

Workplace Concentration of Immigrants: Web appendix[†]

Fredrik Andersson, Mónica García-Pérez, John Haltiwanger,
Kristin McCue, and Seth Sanders*

October 29, 2010

This web appendix includes supplementary tables, figures and analysis. It is organized into four sections. Section A contains tables of summary statistics that compare the full sample (all workers from the UI wage records in our sample of MSAs) to the matched sample (workers with records matched to the Decennial long form data), along with some summary statistics on differences between recent and more established immigrants. Section B includes supplementary tables on the contribution of covariates to immigrant concentration. Section C has some supplementary tables for the country of origin analyses. Section D includes an analysis of a statistical artifact that arises in examining concentration by employer size for very small firms (D.1), and some figures with additional detail on differences by firm size (D.2).

A Summary Statistics

Table W-1 reports variation in immigrant share across sample MSAs using the full sample of UI wage records for the 11 states.

The next set of tables provides some information on the representativeness of our matched sample. Some of the tables also include information on mean differences between recent and established immigrants, where the split is based on the year in which an immigrant first applied for a Social Security Number (SSN), which we use as a proxy

[†] Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau, the Comptroller of the Currency, or the U.S. Department of the Treasury. All results have been reviewed to ensure that no confidential information is disclosed.

* Credit Risk Analysis Division, Office of the Comptroller of the Currency; Saint Cloud State University; University of Maryland; Census Bureau; and Duke University, respectively.

WEB APPENDIX

for the date of their arrival. While this does not precisely identify the actual year of arrival, a comparison to reported year of arrival in the long-form census (where available) suggests that this provides a reasonable proxy. Comparing our recent/established classification based on date of SSN application to one based on reported year of arrival from the 2000 census, 92% of immigrants are classified in the same way according to both sources. The two measures disagree most often for Mexican immigrants: 4% report arriving in the country between 1995 and 2000 while 10% applied for an SSN in that window. The differences between recent arrivals and other immigrants found in these tables confound the effects of time in the U.S. with changes in immigrant characteristics across entering cohorts. For this reason, we include these results for background purposes only—a complete analysis would require developing a panel version of our database.

These additional tables and figure provide the following information:

- Comparisons between the full and matched samples on an unweighted basis for all immigrants are in Table W-2 and for all natives are in Table W-3.
- A comparison of Table W-4 to Table W-5 illustrates how closely the weighted matched sample lines up with the full sample. (The last column of Table W-5 giving native means is identical to the last column of Table 1 in the main text. The immigrant column in Table 1 is a weighted average of the Recent/Established columns in Table W-5.)
- Mean characteristics for the full and unweighted matched samples with a split between recent and established immigrants can be made by comparing Tables W-4 and W-6. Table W-7 provides means for that split for the additional variables that are available only for the matched sample. (The native means in this table match those for natives in Table 1 in the main text.)

WEB APPENDIX

- Figure W-1 shows the cumulative distribution of coworker share for recent/established immigrants and for natives. Comparison to Figure 1 in the main paper shows that patterns in the full sample distribution are similar to those in the weighted distribution for the matched sample.

Table W-1: Variation in Immigrant Share of Workforce across Sample MSAs

	Percent Immigrant		
	Total	Recent	Established
Mean	18.86	3.40	15.46
Standard Deviation	10.27	1.85	8.57
P25	10.57	1.94	8.52
Median	16.26	2.92	13.54
P75	26.60	4.37	22.82
P90	32.58	6.03	27.23

Source: Authors calculations based on LEHD UI-ES202 database.

Note: Unit of observation is an MSA. Immigrant shares are measured as of the second quarter of 2000, and recent immigrants are those arriving between 1995 and 2000. The table presents fuzzed percentiles values.

WEB APPENDIX

Table W-2: Characteristics of Immigrants in Full and Matched Samples (Un-weighted)

		Full	Matched
Coworker share		37.7	36.3
Worker age	Age<30	23.3	22.3
	30< Age <40	33.7	33.3
	Age>40	42.9	44.4
Male		56.1	55.0
Age at arrival	<= 12	12.4	12.5
	13-25	47.5	47.6
	26-35	26.9	27.1
	36+	13.2	12.8
Education	High school drop-out		31.8
	High school graduate		18.4
	Some college		17.1
	Bachelor's degree		22.2
	Advanced degree		10.5
Does not speak English well			20.4
Log quarterly earnings on primary job		8.5	8.5
Continuity of 2000-Q2 job	Q1 and Q3	69.3	71.1
	Q1 or Q3	24.6	23.3
	Neither Q1 nor Q3	6.0	5.6
Establishment size	2-9 employees	9.3	8.7
	10-49	23.0	21.9
	50-99	13.3	12.9
	100-499	31.0	30.8
	500 or more	23.4	25.6
Firm has multiple establishments		33.5	35.8
Establishment age	<=1 year	12.4	11.4
	2-4 years	23.5	22.6
	5+ years	64.1	66.1
Sector	Construction	5.2	5.2
	Manufacturing	20.8	21.3
	Transportation/utilities	4.0	3.6
	Wholesale	7.2	6.6
	Retail	19.2	19.3
	FIRE	5.0	4.9
	Services	38.5	39.1
Immigrant share of workers in residence tract		36.7	35.9
Neighborhood network index		1.8	1.9
Shared commute index		0.3	0.3

Notes: The unit of observation is a worker. N=600,761 for the matched sample and N=6.2 million for the full sample. All figures except log earnings represent percentages.

WEB APPENDIX

Table W-3: Characteristics of Natives in Full and Matched Samples (Unweighted)

		Full	Matched
Coworker share		14.5	13.6
Worker age	Age<30	32.3	30.9
	30< Age <40	26.8	25.9
	Age>40	40.8	43.1
Male		51.3	50.5
Education	High school drop-out		15.8
	High school graduate		25.5
	Some college		25.7
	Bachelor's degree		24.7
	Advanced degree		8.2
Does not speak English well			0.8
Log quarterly earnings on primary job		8.4	8.4
Continuity of 2000-Q2 job	Q1 and Q3	65.6	67.3
	Q1 or Q3	26.6	25.6
	Neither Q1 nor Q3	7.8	7.2
Establishment size	2-9 employees	8.3	7.8
	10-49	23.8	22.7
	50-99	13.5	13.1
	100-499	29.6	29.6
	500 or more	24.9	26.8
Firm has multiple establishments		42.2	44.7
Establishment age	<=1 year	12.0	11.1
	2-4 years	24.8	23.9
	5+ years	63.3	65.1
Sector	Construction	5.7	5.7
	Manufacturing	12.1	13.3
	Transportation/utilities	5.4	5.0
	Wholesale	6.6	6.2
	Retail	22.0	22.3
	FIRE	7.1	6.7
	Services	41.2	40.8
Immigrant share of workers in residence tract		15.7	14.0
Neighborhood network index		1.7	1.9
Shared commute index		0.5	0.5

Notes: The unit of observation is a worker. N=3.0 million for the matched sample and N=26.4 million for the full sample. All figures except log earnings represent percentages

WEB APPENDIX

Table W-4: Characteristics of Immigrant and Native Workers, Full Sample

	Immigrants		
	Recent	Established	Native
Coworker share	42.1	36.8	14.5
Age			
Age<30	42.8	19.6	32.3
30<Age<40	36.2	33.3	26.8
Age>40	21.0	47.1	40.8
Male	56.3	56.1	51.3
Age at arrival (*)			
<= 12	1.0	14.5	
13-25	36.5	49.6	
26-35	36.9	25.0	
36+	25.6	10.9	
Establishment size			
2-9 employees	8.7	9.4	8.3
10-49	23.8	22.8	23.8
50-99	14.2	13.2	13.5
100-499	31.8	30.9	29.6
500 or more	21.5	23.8	24.9
Firm has multiple establishments	32.0	33.8	42.2
Establishment age			
0-1	13.6	12.1	12.0
2-4	26.4	23.0	24.8
Age 5 or more	60.0	64.9	63.3
Sector			
Construction	4.5	5.4	5.7
Manufacturing	19.0	21.2	12.1
Transportation & utilities	3.1	4.2	5.4
Wholesale	7.1	7.2	6.6
Retail	23.2	18.5	22.0
FIRE	3.2	5.3	7.1
Services	39.8	38.2	41.2
Log quarterly earning on primary job	8.2	8.5	8.4
Consecutive quarters on 2000-Q2 job			
Quarter before AND after	59.7	71.2	65.6
Quarter before OR after (not both)	32.1	23.2	26.6
Neither quarter before NOR after	8.2	5.6	7.8
Immigrant share of workers in residence tract	38.2	36.5	15.7
Neighborhood network index	2.0	1.8	1.7
Shared commute index	0.3	0.3	0.5

Notes: (*) Year of application for a SSN is used as a proxy for time of arrival in the U.S. Based on authors' calculations from LEHD database. The unit of observation is a worker (N=35,966,1450). All figures except log earnings represent percentages.

WEB APPENDIX

Table W-5: Characteristics of Weighted Matched Sample

	Immigrants		
	Recent	Established	Native
Coworker share	41.0	36.4	14.3
Age			
Age<30	43.7	19.9	33.0
30<Age<40	36.4	33.5	26.1
Age>40	19.9	46.6	40.9
Male	55.5	55.6	51.2
Age at arrival (*)			
<= 12	1.2	14.8	
13-25	37.0	49.5	
26-35	37.2	25.2	
36+	24.6	10.5	
Establishment size			
2-9 employees	8.2	9.0	7.8
10-49	23.5	22.3	23.2
50-99	14.3	12.9	13.4
100-499	31.5	30.5	29.5
500 or more	22.5	25.3	26.1
Firm has multiple establishments	33.5	34.8	43.3
Establishment age			
0-1	13.0	11.6	11.5
2-4	25.9	22.3	24.4
Age 5 or more	61.1	66.2	64.1
Sector			
Construction	5.1	5.6	6.0
Manufacturing	18.0	20.7	12.4
Transportation & utilities	2.7	3.7	4.9
Wholesale	6.3	6.6	6.1
Retail	24.2	19.1	23.1
FIRE	3.1	5.0	6.5
Services	40.6	39.3	41.1
Log quarterly earning on primary job	8.1	8.5	8.3
Consecutive quarters on 2000-Q2 job			
Quarter before AND after	58.2	70.5	64.4
Quarter before OR after (not both)	32.4	23.4	27.1
Neither quarter before NOR after	9.4	6.1	8.4
Immigrant share of workers in residence tract	38.2	36.4	14.8
Neighborhood network index	2.0	1.8	1.9
Shared commute index	0.3	0.3	0.5

Notes: (*) Year of application for a SSN is used as a proxy for time of arrival in the U.S. Based on authors' calculations from LEHD database. The unit of observation is a worker (N=3,549,111). All figures except log earnings represent percentages.

WEB APPENDIX

Table W-6: Characteristics of Matched Sample Workers (Unweighted)

	Immigrants		
	Recent	Established	Native
Coworker share	40.2	35.7	13.6
Age			
Age<30	42.5	18.8	30.9
30<Age<40	36.6	32.7	25.9
Age>40	20.9	48.5	43.1
Male	55.0	55.0	50.5
Age at arrival (*)			
<= 12	1.1	14.5	
13-25	36.0	49.6	
26-35	37.3	25.4	
36+	25.6	10.6	
Establishment size			
2-9 employees	8.0	8.9	7.8
10-49	22.8	21.8	22.7
50-99	14.0	12.7	13.1
100-499	31.9	30.6	29.6
500-high	23.4	26.0	26.8
Firm has multiple establishments	34.7	36.0	44.7
Establishment age			
0-1	12.4	11.2	11.1
2-4	25.7	22.0	23.9
Age 5 or more	61.9	66.8	65.1
Sector			
Construction	4.8	5.3	5.7
Manufacturing	19.2	21.7	13.3
Transportation & utilities	2.7	3.7	5.0
Wholesale	6.5	6.6	6.2
Retail	23.5	18.6	22.3
FIRE	3.2	5.2	6.7
Services	40.3	38.9	40.8
Log quarterly earning on primary job	8.2	8.5	8.4
Consecutive quarters on 2000-Q2 job			
Quarter before AND after	61.2	72.8	67.3
Quarter before OR after (not both)	30.8	22.0	25.6
Neither quarter before NOR after	8.0	5.2	7.2
Immigrant share of workers in residence tract	37.3	35.7	14.0
Neighborhood network index	2.0	1.8	1.9
Shared commute index	0.3	0.3	0.5

Notes: (*) Year of application for a SSN is used as a proxy for time of arrival in the U.S. Based on authors' calculations from LEHD database. The unit of observation is a worker (N=3,549,111). All figures except log earnings represent percentages.

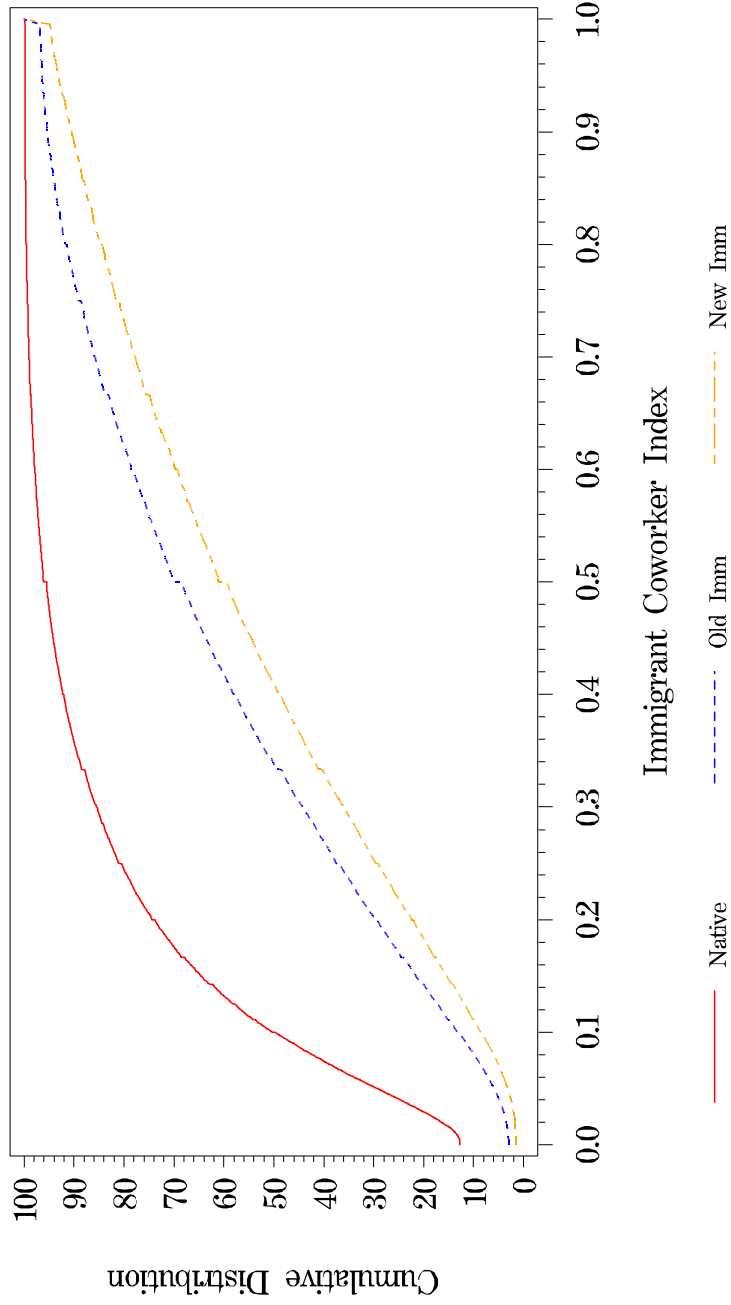
WEB APPENDIX

Table W-7: Means for Variables Available from Match (Weighted)

	Immigrants		
	Recent	Established	Native
Education categories			
High school drop-out	34.3	32.3	17.0
High school graduate	19.6	18.4	25.3
Some college	14.2	17.7	25.8
Bachelor's degree	20.4	21.9	24.1
Advanced degree	11.4	9.7	7.8
Does not speak English well	30.3	19.2	0.9

Notes: Year of application for a SSN is used as a proxy for time of arrival in the U.S. Estimates use weights based on propensity score model. Based on authors' calculations from LEHD database. The unit of observation is a worker (N=3,549,111).

Figure W-1: Cumulative Distribution of Coworker Share by Worker Type



Source= LEHD database, year 2000 second quarter

B Supplementary Analysis of Decompositions of Immigrant Concentration

Tables W-8 and W-9 report concentration estimates and R-squared values from the main effects model for recent and established immigrants, and for both the full and matched samples. Rows 1-3 in Table W-9 correspond to the 1st, 2nd, and last rows of Table W-8. A comparison of these rows across the two tables shows that the results are quite similar for the full and matched samples. In Table W-8, controls for industry and residential segregation have relatively large effects on both the measures of concentration and the R-squared values. Table W-9 adds the education and English language variables available through the match to decennial data to the set of controls, showing that English language skills are also important. Note that these are also the variables identified as having important effects using the Gelbach decomposition in Table 3.

Table W-10 reports results for the Gelbach decomposition applied to the main effects model. These results are much like those in Table 3 in the main paper but are broken out by recent and established immigrants. The last column of Table W-10 will not exactly match the results in the main paper since for this supplementary analysis we estimated the models using a 1/5 subsample to ease computation of the decomposition.

Tables W-11 and W-12 present supplementary versions of Table 7 in the main paper. Table W-11 gives the decomposition of the fully interacted model evaluated at immigrants means. The contribution of coefficients is close to but not exactly zero because of our slightly asymmetric treatment of MSA interactions. These results are based on a 1/5 random subsample of our data to speed up estimation. For purposes of comparison, that table also includes estimates evaluated at native means. These differ from those in Table 7 only because of the subsampling.

As with the Oaxaca decomposition, there are two versions: the differences in means

WEB APPENDIX

can be multiplied by the immigrant or native coefficients, with the differences in coefficients multiplied by the means for the other group. Table 7 uses the native coefficients for the first term and immigrant means for the second (as does Table W-10) which fits more readily with our set up. Table W-12 presents the other version, which corresponds to a model where immigrants are the omitted category and the coefficient on the native dummy gives a measure of immigrant concentration.

WEB APPENDIX

Table W-8: Contribution of Covariates to Immigrant Concentration (Full Sample)

Covariates	Recent immigrant	Established immigrant	R-square
No covariates	0.273	0.221	0.209
MSA dummies	0.224	0.163	0.387
MSA + following controls (1 at a time)			
Worker age	0.225	0.162	0.388
Worker sex	0.224	0.162	0.387
Log earnings and full-quarter controls	0.224	0.162	0.387
Employer size	0.224	0.162	0.389
Employer age	0.224	0.163	0.387
Employer age * Multi-unit	0.220	0.160	0.395
Industry detail	0.192	0.136	0.469
Size and industry	0.192	0.136	0.470
Neighborhood network index	0.224	0.162	0.388
Shared commute index	0.224	0.162	0.388
Immigrant share in residential tract	0.183	0.128	0.418
All of the above	0.155	0.103	0.501

Notes: Figures in the first two columns give the predicted difference in mean coworker share between the immigrant group and natives. As a point of reference, the mean coworker share for natives in the first line is .145 (as in the first column of Table W-3). It is also .145 for all the other specifications if evaluated at the native mean for all included covariates, but somewhat higher if evaluated at the pooled sample mean. The unit of observation is a worker. N=35,966,450 for the full sample. The variables are as described in Table W-3, except that we use 185 detailed industry categories in place of sector, and use more detailed size categories for establishments with fewer than 50 employees.

WEB APPENDIX

Table W-9: Contribution of Covariates to Immigrant Concentration in Matched Sample

Covariates	Recent immigrant	Established immigrant	R-square
Matched sample			
No covariates	0.269	0.214	0.201
MSA dummies	0.219	0.156	0.382
Full sample specification	0.151	0.097	0.499
Full sample specification +:			
Education controls	0.148	0.095	0.501
English language control	0.131	0.087	0.506
Education and English controls	0.131	0.086	0.507

Notes: Figures in the first two columns give the predicted difference in mean coworker share between the immigrant group and natives. As a point of reference, the mean coworker share for natives in the first line is .136 (as in column2 of Table W-3). The unit of observation is a worker. N=3,549,111. The variables are as described in Table W-3, except that we use 185 detailed industry categories in place of sector, and use more detailed size categories for establishments with fewer than 50 employees.

WEB APPENDIX

Table W-10: Decomposition of Contribution of Covariates to Immigrant Concentration in Matched Sample

Covariates	Immigrant group		
	Recent	Established	All
Mean immigrant-native difference in model with:			
1. No covariates	0.268	0.220	0.228
2. MSA dummies	0.214	0.160	0.170
3. Full set of controls	0.128	0.091	0.097
Contribution to reduction in coefficient between rows 2. and 3.			
		Percents	
Individual characteristics (total)	26.1	22.8	23.9
Log earnings	-0.4	0.4	0.2
Quarters of work	0.0	0.0	0.0
Age and sex	-0.9	0.7	0.2
Language	24.5	18.6	20.5
Education	3.0	3.0	2.9
Employer characteristics (total)	35.4	36.5	36.0
Firm size	0.0	0.4	0.3
Firm age and multi-unit status (interacted)	3.2	3.5	3.4
Industry	32.2	32.6	32.3
Sector	13.1	17.1	16.1
Detail	19.1	15.6	16.2
Manufacturing detail (73 3-digit industries)	2.1	-1.1	-0.5
Transportation, commun, utilities (14 inds)	0.1	0.2	0.2
Wholesale (18 industries)	1.0	1.0	1.0
Retail (33 industries)	2.7	3.6	3.4
FIRE (4 industries)	-0.4	0.0	-0.1
Services (51 industries)	13.5	11.8	12.1
Neighborhood characteristics (total)	38.5	40.7	40.1
Immigrant share in residential tract	37.9	40.1	39.4
Network index	0.2	0.2	0.2
Shared commute index	0.4	0.5	0.5

Notes: Figures in the first three rows give the predicted difference in mean coworker share between the immigrant group and natives. Estimates for recent and established immigrants are based on pooled regressions with only main effects identifying the two groups. Estimates in this table are based on a 1/5 subsample of our full matched sample to ease computation of the decomposition. The rows in the bottom panel of the table give the percentage of the difference in coefficients between rows 2 and 3 accounted for by that particular set of controls.

Table W-11: Decomposition of difference between raw and conditional concentration measures

	Evaluated at	
	Immigrant means	Native means
Mean immigrant-native difference in model with:		
1. MSA dummies	0.170	0.170
2. Full set of controls with interactions	0.115	0.070
Contribution to reduction in concentration between rows 1. and 2. (percents)		
	Xs	Differences in Coefficients
MSA (main effects in base, but interactions not)		
Individual characteristics (total)	19.2	-3.4
Log earnings	0.6	-0.1
Quarters of work	-0.3	0.0
Age and sex	-0.7	-0.1
Language	14.8	0.0
Schooling	4.9	-0.1
Employer characteristics (total)	43.0	8.1
Firm size	0.4	2.7
Firm age and multi-unit status (interacted)	2.7	-1.2
Industry (total)	40.0	-0.7
Sector (categories below, + construction)	18.6	-0.3
Detail (modal ind excluded in each sector)	21.4	-0.2
Manufacturing detail (73 3-digit indus)	0.9	0.1
Transportation, commun, utilities (14 3-dig)	0.3	0.5
Wholesale (18 3-digit industries)	1.1	0.2
Retail (33 3-digit industries)	3.8	0.6
FIRE (4 3-digit industries)	-0.1	2.1
Services (51 3-digit industries)	15.4	-0.1
Neighborhood characteristics (total)	43.8	-0.1
Percent immigrant in residential tract	43.9	8.4
Network index	-0.4	-1.4
Shared commute index	0.3	-1.6
		24.0
		-0.2
		0.2
Total column share	106.1	-6.1
		58.0
		42.0

Notes: Regressions use controls deviated from the indicated mean so the coefficient on the immigrant indicator gives the predicted difference in coworker shares at that mean. The differences in Xs are evaluated using the coefficients for natives, while differences in coefficients are evaluated using immigrant Xs deviated from the indicated mean.

Table W-12: Alternative decomposition of difference between raw and conditional concentration

	Evaluated at	
	Immigrant means	Native means
Mean immigrant-native difference in model with:		
1. MSA dummies	0.170	0.170
2. Full set of controls with interactions	0.115	0.070
Contribution to reduction in concentration between rows 1. and 2. (percents)		
	Xs	Differences in Coefficients
MSA (main effects in base, but interactions not)		
Individual characteristics (total)		
Log earnings	30.0	-3.4
Quarters of work	-0.8	-10.9
Age and sex	1.6	1.4
Language	3.3	-2.0
Schooling	23.7	-4.0
Employer characteristics (total)	2.2	-9.0
Firm size	73.6	2.8
Firm age and multi-unit status (interacted)	0.7	-31.8
Industry (total)	12.1	-1.1
Sector (categories below, + construction)	60.8	-9.7
Detail (modal ind excluded in each sector)	30.7	-21.0
Manufacturing detail (73 3-digit indus)	30.1	-12.3
Transportation, commun, utilities (14 3-dig)	-4.3	-8.7
Wholesale (18 3-digit industries)	0.7	5.4
Retail (33 3-digit industries)	1.8	-0.4
FIRE (4 3-digit industries)	7.8	-0.7
Services (51 3-digit industries)	0.1	-4.0
Neighborhood characteristics (total)	24.0	-0.3
Percent immigrant in residential tract	63.2	-8.7
Network index	59.9	-20.8
Shared commute index	1.9	-17.6
	1.4	-2.0
		-1.2
Total column share	166.7	-66.7
		91.2
		8.8

Notes: Regressions use controls deviated from the indicated mean so the coefficient on the immigrant indicator gives the predicted difference in coworker shares at that mean. The differences in Xs are evaluated using the coefficients for immigrants, while differences in coefficients are evaluated at native Xs deviated from the indicated mean.

C Supplemental Tables for Country of Origin Analyses

Table W-13 presents some summary statistics by country of origin. Table W-14 presents the Gelbach decomposition by country of origin for the main effects model. Table W-15 presents the language cross-effects by country of origin for the interactive model.

WEB APPENDIX

Table W-13: Selected sample characteristics by country of origin

	Share of workers	Share not speaking		Share of neighborhood workers who are:			
		English well		Same group	Mexican	Other immigrants	Natives
Mexico	4.9	36.2		26	.	15	59
El Salvador	0.7	35.4		8	19	22	51
Cuba	0.7	34.0		31	2	25	42
India	0.8	4.8		6	3	22	69
Philippines	1.3	3.6		9	8	21	62
China	0.5	33.5		10	4	29	57
Japan	0.2	9.6		1	4	20	75
Korea	0.4	19.2		4	4	25	67
Vietnam	0.8	30.1		12	9	21	59
Other	8.4	12.9		17	4	9	69

Notes: The figures in the table are percentages. Those in the first column would add to 100 if natives were included. Those in the right panel add to 100 for each row. In the "Other" row, the same group share includes all neighborhood immigrant workers who are not from one of the listed countries, while "Other immigrants" includes all neighborhood immigrant workers who are from the listed countries aside from Mexico.

WEB APPENDIX

Table W-14: Main-effects decomposition by country of origin

	Individual char			Employer char		Neighborhood char		
	Language	Education	Other	Industry	Other	Own res seg	Other res seg	Other
Own								
Mexico	12	8	0	39	2	39	0	-0
El Salvador	7	3	0	22	1	66	0	0
Cuba	2	0	0	2	1	93	2	0
India	-0	3	2	51	1	44	-2	1
Philippines	-0	-0	1	59	2	37	2	0
China	9	-0	0	14	2	78	-3	0
Japan	3	3	2	28	3	59	2	0
Korea	6	1	-0	14	9	72	-1	1
Vietnam	11	0	0	34	1	53	0	0
Other								
Mexico	54	5	1	22	4	8	5	0
El Salvador	24	6	-0	33	4	-1	33	0
Cuba	34	3	2	19	7	10	25	1
India	9	-2	2	44	6	-14	52	2
Philippines	7	-4	4	42	-1	-0	51	1
China	28	3	1	40	5	-7	29	1
Japan	27	-5	6	30	9	-1	32	2
Korea	24	-1	1	35	10	1	28	2
Vietnam	27	3	0	43	5	-4	26	1

Notes: The figures in the table are contributions based on Gelbach decompositions for the regressions used in Table 4. The “Own” panel gives shares of the difference in coefficients between columns (1) and (3), while the “Other” panel gives shares of the difference in coefficients between columns (2) and (4).

Table W-15: Language cross-effects by country of origin

Share of coworkers from:	Main effect	Difference in coworker share (relative to share for natives who do not speak English well) for immigrants who do not speak English well and are from:									
		Mexico	El Salvador	Cuba	India	Philippines	China	Japan	Korea	Vietnam	
Mexico	1.9	1.3	1.8	-3.0	-1.9	-1.4	-4.4	-2.2	-0.9	-3.0	
El Salvador	0.2	0.1	1.0	-0.3	-0.2	-0.3	-0.9	-0.8	-0.2	-0.5	
Cuba	0.2	-0.2	-0.1	5.5	-0.1	-0.2	-0.3	-0.2	-0.3	-0.3	
India	0.0	-0.0	-0.0	0.0	6.5	0.0	-0.6	-0.4	-0.2	-0.2	
Philippines	0.0	-0.3	-0.5	0.3	0.1	2.6	-1.0	-0.2	-0.7	-0.6	
China	0.0	-0.1	-0.1	-0.1	-0.2	0.1	6.5	-0.5	-0.5	1.7	
Japan	0.0	-0.0	-0.0	-0.0	-0.1	-0.1	-0.3	11.4	-0.1	-0.2	
Korea	0.1	0.0	0.1	-0.0	-0.1	-0.0	-0.2	1.1	8.3	-0.3	
Vietnam	0.1	-0.3	-0.2	-0.2	0.3	0.4	3.0	-0.2	-0.1	6.7	

Notes: Each of the figures in the table is a coefficient from a regression with the share of coworkers from the country listed in the left-most column as the dependent variable. The main effect column gives the coefficient on the dummy for not speaking English well. The other columns contain the coefficients on the interaction between a dummy variable for the country listed above the column and the dummy for not speaking English well. The coefficients have been rescaled to represent percentages (i.e. they are multiplied by 100). The regression specifications are as in Table 6, except that the language interactions are with our full set of country of origin dummies rather than the own/other grouping used in Table 6.

D Additional analysis of employer size effects

D.1 Simulations of employer size effects in a statistical model with segregation

If immigrants and natives are randomly allocated to jobs in proportion to their presence in the working population, the expected difference between immigrants and natives in the share of coworkers who are immigrant is zero regardless of employer size. However, we find that the distribution of immigrants across workplaces is inconsistent with random allocation, and that concentration is particularly high in small businesses. This raises the question of whether we should expect a general tendency to segregate to have the same effects on measured concentration in small and large businesses. The following sets up a statistical model that incorporates a tendency to segregate. The model is then used to simulate concentration by employer size. Under this model, the tendency to segregate has a much larger effect on concentration for very small employers than for those of modest or large size.

Suppose that employers of size s draw their workforces randomly from the population, but that some fraction of initial draws that involve an integrated workforce (i.e. some natives and some immigrants) are rejected and replaced with a new draw. For simplicity, we treat these draws as with replacement and assume that all employers are the same size, rather than dealing with a distribution of employer sizes. Assume that the outcome of each draw can be described using the binomial probability mass function:

$$b(i, s) = \binom{s}{i} p_D^i (1 - p_D)^{s-i} \quad (\text{D.1})$$

where i represents the number of immigrants in the workforce draw, s represents employer size, and p_D represents the fraction of workers who are immigrants in the group

WEB APPENDIX

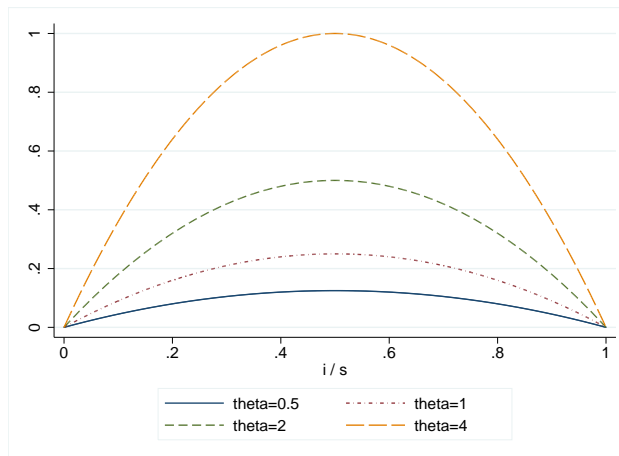
being sampled in draw D . For the initial draw, the parameter p_0 will equal the overall share of immigrants in the workforce.

Suppose that employers discard a draw with probability d which depends on workforce composition and a parameter θ that indexes the tendency to segregate ($0 \leq \theta \leq 4$).

$$d(i; s, \theta) = \frac{i}{s} \left(\frac{s-i}{s} \right) \theta \tag{D.2}$$

If an employer draws only immigrants or only natives, then $d = 0$ and the original draw is kept. If there are some of both types of employees, then the workforce is redrawn with probability d . This shifts some of the probability mass from more integrated towards more segregated types of employee mixes. Figure W-2 illustrates the shape of $d(\cdot)$ for various values of θ .

Figure W-2: Shape of function d



For $\theta = 4$, all draws with immigrants making up exactly half the workforce ($i/s = .5$) are discarded in the first round. However, even with $s = 2$, the final distribution includes some workforces with $i/s = .5$ because 1 immigrant and 1 native can be drawn in the second round.

If immigrants account for a small share of the population, they are disproportionately included in integrated workforces in the first draw. Because of this, the population that

WEB APPENDIX

the second draw is taken from has a somewhat higher share of immigrants than the initial population. For example, with $s = 2$ immigrants are always half of the workers in discarded first round draws, no matter what p_0 is.

Thus while we assume that the final draw is also binomial, the relevant immigrant share is given by:

$$p_1 = \frac{\sum_{j=1}^s b(j, s; p_0) * d(j; s, \theta) * j}{\sum_{j=1}^s b(j, s; p_0) * d(j; s, \theta) * s} \quad (\text{D.3})$$

and

$$Pr(i; s, p_0, \theta) = b(i, s; p_0) * (1 - d(i; s, \theta)) + b(i, s; p_1) * \left(\sum_{j=0}^s b(j, s; p_0) * d(j; s, \theta) \right) \quad (\text{D.4})$$

where the first term represents the probability that the initial draw has i immigrants and is not discarded, and the second term represents the probability that the final draw has i immigrants and that an initial draw was discarded.

For the simple case $s = 2$ and $\theta = 4$ (so $d = 1$ for the only integrated workforces—those with 1 immigrant, 1 native), $p_1 = .5$, and the probability of observing a workforce with 1 immigrant and 1 native in the final distribution simplifies to $p_0(1 - p_0)$ (half the binomial probability). Figure W-3 illustrates the difference between the distribution of the coworker mean with segregation and without for employers of varying size. It uses parameter values $\theta = 4$ and $p_0 = .25$. Smaller values of θ would reduce the shift in the distribution, while smaller values of p_0 shift the weight of both distributions to the left.

For immigrants, mean share of coworkers who are immigrant for employer size s is:

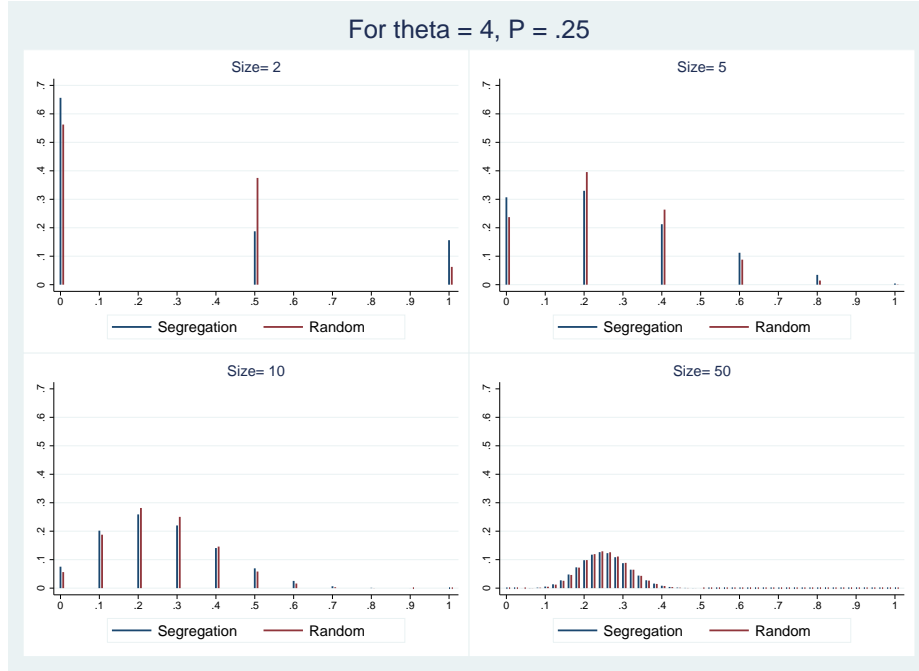
$$E(cw_I|s) = \sum_{i=0}^s \left(Pr(i|I_j = 1; s, p_0, \theta) * \frac{i-1}{s-1} \right) = \sum_{i=0}^s \left(Pr(i; s, p_0, \theta) * \frac{i}{sp_0} * \frac{i-1}{s-1} \right) \quad (\text{D.5})$$

and for natives,

WEB APPENDIX

$$E(cw_N|s) = \sum_{i=0}^s \left(Pr(i; s, p_0, \theta) * \frac{(s-i)}{s(1-p_0)} * \frac{i}{s-1} \right) \quad (D.6)$$

Figure W-3: Immigrant share distribution with and without segregation



The difference is then:

$$E(cw_N - cw_I|s) = \sum_{i=0}^s \left(Pr(i; s, p_0, \theta) * \frac{i[p_0(s-i) - (i-1)(1-p_0)]}{s(s-1)p_0(1-p_0)} \right) \quad (D.7)$$

Figures W-4 to W-6 plot out the relationship between employer size and coworker means for various values of the immigrant share of the overall workforce p (different colored lines in each graph), using segregation parameter $\theta = 4$. Figure W-4 graph gives the mean by firm size for immigrants, Figure W-5 is for natives, and Figure W-6 gives the difference between them. Figure W-7 repeats Figure W-6, except that it is parameterized to represent a lower level of segregation ($\theta = 1$). Examination of these figures makes a couple of patterns clear: (i) For very small employers (< 10 employees), the model can generate a large difference in coworker means, even with a relatively mild tendency to

WEB APPENDIX

segregate. (ii) Even for large theta, this model generates essentially no segregation in large firms.

Because the change in variance with sample size falls off quite quickly as size increases, we think that the statistical effect is unlikely to account for size effects among firms with more than 20 employees. Thus it might be reasonable to think of size effects based on the portion of our sample with at least 20 employees as representing the economic size relationship, while in smaller firms the size effect combines the economic and statistical relationships. Based on this assumption, we fit a flexible functional form to the size effect for the portion of our sample with at least 20 employees, and then use the fitted model to predict the size effect for smaller firms.¹ The lower panel of Figure W-8 superimposes this estimated/extrapolated relationship on the actual size-specific means. The analysis in Figure W-8 uses the full sample and accordingly only the variables in the full sample. It also breaks out the results for natives, recent and established immigrants.

For each of our three groups, we separately fit the relationship between mean coworker share and firm size over the range of firm size above 20. The points marked on each line represent the mean predicted coworker share for that employer size grouping. For example, in the lower graph, the 23% marked on the established immigrant line for the 500+ size group is the mean predicted value for established immigrants in this size range—a bit lower than the actual 27% share which is labeled in the upper graph. For groups 2-4, 5-9, and 10-19, the actual coworker share does not influence the fit of the model. The model projection fits the native means closely, which is unsurprising given that the native mean varies little with size. For immigrants, the projections underpredict the coworker means, with a particularly large gap for recent immigrants in the smallest firm size classes. If we take the projection as tracing out the real size effect, the

¹We use linear, quadratic and cubic functional forms to predict the size effect for smaller firms. The quadratic and cubic specifications gave very similar results. We show the quadratic results here.

WEB APPENDIX

evidence is consistent with a modest underlying size effect. Given that interpretation, the gap between the actual and projected mean then represents the purely statistical effect of size. Consistent with the statistical model above, this effect is large for very small firms, but rapidly decreases with size.

Figure W-4: Immigrant coworker mean and employer size ($\theta = 4$)

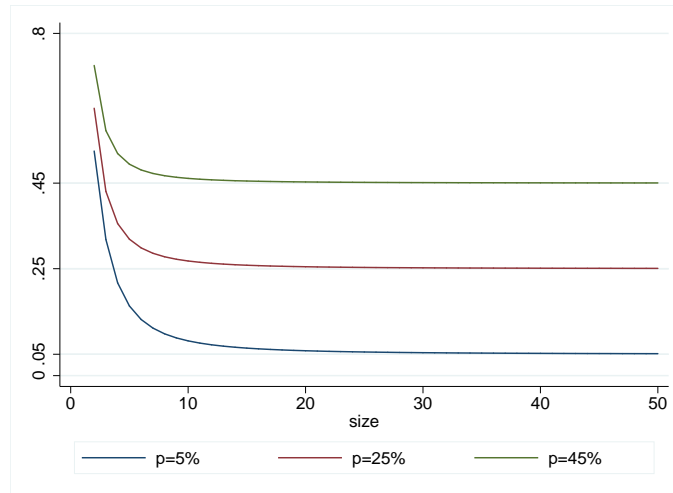
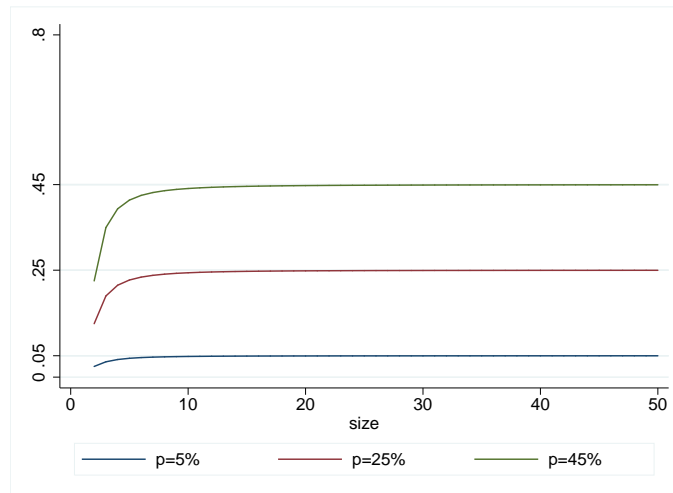


Figure W-5: Native coworker mean and employer size ($\theta = 4$)



WEB APPENDIX

Figure W-6: Immigrant-native difference in coworker mean and employer size ($\theta = 4$)

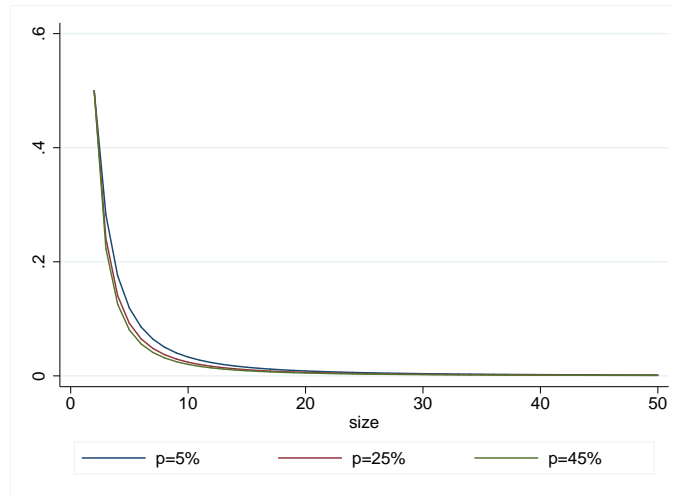
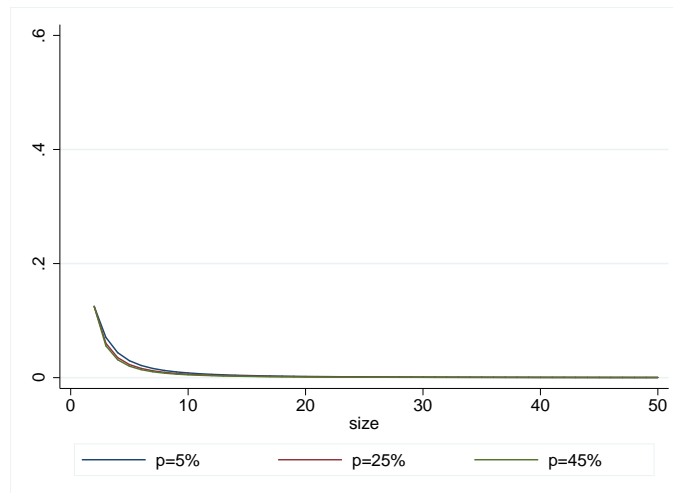


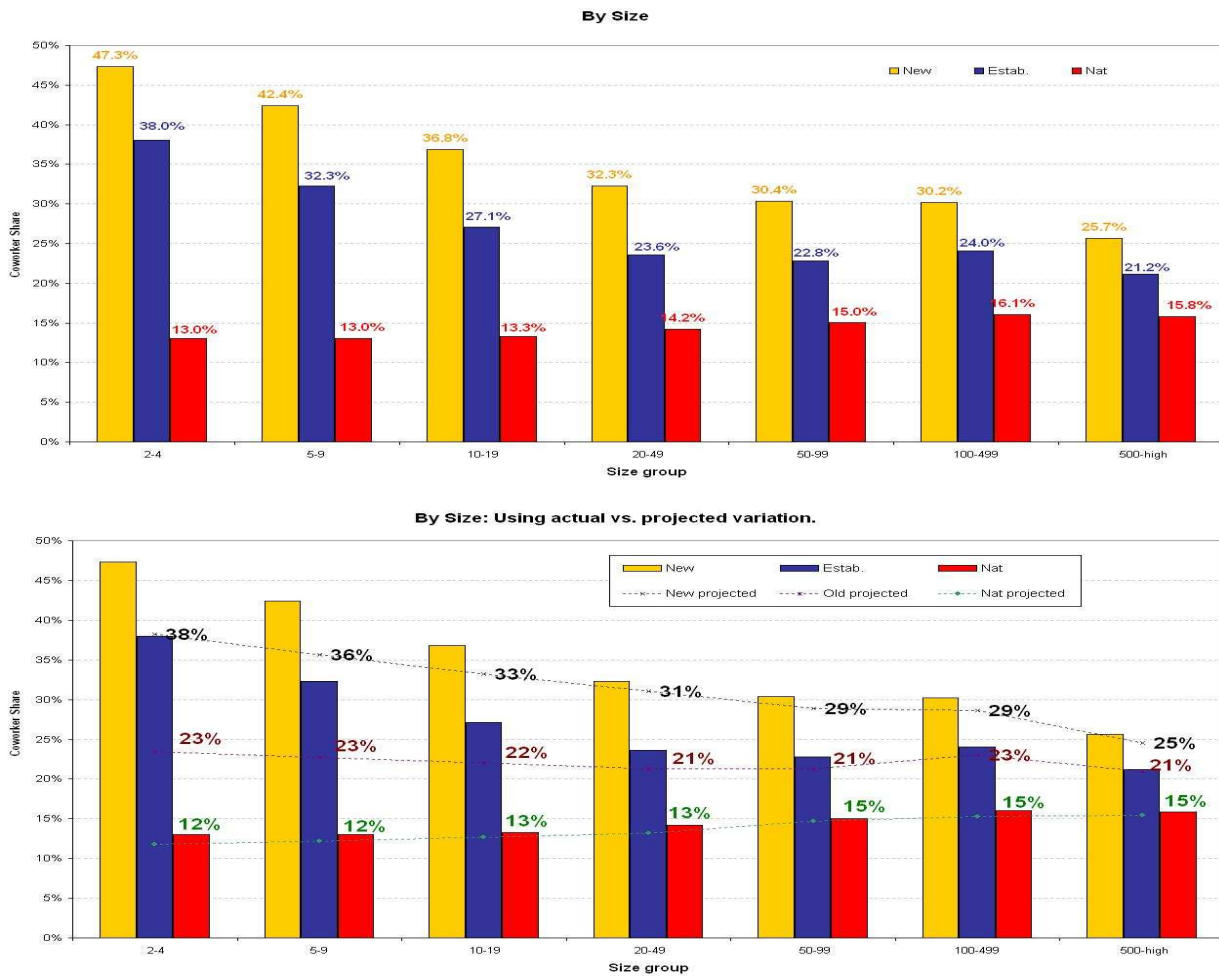
Figure W-7: Immigrant-native difference in coworker mean and employer size ($\theta = 1$)



WEB APPENDIX

D.2 Supplementary evidence on employer size differences

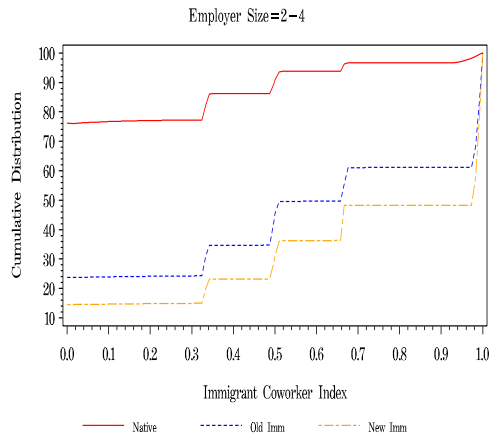
Figure W-8: Coworker share by employer size



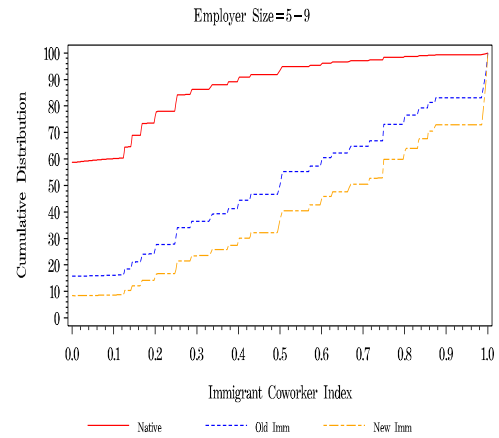
Notes: Evaluated at pooled mean for other control variables—MSA, sector, immigrant demographics, establishment age interacted with multi-unit status. Sector, individual’s age, establishment age, sex, units and MSA groups use total population distribution. Using full two-way interactions with individual status.

WEB APPENDIX

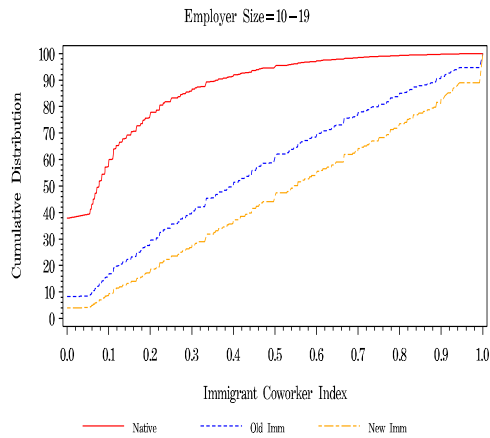
Figure W-9: Cumulative Distribution of Coworker Share by Worker Type and Employer Size



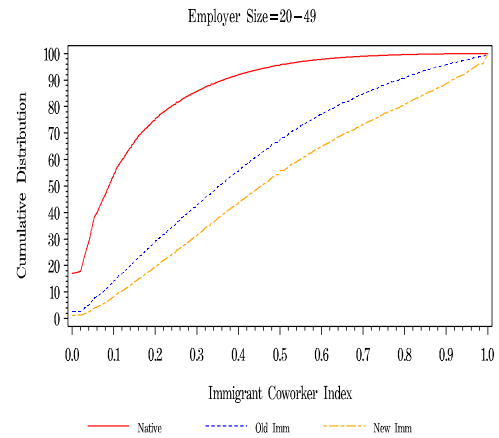
Source= LEHD database year 2000 second quarter



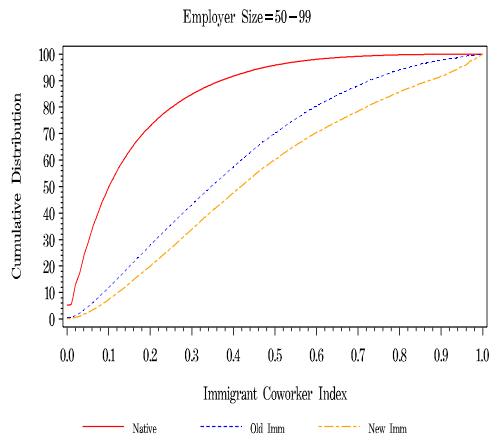
Source= LEHD database year 2000 second quarter



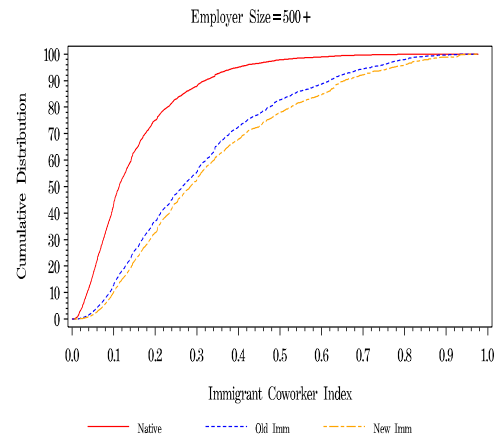
Source= LEHD database year 2000 second quarter



Source= LEHD database year 2000 second quarter



Source= LEHD database year 2000 second quarter



Source= LEHD database year 2000 second quarter

Source=LEHD database. Year 2000 second quarter.