub premium resulting from coverage of an airport by AIR-21.

To identify the impact of AIR-21 on the hub premium, and fares more generally, we exploit the sharp discontinuity in the relationship between coverage and concentration:

".....'covered airport' means a commercial service airport....that has more than .25 percent of the total number of passenger boardings each year at all such airports.....at which one or two air carriers control more than 50 percent of the passenger boardings."

These concentration thresholds create treatment and control groups, airports "very near" either side of the discontinuity, which can be used to measure the impact of the legislation on competition.[5] An airport is covered by the legislation if it qualifies in both the size and concentration dimensions.[6] In Section 4, we discuss how we exploit this feature of the legislation using a regression discontinuity approach to measure a (local) treatment effect, or impact from coverage at the concentration cutoff. Tables 1 and 2 show the show the two-firm enplanement concentration and the fraction of total domestic enplanement at covered and non-covered airports, respectively. While concentration and size are positively correlated, it is far from a perfect relationship. For example, Newark (EWR) is covered while New York (JFK) is not. Similarly, San Fransisco (SFO) is covered while Los Angeles (LAX) is not.

[5] As with any analysis examining treatment effects, the treatment must be exogenously applied. In the context of our study, endogeneity of treatment might arise if airports are able to lower concentration of enplanements and/or total enplanements to avoid being covered by the legislation. In Section 4.4.2, we show that there is little or no support for the claim that enplanements were strategically manipulated by carriers with the intention of avoiding coverage.

[6] The discontinuity along the size dimension also presents an opportunity to identify an effect from coverage, but the small number of airports near this cutoff limit our ability to exploit this feature of the law.

2.2 Implementation of Competition Plans

The majority of the competition plans and subsequent updates are available on the respective airport's website. The details of each competition plan are too vast to review here. However, a 2006 FAA report highlights specific actions taken by airports in a variety of areas to increase competition.[7]

In terms of improving availability of gates and related facilities, airport responses included: asserting control over under-utilized gates, designating Competition Access committees, adopting more entry-friendly leasing terms, removing specific access protections for signatory carriers, streamlining a forced accommodation process. Specific actions included, Hartsfield-Jackson Atlanta International Airport (ATL) invoking recapture authority to convert a leased gate to common-use, Cincinnati-Northern Kentucky International Airport (CVG) negotiating conversion of exclusively leased gates to common and preferentially leased gates, and San Francisco International Airport (SFO) invoking a forced accommodation clause to ensure that temporary needs of new entrant airlines were met. In terms of subleasing agreements, covered airports also began to assert more control and oversight over sublease fees, terms, and conditions, impose sublease caps on administrative fees, review and/or pre-approve subleases, and notify carriers of gates available for subleases.

Improving access to passenger boarding gates were clearly the focus of a large proportion of each competition plan. However, covered airports put forth effort in a variety of other ways to increase competition. For example, both Charlotte Douglas International Airport (CLT) and San Antonio International Airport (SAT) implemented a marketing plan to attract additional low fare carrier service. In order to make more efficient use of existing common-use facilities, ATL now enforces maximum turnaround times. Oakland International Airport (OAK) installed common use ticketing equipment (CUTE) at ticket counters and gates so that all airlines operating there will use identical facilities, providing maximum flexibility to airport administrators. CLT reduced landing fees for non-signatory and new entrant carriers to the same level as signatory airlines. Nearly all covered airports implemented measures to record gate utilization, impose minimum-use standards, and notify airlines of gate availability in order to make more efficient use of existing gates. Many airports also amended majority-in-interest (MII) agreements to exempt capital projects necessary for competition from MII votes.

[7] This report is available through the FAA website at:
http://www.faa.gov/airports/aip/guidance_letters/media/pgl_04_08b_competition_highlights_2006.pdf

3 Data

3.1 Sources

The majority of our data for this study is taken from the Data Bank 1B (DB1B) of the U.S. Department of Transportation's Origin and Destination Survey for the years 1993 through 2008. The DB1B data is a 10 percent random sample of all domestic itineraries. The unit of observation is the passenger level. The data contains information on the ticketing and operating carrier, details of any connections made by the passenger, and the fare paid for the itinerary used by the passenger. Following Evans and Kessides (1994), we consider round-trip tickets to be two equally priced one-way tickets and drop any inter-line tickets. Due to key punch errors or redemption of frequent flier miles, there are some unusually large and small ticket prices in the DB1B data. For this reason, we drop any fares greater than \$2500 and less than \$25.[8] In addition, we drop itineraries with more than 6 coupons (4 connections) for roundtrip itineraries and 3 coupons (two connections) for one-way itineraries. Following Borenstein (1989), we define a market as travel between a unique airport-pair.

We also collected the enplanement data used by the FAA to determine coverage by AIR-21. There are significant differences between this data and the enplanement data that is publicly available through the DOT's T100 database. These differences arise because the T100 data does not include on-demand (e.g. charter flights) and in-transit (e.g. plane stops to refuel does not deplane) passengers which are a significant source of enplanements at many airports. The differences are significant enough that the determination of coverage for a handful of airports would change depending on the source of enplanement data.

Our final source of data is a survey conducted jointly with the ACI-NA. The survey, completed by 47% of all medium and large hubs, those enplaning more than 0.25% of all enplanments at primary airports in the US, focused on gathering information on carrier-airport specific leasing agreements for boarding gates. For each airport, we observe the total number of gates, number of gates leased by each carrier on an exclusive and preferential basis, and the number of gates reserved for common-use by the airport authority.

[8] We also drop all itineraries the for which the DOT questions the credibility of the reported fare, as indicated by the tktdollarcred variable.

3.2 Descriptive Statistics

We summarize the FAA and survey data for medium and large hubs in Tables 1 and 2. Column 1 in Tables 1 and 2 list the covered and non-covered airports, respectively. The second column of Table 1 lists the year in which each airport was first covered by the legislation. Due to the lag in data collection, coverage in any particular year is determined by enplanement data from two years earlier. For example, the set of airports first covered by the legislation in 2000 was determined using enplanement data from 1998. This is important for our purposes, since it would be very unlikely that an airline could perfectly foresee the details of the legislation two years in advance and manipulate enplanements to avoid coverage of a particular airport. Of covered airports, LAS was the only airport not covered retroactively by the legislation. In Section 4.4.2, we test whether the lack of a significant number of airports first covered in later years is due to potential manipulation of enplanements by carriers.

The next 3 columns of both Table 1 and 2 reports; the mean fraction of all US enplanements performed at the airport, the mean and max share of the top-2 carriers from 1998 to 2006 (determines coverage from 2000 to 2008). The maximum of the top-2 carriers' shares during this period serves as the predictor of coverage by the legislation. Thus, for each airport in Table 1 (2) this variable is greater (less) than .5. It is also important to note that coverage is not a proxy for the size of the airport. Examining the means at the bottom of Tables 1 and 2 for the fraction of all US enplanements, there is little difference in size between covered and non-covered airports. This is important as it alleviates some concerns over the homogeneity of the treatment and control groups in our analysis.

The final columns of Tables 1 and 2 report the fraction of gates reserved by the airport authority for common-use, fraction leased on a preferential or exclusive basis by legacy carriers, and fraction leased on a preferential or exclusive basis by low-cost carriers[9]. Examining the respective means in 2001 and 2008 of these variables at the bottom of Tables 1 and 2, there is little evidence that gates moved differentially at covered and non-covered airports. However, the large amount of missing data makes drawing any strong conclusions difficult. The lack of a significant movement in the allocation of gates for most airports from 2001 to 2008 suggests that the FAA and DOT largely followed the recommendations put forth by GAO (2001). GAO (2001) cautioned that AIR-21 should not be used as a means to force the divestiture of assets (e.g. boarding gates) from

[9]Low-cost carriers include B6, FL, F9, G4, J7, KP, KN, N7, NJ, NK, P9, QQ, SY, SX, TZ, U5, VX, W7, W9, WN, WV, XP, and ZA.

dominant carriers at an airport for two reasons. First, the reallocation of assets among competing carriers may have little to no benefit if the gates were not allocated to a low-cost competitor, see Brueckner (2010) and Ciliberto and Tamer (2009) for strong support for this statement. Second, service in smaller markets would likely be the first affected by divestiture of a dominant carriers assets. This is intuitive, we expect a firm to eliminate or cut service in the least profitable markets and significant economies of density in the industry, see Brueckner and Spiller (1994), ensures a strong correlation between profitability and size. The lack of a significant difference in the reallocation of gates among carriers at covered and non-covered airports foreshadows our finding that coverage by AIR-21 has little effect on the network of destinations offered out of an airport. It also suggests that if we are to find a significant effect from coverage by AIR-21 on other dimensions of service, it is due to more efficient use of existing assets (the focus of most competition plans) rather than a redistribution of assets among carriers.

Table 3 summarizes the variables we construct from the DB1B data and other sources, before and after AIR-21, separately for the set of covered and non-covered airports. To motivate our approach in Section 4 and emphasize the importance of controlling for trends in the data prior to coverage by AIR-21, we summarize the first difference for each variable. More precisely, for each variable, the difference before AIR-21 is calculated as the level in the first quarter of 2000 minus the level in the first quarter of 1993 while the difference after AIR-21 is calculated as the level in the first quarter of 2008 minus the level in the first quarter of 2001.

The majority of our variables are calculated at the market-carrier level, where we classify a carrier's service into two types, nonstop or connecting. For each type of service in a market, $Avg.Fare$ is calculated as the average fare across passengers choosing a type of service. 20^{th}% Fare, 50^{th}% Fare, 80^{th}% Fare are constructed similarly for different quantiles of the fare distribution for each carrier, market and type of service. Table 3 shows there has been a significant downward trend in fares in both covered and non-covered markets. However, prior to AIR-21 fares were falling less rapidly at covered airports, while after AIR-21, fares fell more rapidly at covered airports. These differential trends are strongest in the upper quantiles of the fare distribution. In Section 4, we attempt to identify a causal relationship between coverage by AIR-21 and these differential trends in fares, while controlling for a variety of time-varying covariates. $Nonstop$ is an indicator for whether or not a carrier's service is nonstop. $DistanceTraveled$ is the average number of miles traveled by passengers purchasing a type of service from a carrier in a particular market. For nonstop service, $DistanceTraveled$ is equal to the direct distance between the market endpoints. For connecting service, *DistanceTraveled* is strictly greater than the direct distance. *FractionRoutes* is the proportion of all the destinations offered out of an originating airport that a carrier offers some type of service on. This variable is intended to measure the extent of a carrier's network out of the originating airport.

From the DB1B data, we also construct a number of market-specific variables. Our measure of the hub premium in a given market is calculated as the difference between the fares charged for a particular type of service by the carriers with the largest share of enplanements at the origin and destination airports and the average of fares charged by all other carriers. For example, in the ATL (Atlanta Hartsfield) to CLE (Cleveland Hopkins), Delta and Continental are regarded as the dominant carriers (those with the largest share of enplanements), and *Avg. Hub Premium* is calculated as the difference between the average fare charged by Delta and Continental and the average fare charged by all other carriers. The hub premium measures for the different quantiles of the fare distribution are constructed similarly, replacing the average fare with the appropriate quantile. These variables are summarized in Table 3 and suggest that coverage is associated with a large decline in the hub premium. To measure the availability of nonstop service, an important dimension to service quality, we calculate *% Nonstop* as the percentage of passengers traveling nonstop in a market. In addition, we construct two measures of competition in a market, *Lcc Penetration* and *Number Firms*. *Lcc Penetration*, summarized in Table 3, is an indicator for whether or not a low-cost carrier is present in the market. As has been well documented, low-cost carrier penetration has been steadily increasing over the previous decade and typically results in intense price competition. In Section 4, and as the descriptives suggest, we show that in markets where one or both endpoint airports are covered by AIR-21, the low-cost penetration rate is significantly higher as a result of coverage. *Number Firms* is the total number of firms serving the market and is a commonly used measure of competition in the Industrial Organization literature, see Berry (1992) and Ciliberto and Tamer (2009).

We supplement the DB1B data with information on the frequency of departures from the DOT's T100 database and the frequency and severity of delays from the DOT's Airline On-Time Performance database. From these data, we construct two variables. *Departures* is calculated as the number of departures per quarter by a carrier on a particular flight segment and *%OnTime* is calculated as the proportion of flights that arrive 15 or more minutes late. In addition to those variables we construct from the DOT sources, we also collected data on both population and per-capita income for each MSA from the Bureau of Economic Analysis to serve as controls throughout our analysis.

4 Empirical Analysis

Our final sample includes data from all airports classified as a medium or large hub by the FAA (enplaning at least 0.25% of total domestic enplanements), including highly concentrated hubs, such as Minneapolis and Dallas. A legitimate concern here is that these highly concentrated airports are significantly different from the control group (non-covered airports) in both observable and unobservable ways. For example, since airport presence is known to be an important factor in airline quality, cost, and price competition it is troubling that we have no airports in the control group that are comparable in terms of presence measures. Similarly, unobserved airport features, such a geographic location, may affect the network economies of an airport leading it to be both highly concentrated and also have different competitive mechanics than less concentrated airports. The results from Berry and Jia (2009) also give an important example of the interaction of unobservable changes in consumer preferences, i.e. decreasing willingness to pay for quality, with observable airport presence differences.

To get around these problems we exploit AIR-21's sharp discontinuity at the 50% two carrier enplanement level. Broadly, we assume that the distribution of market level, our level of observation, unobservables changes smoothly across the policy discontinuity. That is, the unobservable features of a randomly chosen market just below the cutoff has the same distribution as the unobservable features of a randomly chosen market just above the cutoff.

With this identification strategy in mind we estimate the local average treatment effects (LATEs) of the law using two approaches. First, we proceed in the spirit of Black (1999) and estimate a series of difference in difference regressions using only those observations in progressively smaller windows around the concentration cutoffs determining coverage. Figure 1 demonstrates this approach. We begin by utilizing the complete sample and then examine the subset of markets within narrower windows around the coverage cutoffs. Using this approach, we identify market outcomes that are impacted by coverage in a statistically and economically significant manner. This approach also allows us to use covariates to control for observable differences in airports and markets. This is potentially useful because, while we have a large number of markets, these markets are drawn from a relatively small number of airports, which may create a small sample problem even if our identifying assumption is correct. For example, New York (JFK) is always included as a control airport and serves markets that are larger, richer and more distant on average than those in the treatment group and due to the large number of markets originating or terminating at the airport it represents a nontrivial fraction of the sample.

In the second step, we employ a true regression-discontinuity approach and allow the window width to collapse to zero. We find that our main conclusions from the first step are robust. In addition, the regression-discontinuity approach allows us to examine variation in the effect of coverage along the cutoffs. This is important as we are able to identify particularly influential steps taken by airports, including those in the control group (ie. JFK), with regards to gate availability for low-cost carriers.

Given our above discussion, an IV design, estimating the effect of airport concentration on fares using the RD assumptions for identification of the first stage, would seem natural. This is not what we do here. First, we think of airport level concentration as being a very noisy measure of the average market level competitive intensity of that airport. Also, we do not feel comfortable excluding the treatment variable from, for example, the equation determining fares and will discuss this more below. For completeness, we include Figure 3 shows the changes in airport concentration. The vertical axis shows the two firm concentration in 2000 and the horizontal axis shows the same measure in 2008. Airports above the 45 degree line saw concentration decrease, while those below it saw concentration increase. Airports in the untreated group appear to have actually decreased in concentration, however, interpretations here are fragile. Most of the cloud of points that increased from around .5 concentration to .6 concentration in the middle of the figure, did so because a low cost carrier entered and/or expanded so much over the period that it came to have a large share.

4.1 Fares

Following Black (1999), we begin under the assumption that coverage is exogenous and homogenous in its effect on fares by estimating the following regression:

$$\Delta_t \log(avg_{ijmt}) = \Delta_t x_{ijmt}\beta + \Delta_t z_{mt}\gamma + \psi Nonstop_{ijmt} + \tau_1 1[1 \text{ cover}_m] + \tau_2 1[2 \text{ cover}_m] + \Delta_t \epsilon_{ijmt}, \quad (1)$$

using the complete sample. The dependent variable is the long second difference, the change from 2001 to 2008 minus the change from 1993 to 2000, of the logarithm of average fares paid by passengers who purchased product j (nonstop or connecting service) from carrier i in market m, where the second difference is constructed identically to the descriptives in Table 3. The vectors Δx_{ijmt} and Δz_{mt} include the second differences of $FractionRoutes$, $DistanceTraveled$, and the population and per-capita income at the market endpoint airports.[10] In addition, we include an indicator for nonstop service to capture the possibility that fares for nonstop service changed differentially relative to connecting service. Using second differences allows us to control for market level linear trends, whose distributions may differ across treatment and control groups. Specifications using first differenced data indicate a small but statistically significant difference in average trends between the control and treatment groups. This suggests using the first differenced data, as we do in the true RD design, will provide conservative estimates.

To capture the impact of coverage by AIR-21 on the time-path of fares, we include indicators for whether one or both of a market's endpoints were covered, $1[1\ \text{cover}_m]$ and $1[2\ \text{cover}_m]$, respectively. Under the assumption that coverage is exogenous and homogenous in its effect on fares, τ_1 and τ_2 measure the causal effect on the dependent variable in a market with one and two endpoints covered, respectively. In order to relax these assumptions and ensure a causal interpretation of τ_1 and τ_2, we estimate the same regression on the subsamples of markets in progressively smaller windows around the coverage cutoffs. For such an approach to give consistent estimates, a significant portion of the data must be located within these windows. Figure 2 gives the number of observations for each combination of the predictors of treatment. The histogram shows that the majority of the data is in fact immediately around the coverage cutoffs. This is of particular importance as we shrink the window further in the regression-discontinuity analysis.

The estimates of Equation 1 are presented in Columns 1, 3, and 5 of Table 4. Robust standard errors are calculated by clustering at the market level to account for the interdependence of observations within a market. Our estimates of τ_1 and τ_2 are negative and statistically and economically significant. From Column 5, where we can reasonably interpret our coefficients in a causal fashion, the results indicate that coverage of a single endpoint by AIR-21 results in a approximately a 10% reduction in average fares, while coverage of both endpoints results in approximately a 20% change in average fares. This result is robust across different window widths, suggesting that unobservable differences across airports that may drive selection into the treatment and control groups are not significant. The remaining results in Columns 5 are straightforward to interpret, we find that fares for nonstop service declined more rapidly than those for connecting service, carriers with a larger market presence are able to charge higher fares, and less direct connections are more costly to provide. These results are robust across subsamples.

[10] See Berry (1990), BCS (2006), and Berry and Jia (2010) for a discussion of the impact of the size of a carrier's network on demand for that carrier's services.

We interpret the results in column 1,3, and 5 as estimates of the total effect, both direct and indirect, of AIR-21 on average fares. Since we typically think of the effect of barriers to entry on fares as being an indirect one, it is important to the credibility of our identification as well as our evaluation of the policy to try to understand the direct channels through which fares are affected. Columns 2, 4, and 6 of Table 4 show specifications designed to partially get at these channels. Specifically, we add a very similar set of time-varying regressors to those employed by Borenstein and Rose (1994) to control for any changes in the competitive environment, including quality measures, in a market. The set of controls includes: airport-level enplanement Herfindahl indices for both endpoints, market-level enplanement Herfindahl indices, market shares for each carrier, an indicator for whether the carrier offers both nonstop and connecting service, the number of competitors in the market, and an indicator for whether or not a low-cost carriers serves the market. One can then interpret changes in the estimates of τ_1 and τ_2, when the controls are included, as evidence that variation in the competitive environment explains some portion of the estimated effect of coverage. The results from these regressions, are presented in Columns 2, 4, and 6 of Table 4. We find that these controls are able to explain between 40% and 50%, depending on the window width, of the effect from coverage we estimated in Columns 1, 3 and 5. The table only shows the coefficients on $LccPenetration$ and $NumberofFirms$. Of all the variables included in the specifications, only the measure of lcc penetration has an economically meaningful effect on fares. As the table indicates, the estimate of the effect is large, up to around 17%. While

Table 5 presents our results when we re-estimate Equation 1, replacing the dependent variable with various quantiles of the fare distribution. For conciseness and due to the similarity of the estimates, we present the estimates for the subset of coefficients of particular interest. Again, Columns 2, 4, and 6 (1, 3, and 5) present the our estimates with(out) the Borenstein and Rose (1994) controls. Consistent with the descriptive evidence in Table 3, we find that the estimated decline in fares resulting from coverage by AIR-21 is increasing in the fare quantile. Column 5 shows that the 20% fare declined approximately 2% (4%) in markets when one (both) endpoint(s) was (were) covered, compared to 7% (13%) and 13% (24%) for the median and 80% fares, respectively. We also find additional supporting evidence for our conclusions reached from the results in Table 4, in particular, that low-cost penetration drove a significant portion of the decline in average fares in covered markets. Low-cost carriers typically target price sensitive consumers when setting fares. As a result, we would expect to observe that the Borenstein and Rose (1994) controls, specifically the low-cost penetration indicator, would explain a larger proportion of the estimated effect from coverage for lower fare quantiles. Column 6 of Table 5 shows that inclusion of these controls completely explains away the coverage effect for the 20% fares, while explaining less than half of the coverage effect for the 80% fare.

The last measure of the mandates' impact on fares we look at is the effect on the hub premium. We measure the hub premium as the difference in the logarithm of the fare charged in a market by the carriers with the largest presence at the market's endpoints with that of its competitors. These premia range from roughly 15-40% in 2000 and, on average, are sharply increasing in the concentration of an airport. Table 6 reports the results of the regressions. The results are consistent across different window widths. If we focus on Column 3 of Table 6, the narrowest window, we find that these premia have fallen significantly faster in markets with one or both endpoints covered. This decline is larger for the upper tail of the fare distribution. More precisely, the premium on the 20% fare declined (9%) 15% in markets with one (both) endpoint(s) covered, while the hub premium on the 80% fare declined (12%) 28% in markets with one (both) endpoint(s) covered. The declines in the hub premium across the entire fare distribution suggests that AIR-21 was successful in reducing operating practices that gave an advantage to dominant carriers.

4.2 Quality

In addition to fares, many other characteristics of service may change as the result of coverage by AIR-21. GAO (2001) suggests that granting authority to regulators to force dominant carriers at certain airports to divest critical assets (e.g. boarding gates) introduces uncertainty and can lead to disinvestment in an airport. In particular, GAO (2001) suggests that smaller markets would be the first to be affected, possibly losing service altogether. If fare reductions are accompanied by diminished service quality, then the welfare consequences of coverage is ambiguous. We focus our attention on four critical dimensions of service quality, the availability of nonstop service (percentage of passengers flying nonstop with a carrier in a market), frequency of service (number of departures in a quarter by a carrier on nonstop flight segments), the on-time performance of carriers (percentage of flights arriving 15 or more minutes late by a carrier on nonstop flight segments), and the number of markets served by a carrier out of an airport (number of destinations served on a connecting or nonstop basis by a carrier out of an airport).[11]

To estimate the impact of coverage on the availability of nonstop service, we estimate the

[11] For a detailed discussion of those dimensions of service quality that have been shown to be the most important to consumers, see Berry (1990), Berry, Carnall and Spiller (2006), and Berry and Jia (2010).

following regression::

$$\Delta_t \% Nonstop_{imt} = \Delta_t z_{mt}\gamma + \tau_1 1[1 \text{ cover}_m] + \tau_2 1[2 \text{ cover}_m] + \Delta_t \epsilon_{imt}$$

where $\Delta\% Nonstop_{imt}$ denotes the second difference, constructed identically to the dependent variable in Equation 1, in the fraction of passengers flying nonstop in market m. To examine the impact of the coverage on the frequency of service and severity of delays on any nonstop flight segment s, we estimating the following regressions:

$$\Delta_t \log(Departures_{ist}) = \Delta_t z_{st}\gamma + \tau_1 1[1 \text{ cover}_m] + \tau_2 1[2 \text{ cover}_m] + \Delta_t \epsilon_{ist}$$

and

$$\Delta_t \log(\% OnTime_{ist}) = \Delta_t z_{st}\gamma + \tau_1 1[1 \text{ cover}_m] + \tau_2 1[2 \text{ cover}_m] + \Delta_t \epsilon_{ist}$$

, respectively, where *Departures* is the number of departures made by carrier i and %OnTime is the fraction of flights by carrier i that arrive 15 or more minutes late. Finally, in order to capture any potential divestiture by carriers in an airport resulting from coverage by AIR-21, we estimate the following regression:

$$\Delta_t \log(\# Routes)_{iat} = \Delta_t z_{at}\gamma + \tau 1[1 \text{ cover}_a] + \Delta_t \epsilon_{iat}$$

where the unit of observation is at the carrier-airport (a) level.

The results of these regressions are presented in Table 8. Robust standard errors are calculated by clustering at either the market, nonstop segment, or airport level to account for the interdependence of observations within the respective group. With the exception of delays, we find no significant declines in the quality of service. With regards to delays, we find a statistically significant increase in the proportion of flights arriving 15 or more minutes late. This is not particularly surprising given the results of Mayer and Sinai (2003) which finds that carriers controlling the majority of the operations at an airport have an incentive to internalize congestion related delays. This result does make conclusions regarding improvements in consumer welfare as a result of the legislation less clear. However, it seems very unlikely that an increase in congestion related delays, which tend to be mild in length relative to weather related delays, would completely offset a 20% reduction in the average fare. Forbes (2008) estimates a one minute increase in delays causes fares to decline by \$1.42 on average and the average delay is around eight minutes. The average fare in our sample is \$262 implying a decline as a result of AIR-21 on the order of \$25. A hedonic interpretation of the \$1.42 number then suggests the back of the envelope order of magnitude for an equivalent increase in average delay would be 17.5 minutes, an increase of over 100%. However, the significant increase in delays also suggests that the lobbying efforts of the ACI-NA and other trade organizations to raise PFC ceilings (or at least adjust the current ceiling for inflation) in order to expand airport facilities at the most congested airports is not misguided.

4.3 Competition

The results in Tables 4 and 5 suggest that increased competition, particular by low-cost carriers, explains a significant portion of the decline in fares in covered markets. However, it is not clear whether this increase in competition is driven by coverage. To test whether the steps taken by covered airports had a significant impact on the number and identify of firms, we estimate two regressions:

$$\Delta_t \log(\#Firms_{mt}) = \Delta_t z_{mt}\gamma + \tau_1 1[1\text{ cover}_m] + \tau_2 1[2\text{ cover}_m] + \Delta\epsilon_{mt}$$

and

$$\Delta_t \log(LccPenetration_{mt}) = \Delta_t z_{mt}\gamma + \tau_1 1[1\text{ cover}_m] + \tau_2 1[2\text{ cover}_m] + \Delta\epsilon_{mt}$$

where the dependent variables in these regressions are the number of firms serving the market and an indicator for whether a low-cost carrier is present, respectively. Ciliberto and Tamer (2009) and Brueckner (2010) provide useful discussions of the intense level of competition that results from the presence of a low-cost carrier.

The estimates of the coefficients on the coverage indicators are presented in Table 8. Robust standard errors are calculated by clustering at the market level to account for the interdependence of observations within a market. We find that for markets with one (both) endpoint(s) covered there is a 0.10 (0.43) increase in the probability of a low-cost carrier serving the market. While there are obvious caveats in interpreting the coefficients of a second differenced linear probability model, at a minimum, this corroborates our finding that variation in the low-cost indicator played a major role in explaining between 40% and 50% of the reduction in fares as a result of coverage. Moreover, our binary measure of low cost penetration may understate low cost penetration at the intensive margin. We find AIR-21 has no significant impact on the average number of firms serving a market.

4.4 Regression Discontinuity Design

As discussed above there are many strengths associated with the approach of Black (1999). The results, however, rely on a number assumptions, including homogeneity of the coverage effect and exogeneity of coverage, to estimate the effects of coverage by AIR-21. These assumptions can be troublesome because more concentrated airports (those with two carriers enplaning more than 50% of the passengers) are treated while less concentrated airports are not. Therefore, any covariation between fares and concentration after the first quarter of 2001 (the time of the treatment) would be empirically indistinguishable from a treatment effect due to AIR-21. While these assumptions are difficult to formally test, it is possible to measure a local-average treatment effect (LATE) around the treatment cutoff in the absence of these assumptions using a regression-discontinuity approach. Examining treatment and control groups "very near" either side of the treatment cutoff allows us to disentangle those movements in fares that are a result of coverage from those that are simply due to correlation between fares and concentration. We discuss our approach below.[12]

Estimation of the LATEs here is complicated by the two dimensional predictor vector. Instead of a point, our LATE estimates are now functions of the market endpoints' concentrations. Figure 1 makes this clear. Our task is essentially to estimate a nonparametric surface in each quadrant of Figure 1, then look for evidence of statistically significant breaks along the cutoffs determining coverage.

Let $Y_{ijmt}(o,d), o,d \in \{0,1\}$ denote the outcome variable when the origin treatment status is o and the destination treatment status is d. For each observation, we get to observe one of the four possible values of the variable. When only one endpoint is treated we define the LATEs as:

$$\begin{aligned}
\tau^1_{orig}(P^{dest}_m) &= E[Y_{imt}(1,0) - Y_{imt}(0,0) | P^{orig}_m = .5 P^{dest}_m < .5] \\
\tau^1_{dest}(P^{orig}_m) &= E[Y_{imt}(0,1) - Y_{imt}(0,0) | P^{orig}_m < .5 P^{dest}_m = .5]
\end{aligned}$$

and when both endpoints are treated:

$$\begin{aligned}
\tau^2_{orig}(P^{dest}_m) &= E[Y_{imt}(1,1) - Y_{imt}(0,1) | P^{orig}_m = .5 P^{dest}_m > .5] \\
\tau^2_{dest}(P^{orig}_m) &= E[Y_{imt}(1,1) - Y_{imt}(1,0) | P^{orig}_m > .5 P^{dest}_m = .5]
\end{aligned}$$

Our definition of treatment effects is motivated by several considerations. First, are identification considerations. Our data is lumpy in the sense that the predictors of coverage do not vary within an airport, so for a sufficiently small window around a given concentration level all the markets in that window will be drawn from a single airport. For example, consider Dallas-Fort

[12] See Imbens and Lemieux (2007) for an introduction to RDD and Hahn, Todd, and Van Der Klaauw (2001) for a detailed discussion of identification of treatment effects within an RDD framework.

Worth (DFW) which has a predictor value of around 0.8, well away from the coverage cutoff. The estimate of $\tau^2_{dest}(0.8)$ compares the path of fares over the period since the passage of AIR-21 in markets originating at DFW and terminating at airports just below the coverage cutoff to those markets originating at DFW and terminating at airports just above the coverage cutoff. This approach allows us to control, to some extent, for fixed unobserved factors associated with given airports that are potentially distant from the coverage cutoffs. Second, in contrast to the window regressions, allowing the treatment effect to vary along the treatment cutoff in addition to the local linear regression implementation, discussed below, we are able to estimate the effect of coverage more flexibly. Figure 2 shows the large number of observations near the treatment cutoff, making such a flexible approach feasible. Moreover, Berry and Jia (2009) suggest there is direct evidence that the treatment effects may differ in airport concentration. Of course, the interpretation of our estimates as a flexible interactive effect is invalid if there is selection inherent in conditioning on the away-from-the-boundary-airport concentration level, which is likely given that a single airport will dominate any small bin. However, even in the presence of such selection, we can still interpret the estimates as an estimate of LATE heterogeneity where the heterogeneity corresponds to interaction with whatever is driving selection.

Our major task in estimation is to adapt the basic regression discontinuity framework to account for a two dimensional predictor vector. This requires flexibly estimating a two dimensional surface that relates Y_{ijmt} to $\{P^{orig}_m, P^{dest}_m\}$. Local linear estimators are particularly attractive for these type of problems, see Imbens and Lemieux (2007). At boundary points of the support for the predictor vector, local linear estimators do not suffer from the inherent bias of kernel estimators and achieve faster rates of convergence. In addition, local linear estimators are easily extended to multiple dimensions. Fan and Gijbels (1996) provides a detailed discussion of the advantages of local-polynomial modeling.

To demonstrate our approach, suppose we are estimating $\tau^1_{orig}(P^{dest}_m)$. This requires us to estimate the conditional expectation, $E[Y_{imt}(1,0)-Y_{imt}(0,0)|P^{orig}_m = .5 P^{dest}_m < .5]$, for each $P^{dest}_m < .5$. For a particular value of P^{dest}_m, $\overline{P}^{dest}$, the estimator is defined as

$$\tau^1_{orig}(P^{dest}_m) = \widehat{\alpha}^{c+} - \widehat{\alpha}^{c-}$$

where

$$\min_{\{\alpha^{c-},\beta^{c-}_{orig},\beta^{c-}_{dest}\}} \sum_{\substack{m: \\ \{P^{orig}_m < .5, P^{dest}_m < .5\}}} [Y_{imt}(0,0) - \alpha^{c-}_0 - \beta^{c-}_{orig}(P^{orig}_m - .5) - \beta^{c-}_{dest}(P^{dest}_m - \overline{P}^{dest})]^2 w^-_m \quad (2)$$

and

$$\min_{\{\widehat{\alpha}^{c+},\beta^{c+}_{orig},\beta^{c+}_{dest}\}} \sum_{\{P^{orig}_m \geq .5, P^{dest}_m < .5\}} [Y_{imt}(1,0) - \alpha_0^{c+} - \beta^{c+}_{orig}(P^{orig}_m - .5) - \beta^{c+}_{dest}(P^{dest}_m - \overline{P}^{dest})]^2 w^+_m \quad (3)$$

The weights, w^+_m, are calculated as

$$w^+_m = \frac{\phi\left(\frac{P^{orig}_m - c^{orig}}{h^{orig}}, \frac{P^{dest}_m - \overline{P}^{dest}}{h^{dest}}\right)}{\sum_{j:\, P^{orig}_j \geq .5, P^{dest}_j < .5} \phi\left(\frac{P^{orig}_j - c^{orig}}{h^{orig}}, \frac{P^{dest}_j - \overline{P}^{dest}}{h^{dest}}\right)}$$

where $\phi(\cdot)$ is the bivariate standard normal pdf and h^{orig} and h^{dest} are bandwidths. The weights, w^-_m, are defined similarly. This process is then repeated for a range of values for $\overline{P}^{dest}$ to get an estimate of the treatment effect, $\tau^1_{orig}(P^{dest}_m)$, along the entire treatment cutoff. The estimators of $\tau^1_{dest}(P^{orig}_m)$, $\tau^2_{dest}(P^{orig}_m)$, and $\tau^2_{dest}(P^{dest}_m)$ are defined similarly.

To simplify the choice of bandwidth in multiple dimensions, we transform the predictors of coverage prior to estimation to have mean zero and identify covariance matrix, see Pagan and Ullah (1999). This allows us to check the sensitivity of our results by varying a single factor of proportionality, k, such that both h^{orig} and h^{dest} are equal to

$$h = kN^{-\frac{1}{4+d}}$$

where N is the number of observations in the quadrant of interest.[13] We find our results to be insensitive to the choice of bandwidth.[14] The results presented in Figures 3, 4, and 5 and Tables 9 and 10 set $k = 3$, which allows for a great deal of flexibility, as we will discuss below, yet adequately smooths the surface.

Figure 3 shows the estimated surface for a representative region of the effect surface as well as the data used in estimation in the same region. The figure makes clear that our RDD is not as clean as many in the literature, in the sense that the discontinuity is not plainly visible. When

[13] In Equations 2 and 3,

$$N^+ = \sum_m 1[P^{orig}_m < .5, P^{dest}_m < .5]$$

and

$$N^- = \sum_m 1[P^{orig}_m \geq .5, P^{dest}_m < .5]$$

, respectively.

[14] We also explored cross-validation methods for choosing k, but we find it performs very poorly in our application by suggesting a bandwidth near zero that overfits the data. A few aspects of our application and the method; interdependence of observations, multiple dimensional predictor vector, and the slow convergence rate of the cross-validation method, make this an unsurprising result.

viewed from above, the data in the figure appear as a grid due to clumping of observations at a single airport. This motivates our two dimensional smoothing procedure.

Calculating asymptotically valid standard errors for our estimates is a nontrivial computational exercise for a number of reasons. First, we are estimating a nonparametric surface in multiple dimensions. Second, we are most interested in the estimates of this nonparametric surface at the coverage cutoffs. Finally, we must account for the dependence in our data resulting from markets having endpoints in common. For these reasons, we appeal to the resampling with dependent data literature to calculate asymptotically valid point-wise standard errors. For a detailed treatment of resampling techniques for dependent data, see Lahiri (2003). The clear dependence structure in our data makes application of these techniques straight-forward. We treat the sample as representative of the population and compute jack-knife standard errors where we leave out blocks of markets with a common endpoint. In particular, for each airport we find all markets with a common endpoint and drop them from the sample. Using the resulting sub-sample, we then reestimate the model. We repeat this process for each market and use the distribution of the estimates across subsamples to infer moments of the asymptotic distribution of our treatment effects.

4.4.1 Results

The results and conclusions of our RDD analysis are consistent and nearly identical to our findings using the window-regression approach. For this reason, we focus the discussion of our RDD results on the impact of coverage on fares and low-cost competition.

Figure 3 and Table 9 presents the results of our RDD analysis of fares. In examining fares, we follow a very similar approach as in the window regressions. The only difference is that we examine the first difference in fares within a carrier, market, and type of service (nonstop and connecting) since the passage of AIR-21. Precisely, we take the difference in the logarithm of fares in the first quarter of 2008 and 2001 to construct our dependent variables. This serves as a robustness check on our conclusions from the window regressions, ensuring that any differential trends in fares prior to the passage of AIR-21 are not driving our findings above. We find this not to be the case, which is important as it provides some validation for our approach to identifying the effect of coverage. To summarize the results, we find the decline in fares resulting from coverage is statistically and economically significant. We also find that the magnitude of the effect is greater for higher quantiles of the fare distribution. These are clearly evident in the surfaces plotted in Figure 3 and the statistical significance of the point estimates in Table 9.

As discussed above, one advantage of employing a true RDD approach in our application is the opportunity to look for heterogeneity in the effect of coverage. This heterogeneity is obvious as one looks at the point estimates of the effect from coverage on average fares in the top-left corner of Table 9. Looking at the estimates of $\tau^1_{orig}(P^{dest}_m)$, we find a 10% reduction in fares resulting from the destination being covered in markets where the origin has a two-firm concentration of 0.4 and only a 3.7% reduction in fares in markets where the origin has a two-firm concentration of .5. This result would not be interesting if it were not for the statistical significance of both estimates, since the latter estimate may suffer from the "curse of dimensionality". More precisely, when estimating a surface non-parametrically, the number of observations falling in any locally defined ball falls exponentially in the number of explanatory variables. This problem is exacerbated when attempting to estimate the surface at the boundary of the support for the explanatory variables. Thus, at first glance, this appears to be a discouraging result that is consistent across all quantiles of the fare distribution.

However, in this case, our estimates of the effect of coverage on low-cost carrier penetration provide a clear explanation for the heterogeneity in the effect of coverage on fares we observe along the coverage cutoffs. Figure 4 and Table 10 report these results. In the top half of Figure 4, we plot the surface which we use to compute the estimates of the coverage effect on low-cost penetration in Table 10. These surfaces are relatively smooth, with one exception, in which we observe a significant jump in the entry behavior of low-cost carriers. The difference in low-cost entry behavior over this portion of the predictors' support is large enough to generate a negative and statistically significant effect on low-cost entry behavior, the top-left portion of Table 10. This suggests that at least one airport not covered by AIR-21, near the coverage cutoff observed substantial low-cost entry from 2001 to 2008. By examining Table 3, we identified those airports near the cutoff and then re-estimated the surface excluding each airport, one at a time. Through this process, we identified JFK as the driver of this finding. The estimates of the effect of coverage on low-cost penetration, excluding JFK, are presented in the bottom halves of Figure 4 and Table 10. After JFK's exclusion, the effect of coverage on low-cost carrier penetration is now strictly positive along each portion of the treatment cutoff. This is not surprising, as JFK provided a low-cost carrier, JetBlue, unprecedented access to airport facilities throughout the period since the passage of AIR-21.

Collectively, the results of our RDD analysis both corroborate and provide additional insights to the findings gleaned from the window regressions. The RD estimates of the effect of going from no endpoints treated to one endpoint treated range from 3 to 10 percent. The effect of going from no coverage to both endpoints covered, if measured as the jump right at the vertex of the four surfaces is about 11-12 percent (3.7 plus 8.1 or 3.2 plus 7.7), depending on which two measures of the jump you add. While, strictly speaking, this is the only correct way to measure this effect, averaging over the surfaces suggests the effect is around 14-15 percent. Unlike the window regressions we don't impose symmetric effects for the one endpoint treated markets, however, table 9 suggests that this is not too far off. As expected, these numbers appear to be in accordance with the window regressions and are more conservative since they do not take account of pre treatment trends which are slightly higher for the treated airports. The same conclusions apply to the other measures of changes in the distribution of fares.

For low cost penetration, the magnitude of the estimates are slightly different, but the difference is as expected and is driven by the apparent negative pretreatment trend in lcc penetration in the treatment group relative to the control group as well as the problem of interpreting a second differenced binary dependent variable. These differences would tend to make the RD estimates more conservative than the window regression estimates. Overall, our RD estimates of the effects on lcc penetration are more sensible but still enormous and are somewhat fragile to the inclusion of (at least) JFK. The estimated effect of going from no coverage to one endpoint covered ranges from 7 to 25 percentage points when JFK is excluded from the sample and from -15 to 15 percentage points, with -15 percentage points corresponding to exactly the concentration level of JFK, when JFK is included. The jump at the vertex of the the surfaces is about 24 or 27 perctage points depending on which of the two measures are used. As is clear from figure 6, there is almost no effect of jumping from no coverage to both endpoints covered where the surfaces meet when JFK is included. The measurements are -.9 percentage points and 1 percentage point. As in the case of fares, the symmetry of one endpoint effects imposed in the window regressions, appears largely justified by the RD results.

In terms of additional insights, the RDD results first show that ignoring heterogeneity in the treatment cutoff is important. The JFK example demonstrates this idea. By assuming a homogenous effect from coverage on low-cost penetration, one actually infers the incorrect sign on the effect of low-cost entry over some range of the support for the predictors and severely biased estimates along the remainder of the coverage cutoffs. Second, the ability of the RDD analysis to be able to essentially identify individual airport specific treatment effects, provides further support for our conclusion that entry by low-cost carriers was the driving force behind the large declines in fares in covered markets.

4.4.2 Regression-Discontinuity Validity

Above, we have discussed why we are comfortable assuming there are no (local) selection effects associated with AIR-21. The validity of our identifying assumption also requires there be no problem with incentive effects. That is, that carriers do not manipulate enplanement levels to avoid treatment. There are a number of reasons why we believe this is a valid assumption. First, coverage is determined at the airport-level, not the airline-level. Therefore, no individual airline can manipulate enplanements and entirely determine coverage, rather it would take a cooperative effort on the part of airlines serving the airport. Second, coverage in each year was determined using FAA enplanement data from two years earlier. An airline(s) attempting to avoid coverage by the legislation would have been required to foresee the exact details of the legislation (including the exact enplanement cutoff) two years in advance of its passage. Finally, manipulating enplanements at any one airport, particularly a large airport, has significant costs to an airline in terms of adjusting traffic in its entire network.

Extending formal tests to check for the strategic manipulation of enplanements, see McCrary (2007), with a two-dimensional predictor vector is not immediately clear. However, we develop an informal test for manipulation of the predictors of treatment and provide evidence that little or no strategic manipulation of enplanements occurred. The test is based on the simple observation that those airports just below the coverage cutoff in which one carrier controls a larger proportion of the traffic will be most vulnerable to strategic manipulation of enplanements. For example, consider two airports where the two largest carriers enplane 49% of the passengers. Suppose at the first airport, the top carrier enplanes 35% of all passengers while the top carrier at the second airport enplanes 25% of all passengers. If an airline was attempting to avoid coverage of an airport by AIR-21 by manipulating enplanements, one would expect this to occur at the first airport. At the first airport, the largest carrier would have greater control in ensuring that the airport were not covered.

One way a carrier can lower enplanements is by raising fares. If a carrier was seeking to raise fares and lower their share of enplanements to avoid coverage, one would expect to see less of a drop in fares in markets near the coverage cutoff where one carrier has a larger share of enplanements. Figure 5 shows that there is no evidence to support a claim that enplanements were manipulated. In the top-half of Figure 5, we plot the joint density of the share of the two largest carriers and the share of the largest carrier for airports below the cutoff. Given the high correlation between these two variables, we are only able to plot the relationship between these variables and changes in fares (difference between average fare in the first quarter of 2008 and 2001), for a small range of values in the bottom-half of Figure 5. If carriers chose to strategically manipulate enplanements, we would expect to see a surface sloping up in the top carrier's share at airports closest to the cutoff. We find no evidence to support this claim.

5 Conclusions

High fares at concentrated airports have been a fact of life in U.S. air travel since the deregulation of the industry in 1979. The welfare implications of these high fares are ambiguous because consumers value both the size and scope, in the form of frequency and network size, of an airline when flying out of their home airport. However size and scope lead to market power due to scarce airport facilities. In 2000, the U.S. congress took a stand, deciding too much market power at highly concentrated airports was generating too much of the fare difference and enacted AIR-21. Among other things, these mandates required concentrated airports to take steps to increase competition and make airport facilities available to all carriers wanting to serve the airport.

In this paper we have provided evidence that the mandates were successful in encouraging new and intensified competition at its targeted airports. Moreover, we have found evidence that Congress was right in concluding that market power contributed too much to high fares from the perspective of consumers. That is, we find little evidence that competition significantly eroded quality provision, either directly by reducing large incumbent size or indirectly by disincentivizing high frequencies. The only unintended consequence of the legislation appears to be additional congestion related delays, which are unlikely to fully offset the substantial declines in fares.

Our quasi-experimental approach to analyzing the impact of barriers to entry is also somewhat novel in the Industrial Organization literature, see Angrist and Pischke (2010), and we think our clean identification strategy represents a significant contribution to it. However, our study also highlights some of the difficulties in implementing such a research design, see Einav and Levin (2010) and Nevo and Whinston (2010). While, we are able to explain between 40% to 50% of the decline in fares in covered markets, a result of intensified competition from low-cost carriers, it remains an open question to identify other determinants. Moreover, if we had arrived at more nuanced results, e.g. more significant declines in quality, we would need more structure to say much about the balance of welfare gains and losses.

The competition plans and subsequent FAA reports provide at least a subset of the actions taken by airports and seems to provide a good source for identifying other possible explanations. A couple candidates that seem likely to have some explanatory power are the reduction of landing fees for smaller carriers to the levels enjoyed by large presence incumbents as well as limits on subleasing fees that can be charged by one carrier to another for the use of under-utilized boarding gates. Both these steps, discussed in the majority of the airports' competition plans, have the potential to be a significant source of cost pass-throughs from carriers to consumers. In addition, carriers may simply reduce fares to generate outcomes that are consistent with the goals of AIR-21 in order to avoid additional oversight in the future. We leave more detailed investigation of these channels for future research.

References

[1] Angrist, Joshua, and Jorn-Steffen Pischke (2010) "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics", *Journal of Economic Perspectives*, 24(2), 3-30.

[2] Berry, Steven (1990) "Airport Presence as Product Differentiation", *American Economic Review P&P*, 80(2), 394-399.

[3] Berry, Steven (1992) "Estimation of a Model of Entry in the Airline Industry", *Econometrica*, 60(4), 889-917.

[4] Berry, Steven, Michael Carnall, and Pablo Spiller (2006) "Airline Hubs: Costs, Markups, and the Implications of Costumer Heterogeneity", in Darin Lee (ed.), *Advances in Airline Economics, Competition Policy and Antitrust*, Vol. I, Elsevier.

[5] Berry, Steven, and Panle Jia (2010) "Tracing the Woes: An Empirical Analysis of the Airline Industry", *American Economic Journal: Microeconomics*, 2(3), 1-43.

[6] Black, Sandra (1999) "Do better schools matter? Parental valuation of elementary education", *Quarterly Journal of Economics*, 114(2), 577-599.

[7] Borenstein, Severin (1989) "Hubs and high fares: dominance and market power in the U.S. airline industry", *Rand Journal of Economics*, 20(3), 344-365.

[8] Borenstein, Severin and Nancy Rose (1994) "Competition and Price Dispersion in the U.S. Airline Industry", *Journal of Political Economy*, 102(4), 653-683.

[9] Bresnahan, Timothy, and Peter Reiss (1991) "Entry and Competition in Concentrated Markets", *Journal of Political Economy*, 99(5), 977-1009.

[10] Brueckner, Jan, and Pablo Spiller (1994) "Economies of Traffic Density in the Deregulated Airline Industry", *Journal of Law and Economics*, 37(2), 379-415.

[11] Brueckner, Jan, Darin Lee, and Ethan Singer (2010) "Airline Competition and Domestic U.S. Airfares: A Comprehensive Reappraisal", *Working Paper*.

[12] Ciliberto, Federico, and Elie Tamer (2009) "Market Structure and Multiple Equilibria in the Airline Industry", *Econometrica*, 77(6), 1791-1828.

[13] Ciliberto, Federico and Jonathan Williams (*forthcoming)* "Limited Access to Airport Facilities and Market Power in the Airline Industry", *Journal of Law and Economics*.

[14] Einav, Liran, and Jonathan Levin (2010) "Empirical Industrial Organization: A Progress Report", *Journal of Economic Perspectives*, 24(2): 145-162.

[15] Evans, Williams, and Ioannis Kessides (1994) "Living by the 'Golden Rule': Multimarket Contact in the U.S. Airline Industry", *Quarterly Journal of Economics*, 109(2), 341-366.

[16] Fan, Jianqing, and Irène Gijbels (1996) "Local Polynomial Modelling and its Applications", *Monographs on Statistics and Applied Probability*, Chapman & Hall.

[17] Goolsbee, Austan, and Chad Syverson (2008) "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines", *Quarterly Journal of Economics*, 123(4), 1611-1633.

[18] Hahn, Jinyong, Petra Todd, and Wilbert Van Der Klaauw (2001) "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design", *Econometrica*, 69(1), 201-209.

[19] Imbens, Guido, and Thomas Lemieux (2008) "Regression Discontinuity Designs: A Guide to Practice", *Journal of Econometrics*, 142(2), 615-635.

[20] Lahiri, Soumendra (2003) "Resampling Methods for Dependent Data", *Springer Series in Statistics*, Springer.

[21] Lederman, Mara (2007) "Do Enhancements to Loyalty Programs Affect Demand? The Impact of International Frequent Flyer Partnerships on Domestic Airline Demand", *Rand Journal of Economics*, 38(4), 1134-1158.

[22] Lederman, Mara (2008) "Are Frequent Flyer Programs a Cause of the 'Hub Premium'", *Journal of Economics and Management Strategy*, 17(1), 35-66.

[23] Mazzeo, Michael (2002) "Product Choice and Oligopoly Market Structure", *Rand Journal of Economics*, 33(2): 221-242.

[24] McCrary, Justin (2008) "Testing for manipulation of the running variable in the regression discontinuity design", *Journal of Econometrics*, 142(2), 698-714.

[25] Nevo, Aviv, and Michael Whinston (2010) "Taking the Dogma Out of Econometrics: Structural Modeling and Credible Inference", *Journal of Economic Perspectives*, 24(2), 69-82.

[26] Pagan, Adrian, and Aman Ullah (1999) "Nonparametric econometrics", *Themes in Modern Econometrics*, Cambridge University Press.

[27] Seim, Katja (2006) "An Empirical Model of Firm Entry with Endogenous Product-Type Choices", *Rand Journal of Economics*, 37(3), 619-640.

[28] Transportation Research Board (1999) "Entry and competition in the U.S. airline industry: issues and opportunities", Washington D.C., Transportation Research Board.

[29] U.S. General Accounting Office (1989) "Barriers to competition in the airline industry", Washington D.C., U.S. General Accounting Office.

[30] U.S. General Accounting Office (2001) "Challenges in Enhancing Competition in Dominated Markets", Washington D.C., U.S. General Accounting Office.

[31] U.S. General Accounting Office (1990) "Airline competition: industry operating and marketing practices limit market entry", Washington D.C., U.S. General Accounting Office.

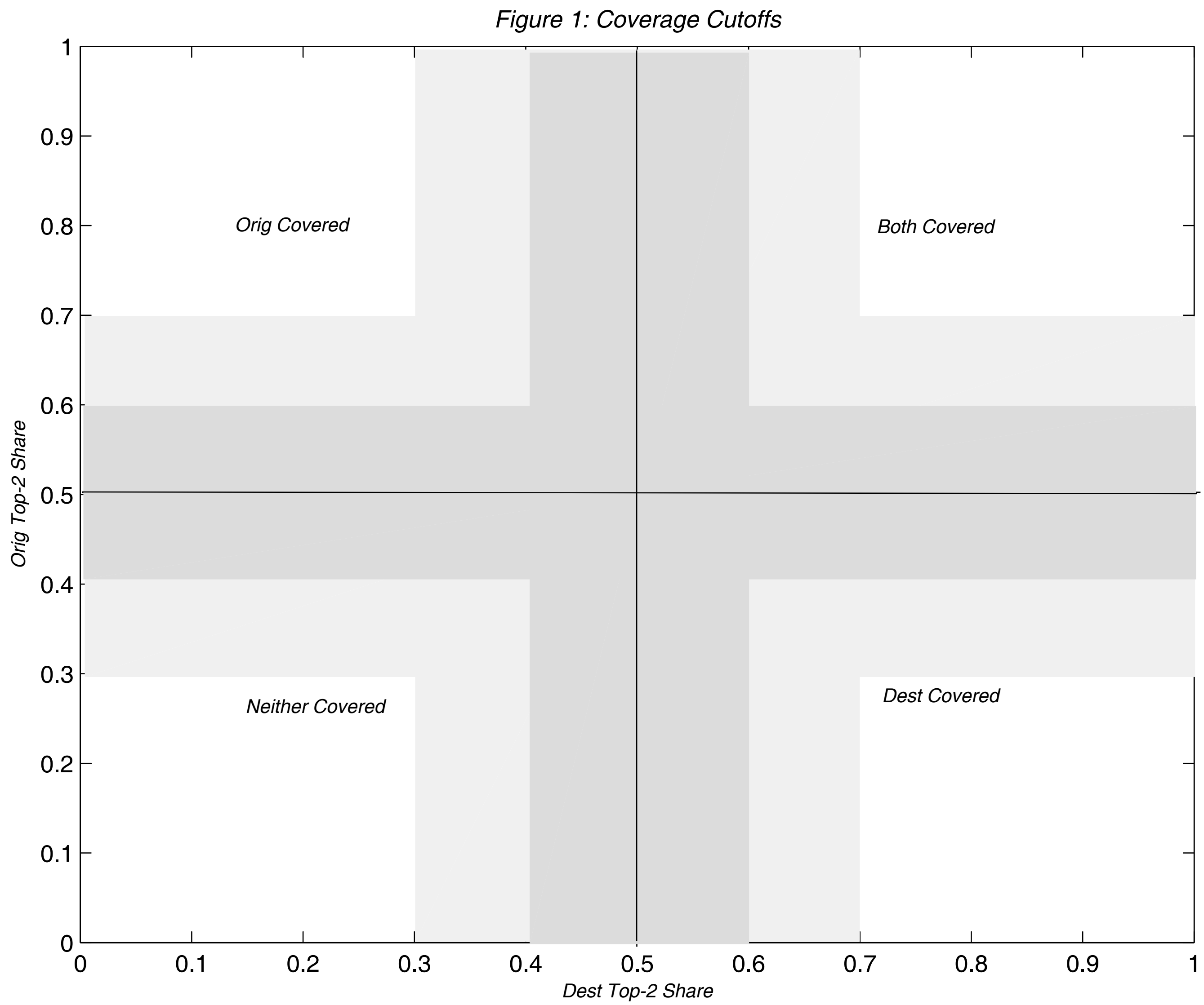


Figure 1: Coverage Cutoffs
Orig Covered
Both Covered
Neither Covered
Dest Covered
Orig Top-2 Share
Dest Top-2 Share
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 2: Density over Support of Predictors

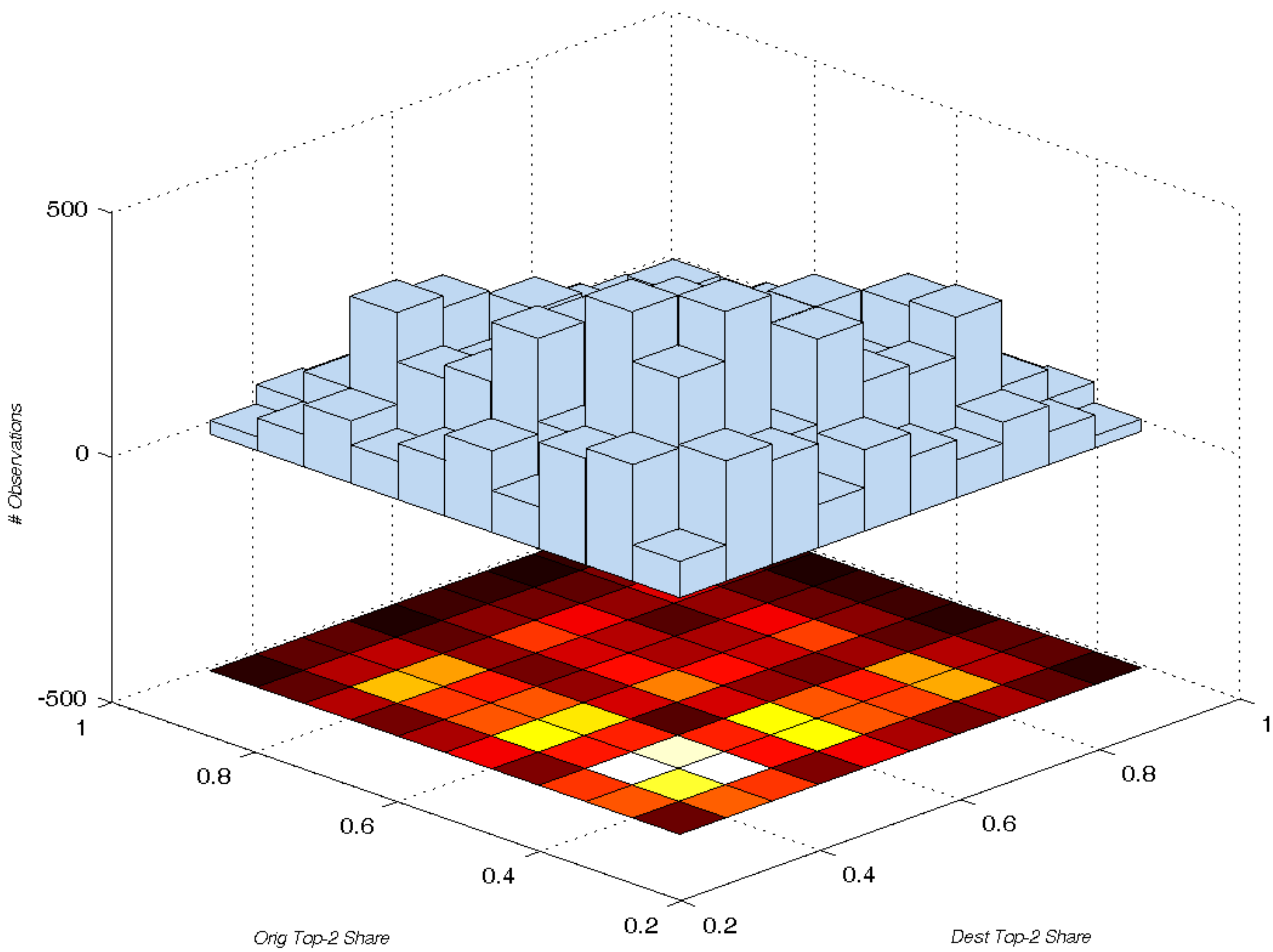


Figure 3: Change in Top-2 Share

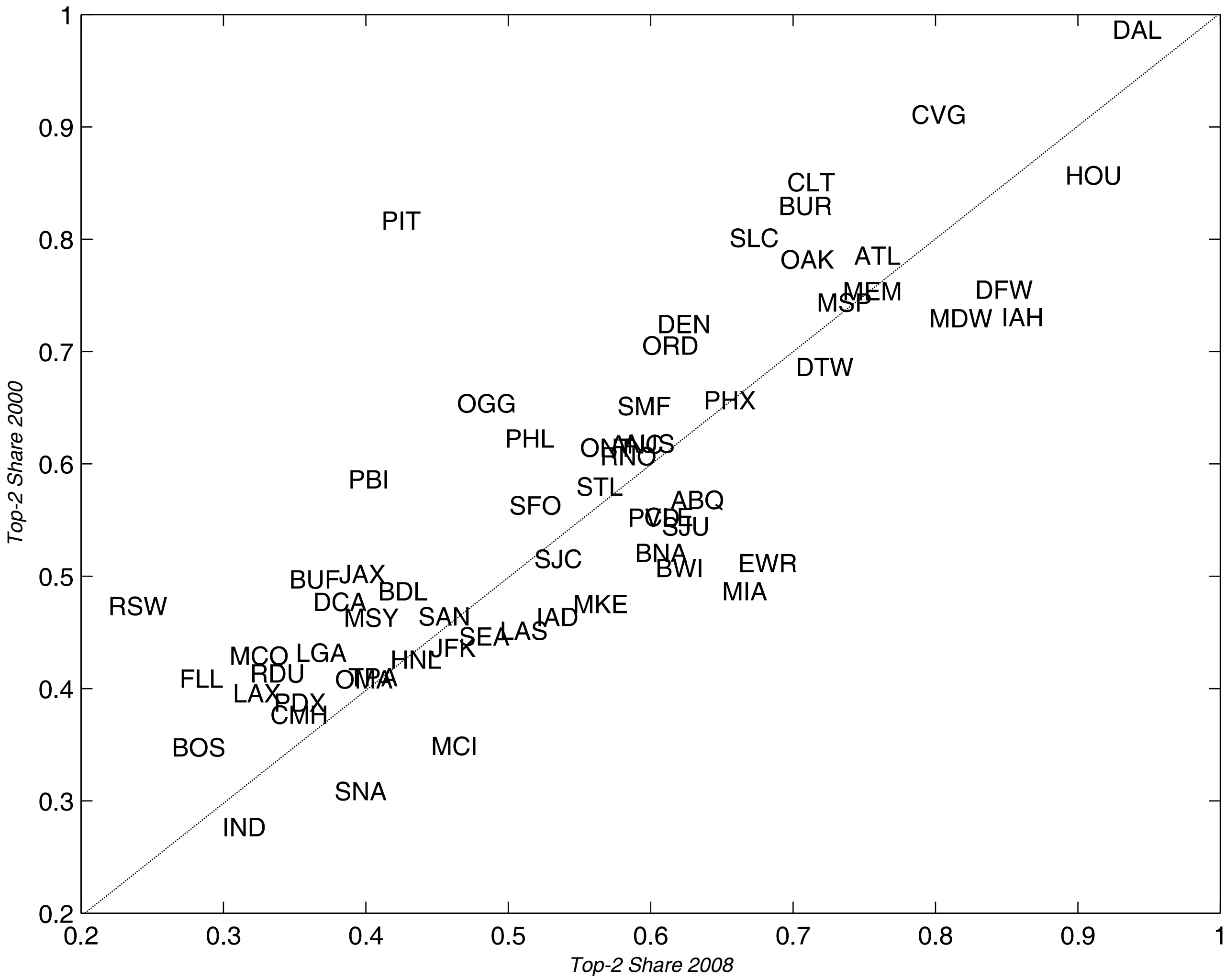


Figure 4: Local Coverage Effect on Fares

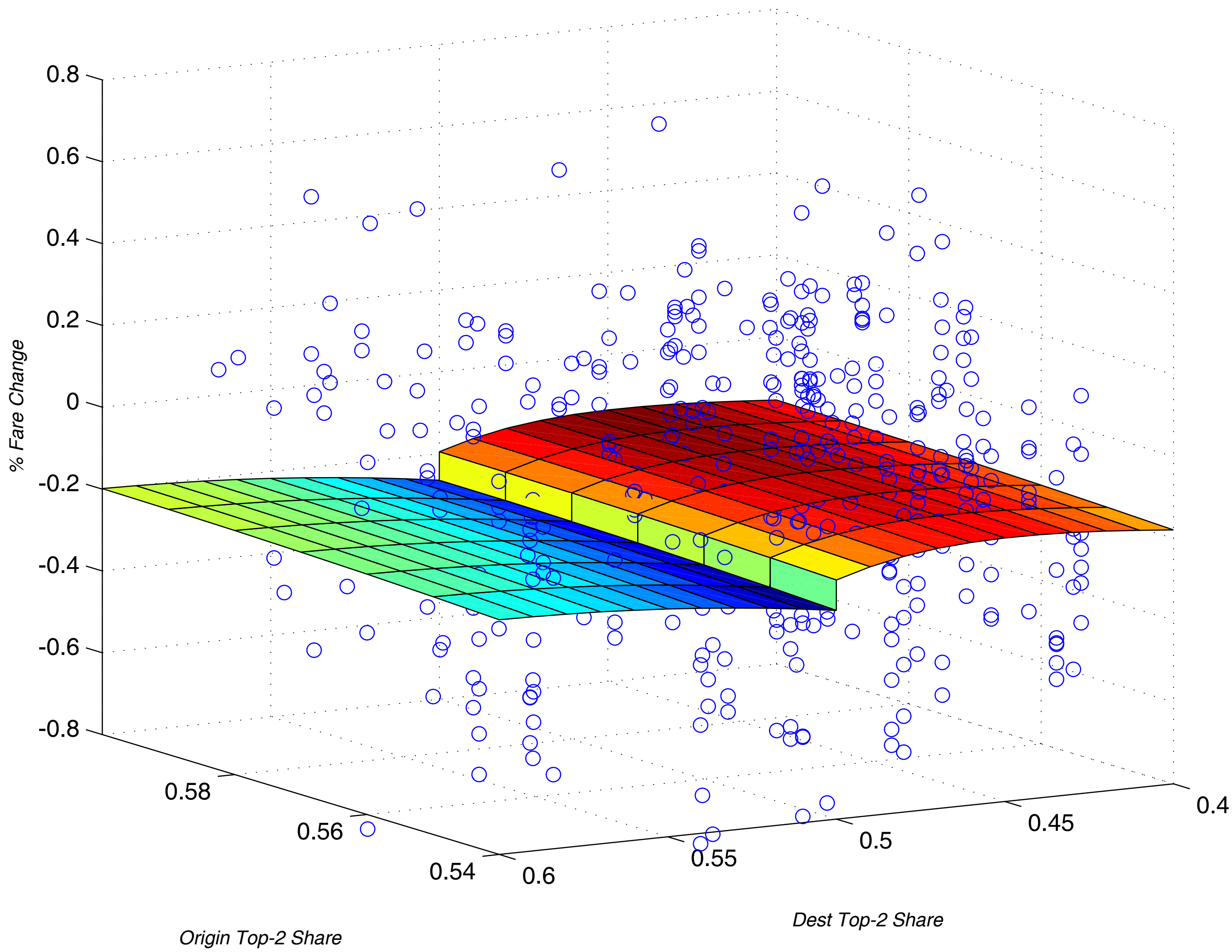


Figure 5: Expectation of Fares

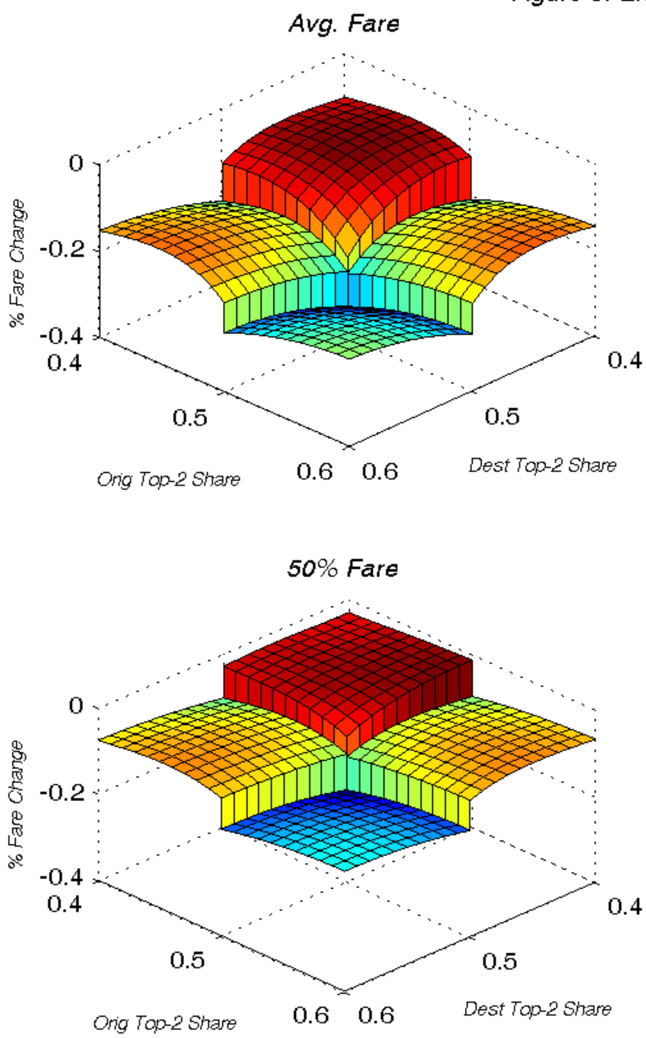


Figure 6: Lcc Presence

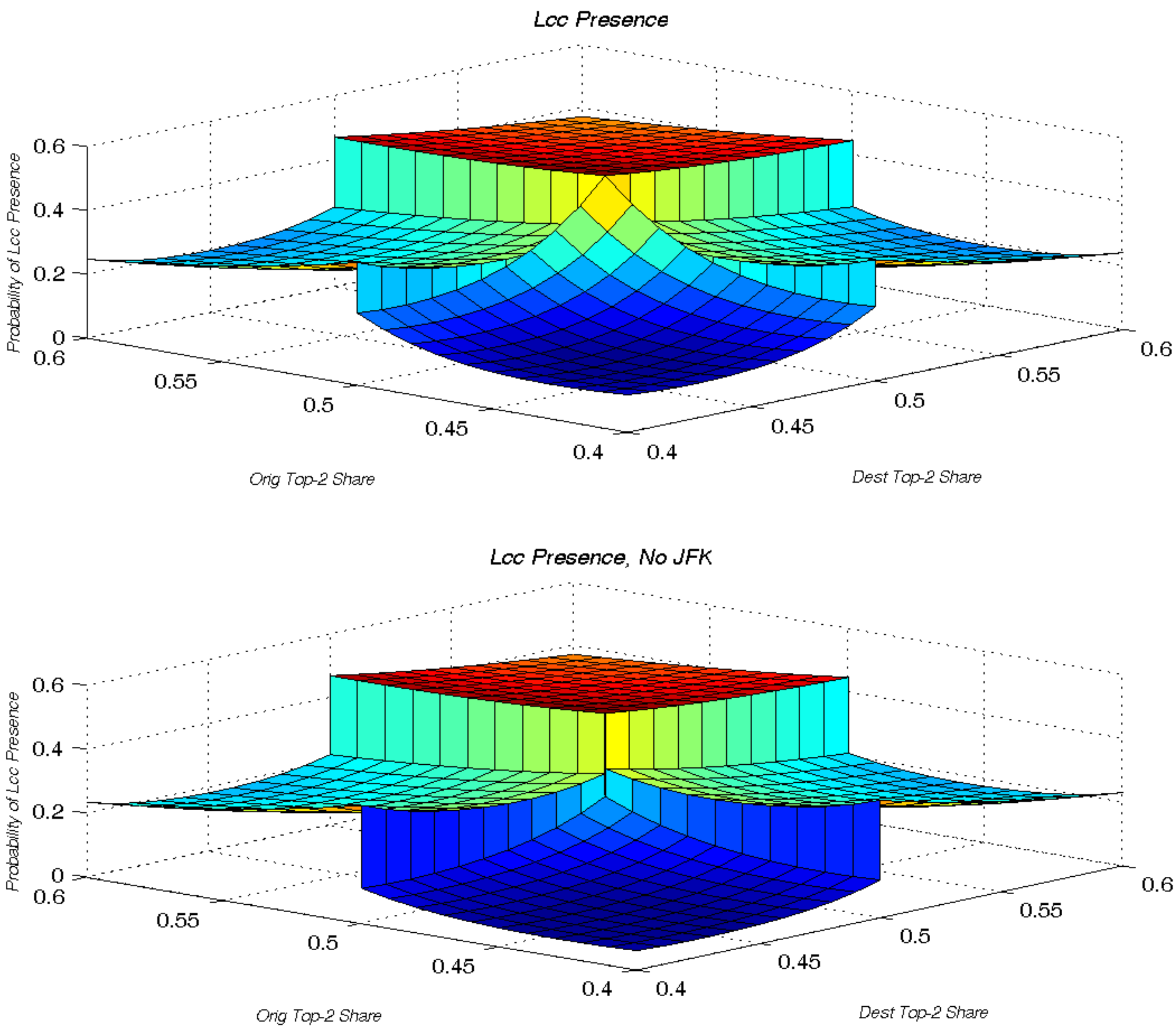
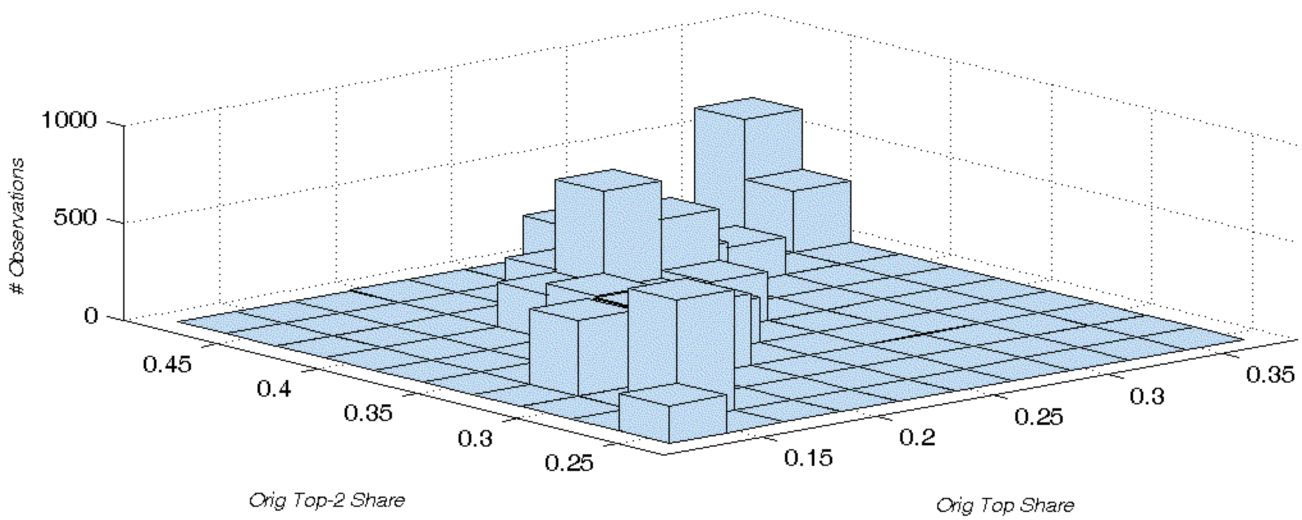


Figure 7: Test for Strategic Manipulation

Density of Observations



Expectation of Fares

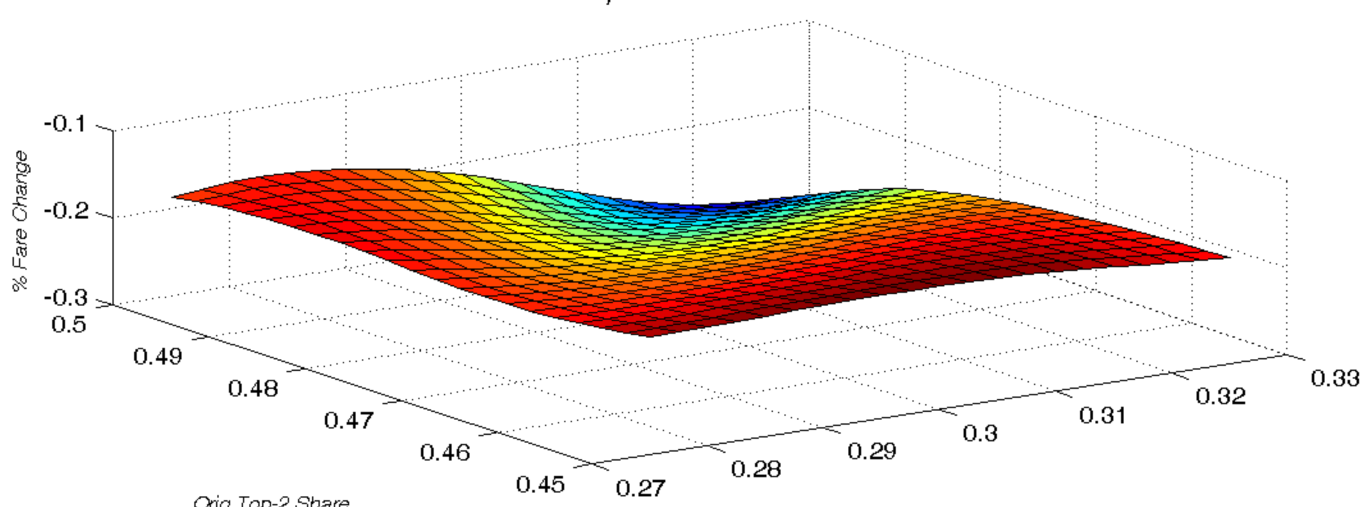


Table 1: Enplanements and Gates for Covered Airports

		Enplanements			Gates					
		US %	Top-2 %		Common %		Legacy %		Lcc %	
Airport	**Yr. Covered**	Mean	Mean	Max	2001	2008	2001	2008	2001	2008
ABQ	2000	0.45%	61.24%	63.97%	26.09%	31.82%	39.13%	63.64%	34.78%	4.55%
ANC	2000	0.36%	55.23%	61.74%	•	•	•	•	•	•
ATL	2000	5.85%	79.17%	82.18%	14.59%	15.08%	72.43%	73.37%	12.97%	11.56%
AUS	2000	0.51%	60.32%	61.80%	28.00%	16.00%	44.00%	52.00%	28.00%	32.00%
BNA	2000	0.64%	59.02%	63.03%	11.48%	9.84%	44.26%	44.26%	44.26%	45.90%
BUR	2000	0.36%	77.98%	83.54%	21.43%	7.14%	28.57%	35.71%	50.00%	57.14%
BWI	2000	1.42%	56.59%	65.95%	•	•	•	•	•	•
CLE	2000	0.84%	58.97%	61.29%	•	•	•	•	•	•
CLT	2000	1.82%	81.43%	86.84%	44.71%	48.35%	55.29%	51.65%	0.00%	0.00%
CVG	2000	1.52%	87.47%	92.87%	•	•	•	•	•	•
DAL	2000	0.48%	97.79%	99.82%	18.75%	0.00%	15.63%	25.00%	65.63%	75.00%
DCA	2001	1.09%	44.06%	50.10%	•	•	•	•	•	•
DEN	2000	2.82%	66.04%	72.44%	•	•	•	•	•	•
DFW	2000	4.06%	77.14%	85.12%	5.47%	17.42%	89.06%	80.00%	5.47%	2.58%
DTW	2000	2.47%	72.71%	76.32%	5.47%	5.08%	84.38%	88.14%	10.16%	6.78%
EWR	2000	2.41%	59.77%	69.90%	•	•	•	•	•	•
HOU	2000	0.61%	89.23%	92.19%	•	•	•	•	•	•
IAD	2001	1.42%	53.70%	59.91%	•	•	•	•	•	•
IAH	2000	2.50%	80.91%	86.12%	•	•	•	•	•	•
JAX	2000	0.37%	46.52%	50.19%	•	•	•	•	•	•
LAS	2005	2.68%	47.81%	52.40%	•	•	•	•	•	•
MDW	2000	1.11%	77.90%	90.37%	•	•	•	•	•	•
MEM	2000	0.80%	72.17%	77.10%	•	•	•	•	•	•
MIA	2001	2.29%	57.22%	68.95%	21.65%	32.04%	74.23%	64.08%	4.12%	3.88%
MKE	2001	0.47%	49.87%	56.49%	17.95%	0.00%	48.72%	21.28%	33.33%	78.72%
MSP	2000	2.44%	75.89%	78.75%	9.52%	8.66%	86.90%	90.55%	3.57%	0.79%
OAK	2000	0.88%	72.26%	78.52%	12.50%	37.93%	16.67%	3.45%	70.83%	58.62%
OGG	2000	0.42%	60.00%	68.59%	•	•	•	•	•	•
ONT	2000	0.48%	59.52%	61.44%	•	•	•	•	•	•
ORD	2000	5.01%	67.79%	74.12%	•	•	•	•	•	•
PBI	2000	0.46%	52.29%	58.64%	50.00%	53.13%	39.29%	34.38%	10.71%	12.50%
PHL	2000	1.92%	60.61%	65.66%	•	•	•	•	•	•
PHX	2000	2.72%	66.85%	68.95%	•	•	•	•	•	•
PIT	2000	1.22%	66.85%	81.65%	•	•	•	•	•	•
PVD	2000	0.39%	56.71%	63.55%	•	•	•	•	•	•
RNO	2000	0.38%	58.91%	62.61%	•	•	•	•	•	•
SAT	2001	0.48%	57.94%	100.00%	16.67%	17.39%	54.17%	52.17%	29.17%	30.43%
SDF	2000	0.27%	45.50%	51.64%	•	•	•	•	•	•
SFO	2000	2.44%	53.42%	56.29%	36.14%	37.04%	60.24%	59.26%	3.61%	3.70%
SJC	2000	0.82%	57.53%	64.18%	•	•	•	•	•	•
SJU	2000	0.74%	62.30%	69.04%	•	•	•	•	•	•
SLC	2000	1.42%	73.60%	80.12%	9.64%	8.43%	80.72%	81.93%	9.64%	9.64%
SMF	2000	0.65%	62.66%	65.90%	21.43%	38.46%	32.14%	19.23%	46.43%	42.31%
STL	2000	1.53%	69.14%	84.04%	4.55%	52.87%	79.55%	34.48%	15.91%	12.64%
Mean	2000.22727	1.45%	64.77%	71.46%	19.79%	22.98%	55.02%	51.29%	25.19%	25.72%

Table 2: Enplanements and Gates for Non-Covered Airports

	Enplanements			Gates					
	US %	Top-2 %		Common %		Legacy %		Lcc %	
Airport	Mean	Mean	Max	2001	2008	2001	2008	2001	2008
BDL	0.49%	44.91%	49.15%	•	•	•	•	•	•
BOS	1.87%	34.99%	38.93%	11.90%	10.31%	82.14%	68.04%	5.95%	21.65%
BUF	0.32%	37.59%	49.72%	6.25%	21.74%	78.13%	56.52%	15.63%	21.74%
CMH	0.49%	32.06%	37.66%	19.44%	22.22%	50.00%	58.33%	30.56%	19.44%
FLL	1.27%	35.05%	40.89%	•	•	•	•	•	•
HNL	1.51%	45.34%	48.07%	•	•	•	•	•	•
IND	0.57%	28.50%	32.86%	26.47%	30.00%	52.94%	57.50%	20.59%	12.50%
JFK	2.50%	41.42%	46.15%	•	•	•	•	•	•
LAX	4.22%	34.44%	40.44%	•	•	•	•	•	•
LGA	1.79%	41.34%	44.43%	•	•	•	•	•	•
MCI	0.81%	42.96%	47.56%	•	•	•	•	•	•
MCO	2.16%	36.66%	42.92%	•	•	•	•	•	•
MSY	0.67%	44.68%	47.59%	•	•	•	•	•	•
OKC	0.24%	41.82%	47.67%	0.00%	23.53%	68.75%	52.94%	31.25%	23.53%
OMA	0.28%	39.14%	41.72%	25.00%	35.00%	45.00%	45.00%	30.00%	20.00%
PDX	0.97%	37.15%	38.80%	19.57%	39.13%	32.61%	28.26%	47.83%	32.61%
RDU	0.63%	35.02%	41.34%	2.08%	19.05%	85.42%	66.67%	12.50%	14.29%
RSW	0.43%	39.12%	47.35%	23.53%	39.29%	58.82%	32.14%	17.65%	28.57%
SAN	1.15%	46.34%	47.80%	32.50%	22.50%	42.50%	43.75%	25.00%	33.75%
SEA	2.04%	45.12%	48.33%	21.62%	35.00%	40.54%	25.00%	37.84%	40.00%
SNA	0.62%	36.73%	39.64%	•	•	•	•	•	•
TPA	1.20%	40.40%	42.66%	18.37%	28.81%	63.27%	44.07%	18.37%	27.12%
Mean	1.19%	39.13%	43.71%	17.23%	27.21%	58.34%	48.19%	24.43%	24.60%

Table 3: Means for Covered and Non-Covered Markets

		Pre-AIR21		Post-AIR21			
Market-Carrier-Product	Covered	# Obs.	Mean	# Obs.	Mean	Diff.	Diff. in Diff.
Δ Avg. Fares	yes	11483	-55.650	11950	-55.362	0.288	-35.081
	no	2011	-67.060	2123	-31.691	35.369	
Δ 20^{th} Pct. Fare	yes	11483	-57.158	11950	-11.469	45.689	-9.796
	no	2011	-61.855	2123	-6.369	55.485	
Δ 50^{th} Pct. Fare	yes	11483	-70.439	11950	-26.472	43.967	-23.047
	no	2011	-76.803	2123	-9.790	67.013	
Δ 80^{th} Pct. Fare	yes	11483	-72.178	11950	-92.266	-20.088	-65.638
	no	2011	-97.468	2123	-51.918	45.550	
ΔFlight Distance	yes	11483	0.016	11950	0.007	-0.009	0.008
(Unit = 1000s of Miles)	no	2011	0.021	2123	0.003	-0.017	
Δ Fraction Routes	yes	11483	-0.010	11950	0.049	0.059	-0.001
	no	2011	-0.007	2123	0.052	0.059	
		Pre-AIR21		Post-AIR21			
Market-Carrier	Covered	# Obs.	Mean	# Obs.	Mean	Diff.	Diff. in Diff.
Δ % OnTime	yes	1873	-0.013	1731	0.044	0.057	0.063
	no	133	-0.003	150	-0.009	-0.006	
Δ Departures	yes	3171	66.560	3171	-57.739	-124.299	-85.177
	no	400	22.115	401	-17.007	-39.122	
		Pre-AIR21		Post-AIR21			
Market	Covered	# Obs.	Mean	# Obs.	Mean	Diff.	Diff. in Diff.
Δ Avg. Hub Premium	yes	1999	13.295	1990	-24.720	-38.014	-23.939
	no	253	4.374	262	-9.701	-14.075	
Δ 20th Pct. Hub Premium	yes	1999	1.747	1990	-12.139	-13.886	-13.530
	no	253	-3.125	262	-3.481	-0.356	
Δ 50th Pct. Hub Premium	yes	1999	9.151	1990	-15.428	-24.578	-30.775
	no	253	-4.300	262	1.897	6.197	
Δ 80th Pct. Hub Premium	yes	1999	28.667	1990	-42.235	-70.903	-54.690
	no	253	10.424	262	-5.788	-16.212	
Δ Lcc Penetration	yes	3171	0.319	3171	0.225	-0.095	0.231
	no	400	0.533	401	0.207	-0.326	
Δ Number Firms	yes	3171	0.942	3171	-0.321	-1.264	0.500
	no	400	1.250	401	-0.514	-1.764	
Δ % Nonstop	yes	3171	0.022	3171	0.034	0.013	-0.050
	no	400	-0.013	401	0.050	0.063	

Table 4: Avg. Fare Regressions

	All Markets		0.2 of Cutoff		0.1 of Cutoff	
	N=9,022		N=8,479		N=5,866	
Log(Avg. Fare)	(1)	(2)	(3)	(4)	(5)	(6)
1[1 cover]	-0.108***	-0.069***	-0.108***	-0.070***	-0.102***	-0.078***
	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)	(0.023)
1[2 cover]	-0.195***	-0.102***	-0.191***	-0.100***	-0.202***	-0.124***
	(0.025)	(0.024)	(0.025)	(0.025)	(0.030)	(0.030)
Nonstop	-0.124***	-0.112***	-0.122***	-0.110***	-0.102***	-0.090***
	(0.013)	(0.013)	(0.014)	(0.014)	(0.017)	(0.017)
Fraction Routes	0.448***	0.530***	0.477***	0.545***	0.488***	0.542***
	(0.036)	(0.035)	(0.038)	(0.036)	(0.045)	(0.043)
Flight Distance	0.239***	0.240***	0.238***	0.235***	0.204***	0.190***
	(0.045)	(0.046)	(0.046)	(0.047)	(0.051)	(0.052)
LccPresence	•	-0.174***	•	-0.173***	•	-0.143***
		(0.011)		(0.011)		(0.013)
NumberFirms	•	-0.001	•	0.001	•	0.001
		(0.004)		(0.005)		(0.005)
R^2	0.096	0.175	0.097	0.173	0.099	0.156
Borenstein-Rose (1994) Controls	no	yes	no	yes	no	yes

Notes

1) Additional controls include Population Origin, Population Dest, Per-Cap Income Origin, Per-Cap Income Dest.

Table 5: Fare Distribution Regressions

	All Markets		0.2 of Cutoff		0.1 of Cutoff	
	N=9,022		N=8,479		N=5,866	
Log(20% Fare)	(1)	(2)	(3)	(4)	(5)	(6)
1[1 cover]	-0.022	-0.004	-0.021	-0.004	-0.017	-0.006
	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)
1[2 cover]	-0.073***	-0.029	-0.082***	-0.038*	-0.041	-0.001
	(0.020)	(0.020)	(0.020)	(0.020)	(0.026)	(0.025)
LccPresence	•	-0.087***	•	-0.086***	•	-0.074***
		(0.009)		(0.010)		(0.012)
NumberFirms	•	-0.018***	•	-0.017***	•	-0.013***
		(0.004)		(0.004)		(0.005)
R^2	0.053	0.100	0.056	0.102	0.046	0.087
Log(50% Fare)						
1[1 cover]	-0.080***	-0.049**	-0.080***	-0.051**	-0.074***	-0.060**
	(0.025)	(0.024)	(0.025)	(0.024)	(0.025)	(0.024)
1[2 cover]	-0.161***	-0.085***	-0.158***	-0.085***	-0.132***	-0.069**
	(0.027)	(0.027)	(0.028)	(0.027)	(0.034)	(0.033)
LccPresence	•	-0.158***	•	-0.154***	•	-0.121***
		(0.012)		(0.013)		(0.016)
NumberFirms	•	-0.019***	•	-0.019***	•	-0.017***
		(0.005)		(0.005)		(0.006)
R^2	0.065	0.123	0.065	0.123	0.058	0.105
Log(80% Fare)						
1[1 cover]	-0.136***	-0.085***	-0.135***	-0.086***	-0.128***	-0.100***
	(0.032)	(0.031)	(0.032)	(0.031)	(0.032)	(0.031)
1[2 cover]	-0.245***	-0.123***	-0.231***	-0.111***	-0.244***	-0.141***
	(0.034)	(0.034)	(0.035)	(0.034)	(0.041)	(0.040)
LccPresence	•	-0.230***	•	-0.230***	•	-0.187***
		(0.015)		(0.016)		(0.019)
NumberFirms	•	-0.002	•	0.000	•	-0.005
		(0.006)		(0.006)		(0.007)
R^2	0.077	0.142	0.077	0.140	0.079	0.132
Borenstein-Rose (1994) Controls	no	yes	no	yes	no	yes

Notes

1) Additional controls include Population Origin, Population Dest, Per-Cap Income Origin, Per-Cap Income Dest.

Table 6: Hub Premium Regressions

	All Markets	0.2 of Cutoff	0.1 of Cutoff
	N=1,491	N=1,350	N=906
Hub Premium Avg. Fare	(1)	(2)	(3)
1[1 cover]	-0.137***	-0.139***	-0.126***
	(0.037)	(0.037)	(0.039)
1[2 cover]	-0.246***	-0.246***	-0.232***
	(0.038)	(0.039)	(0.047)
R^2	0.125	0.123	0.148
Hub Premium 20% Fare			
1[1 cover]	-0.106***	-0.105***	-0.095***
	(0.033)	(0.033)	(0.033)
1[2 cover]	-0.195***	-0.198***	-0.153***
	(0.034)	(0.035)	(0.042)
R^2	0.041	0.039	0.036
Hub Premium 50% Fare			
1[1 cover]	-0.190***	-0.187***	-0.178***
	(0.044)	(0.044)	(0.046)
1[2 cover]	-0.328***	-0.318***	-0.276***
	(0.047)	(0.047)	(0.058)
R^2	0.078	0.070	0.074
Hub Premium 80% Fare			
1[1 cover]	-0.145***	-0.147***	-0.118**
	(0.055)	(0.055)	(0.057)
1[2 cover]	-0.285***	-0.277***	-0.278***
	(0.056)	(0.057)	(0.070)
R^2	0.069	0.062	0.080

Notes

1) Additional controls include Nonstop, FractionRoutes, Flight Distance, Population Origin, Population Dest, Per-Cap Income Origin, Per-Cap Income Dest.

Table 7: Quality Regressions

	All Markets	0.2 of Cutoff	0.1 of Cutoff
% Nonstop	(1)	(2)	(3)
1[1 cover]	-0.014	-0.014	-0.012
	(0.010)	(0.010)	(0.010)
1[2 cover]	-0.015	-0.016	-0.016
	(0.010)	(0.010)	(0.012)
R^2	0.000	0.001	0.001
N	12,216	11,435	7,718
Log(Departures)			
1[1 cover]	0.328	0.321	0.420*
	(0.227)	(0.227)	(0.244)
1[2 cover]	0.341	0.340	0.344
	(0.229)	(0.232)	(0.262)
R^2	0.003	0.003	0.006
N	1,465	1,246	771
% OnTime			
1[1 cover]	0.167***	0.169***	0.166***
	(0.029)	(0.029)	(0.029)
1[2 cover]	0.197***	0.188***	0.188***
	(0.029)	(0.030)	(0.033)
R^2	0.052	0.051	0.057
N	1,267	1,066	675
Log(Number Routes)			
1[1 cover]	-0.034	-0.058	-0.092
	(0.075)	(0.080)	(0.104)
R^2	0.011	0.010	0.012
N	366	272	146

Notes

1) Additional controls include Population Origin, Population Dest, Per-Cap Income Origin, Per-Cap Income Dest.

Table 8: Competition Regressions

	All Markets	0.2 of Cutoff	0.1 of Cutoff
	N=2,979	N=2,747	N=1,826
Lcc Presence	(1)	(2)	(3)
1[1 cover]	0.142***	0.145***	0.096**
	(0.045)	(0.045)	(0.046)
1[2 cover]	0.395***	0.404***	0.428***
	(0.046)	(0.047)	(0.058)
R^2	0.033	0.036	0.038
Log(Number Firms)			
1[1 cover]	0.107	0.113	0.109
	(0.138)	(0.138)	(0.148)
1[2 cover]	0.043	0.102	0.190
	(0.143)	(0.144)	(0.179)
R^2	0.003	0.002	0.003

Notes

1) Additional controls include Population Origin, Population Dest, Per-Cap Income Origin, Per-Cap Income Dest.

Table 9: RDD Coverage Estimates, Fares

	$\tau(P^{orig}<0.5,P^{dest}=0.5)$					$\tau(P^{orig}=0.5,P^{dest}<0.5)$			
Predictor	Log(Avg.)	Log(20%)	Log(50%)	Log(80%)	Predictor	Log(Avg.)	Log(20%)	Log(50%)	Log(80%)
0.4	-0.1***	-0.051***	-0.084***	-0.112***	0.4	-0.088***	-0.046***	-0.073***	-0.109***
	(0.011)	(0.007)	(0.009)	(0.012)		(0.01)	(0.007)	(0.008)	(0.012)
0.42	-0.09***	-0.046***	-0.076***	-0.103***	0.42	-0.079***	-0.042***	-0.063***	-0.098***
	(0.01)	(0.007)	(0.008)	(0.011)		(0.009)	(0.006)	(0.008)	(0.011)
0.44	-0.077***	-0.041***	-0.067***	-0.092***	0.44	-0.067***	-0.038***	-0.054***	-0.088***
	(0.01)	(0.007)	(0.008)	(0.011)		(0.009)	(0.007)	(0.008)	(0.011)
0.46	-0.064***	-0.034***	-0.059***	-0.082***	0.46	-0.056***	-0.033***	-0.047***	-0.08***
	(0.01)	(0.008)	(0.009)	(0.012)		(0.01)	(0.007)	(0.009)	(0.012)
0.48	-0.051***	-0.025***	-0.052***	-0.074***	0.48	-0.045***	-0.027***	-0.043***	-0.073***
	(0.012)	(0.009)	(0.01)	(0.014)		(0.011)	(0.008)	(0.01)	(0.014)
0.5	-0.037***	-0.01	-0.047***	-0.065***	0.5	-0.032***	-0.016	-0.04***	-0.066***
	(0.015)	(0.011)	(0.012)	(0.019)		(0.015)	(0.011)	(0.013)	(0.019)
	$\tau(P^{orig}>0.5,P^{dest}=0.5)$					$\tau(P^{orig}=0.5,P^{dest}>0.5)$			
Predictor	Log(Avg.)	Log(20%)	Log(50%)	Log(80%)	Predictor	Log(Avg.)	Log(20%)	Log(50%)	Log(80%)
0.5	-0.081***	-0.048***	-0.083***	-0.054***	0.5	-0.077***	-0.053***	-0.077***	-0.055***
	(0.012)	(0.009)	(0.011)	(0.016)		(0.012)	(0.008)	(0.011)	(0.016)
0.52	-0.077***	-0.044***	-0.078***	-0.056***	0.52	-0.075***	-0.052***	-0.073***	-0.059***
	(0.012)	(0.008)	(0.01)	(0.015)		(0.012)	(0.008)	(0.01)	(0.015)
0.54	-0.074***	-0.042***	-0.074***	-0.057***	0.54	-0.073***	-0.051***	-0.071***	-0.062***
	(0.011)	(0.007)	(0.01)	(0.014)		(0.011)	(0.007)	(0.01)	(0.014)
0.56	-0.071***	-0.042***	-0.072***	-0.057***	0.56	-0.071***	-0.052***	-0.069***	-0.064***
	(0.011)	(0.007)	(0.01)	(0.014)		(0.011)	(0.007)	(0.01)	(0.013)
0.58	-0.07***	-0.042***	-0.071***	-0.058***	0.58	-0.071***	-0.053***	-0.069***	-0.066***
	(0.011)	(0.007)	(0.01)	(0.013)		(0.011)	(0.007)	(0.01)	(0.013)
0.6	-0.069***	-0.044***	-0.072***	-0.06***	0.6	-0.071***	-0.056***	-0.07***	-0.069***
	(0.011)	(0.007)	(0.009)	(0.013)		(0.011)	(0.007)	(0.009)	(0.013)

Table 10: RDD Coverage Estimates, Lcc Presence

	$\tau(P^{orig}<0.5, P^{dest}=0.5)$			$\tau(P^{orig}=0.5, P^{dest}<0.5)$	
Predictor	JFK	No-JFK	Predictor	JFK	No-JFK
0.4	0.158***	0.26***	0.4	0.147***	0.249***
	(0.021)	(0.018)		(0.02)	(0.016)
0.42	0.115***	0.213***	0.42	0.105***	0.206***
	(0.019)	(0.015)		(0.018)	(0.014)
0.44	0.074***	0.177***	0.44	0.067***	0.173***
	(0.02)	(0.014)		(0.02)	(0.014)
0.46	0.023	0.145***	0.46	0.02	0.144***
	(0.022)	(0.015)		(0.023)	(0.017)
0.48	-0.049	0.111***	0.48	-0.047	0.117***
	(0.028)	(0.02)		(0.03)	(0.022)
0.5	-0.153***	0.072***	0.5	-0.143***	0.087***
	(0.039)	(0.028)		(0.042)	(0.032)
	$\tau(P^{orig}>0.5, P^{dest}=0.5)$			$\tau(P^{orig}=0.5, P^{dest}>0.5)$	
Predictor	JFK	No-JFK	Predictor	JFK	No-JFK
0.5	0.144***	0.173***	0.5	0.153***	0.188***
	(0.022)	(0.024)		(0.02)	(0.021)
0.52	0.159***	0.19***	0.52	0.168***	0.202***
	(0.021)	(0.022)		(0.018)	(0.019)
0.54	0.173***	0.204***	0.54	0.182***	0.214***
	(0.02)	(0.02)		(0.017)	(0.018)
0.56	0.185***	0.217***	0.56	0.195***	0.226***
	(0.019)	(0.02)		(0.017)	(0.017)
0.58	0.197***	0.228***	0.58	0.207***	0.236***
	(0.019)	(0.019)		(0.017)	(0.017)
0.6	0.207***	0.237***	0.6	0.219***	0.247***
	(0.019)	(0.019)		(0.017)	(0.017)