Inferring Labor Income Risk From Economic Choices: An Indirect Inference Approach*

Fatih Guvenen† Anthony Smith‡

June 25, 2009

Abstract

This paper uses the information contained in the joint dynamics of households' labor earnings and consumption-choice decisions to quantify the nature and amount of income risk that households face. We accomplish this task by estimating a structural consumption-savings model using panel data from the Panel Study of Income Dynamics and the Consumer Expenditure Survey. Specifically, we estimate the persistence of labor income shocks, the extent of systematic differences in income growth rates, the fraction of these systematic differences that households know when they begin their working lives, and the amount of measurement error in the data. Income processes that differ along these dimensions can have vastly different implications for economic behavior as well as for a variety of policy questions. Although data on labor earnings alone can shed light on some of these dimensions, to assess what households know about their income processes requires using the information contained in their economic choices (here, consumption-savings decisions). In addition, consumption data increases the precision of our parameter estimates. To estimate the consumption-savings model, we use indirect inference, a simulation method that puts virtually no restrictions on the structural model and allows the estimation of income processes from economic decisions with general specifications of utility, frequently binding borrowing constraints, and missing observations. The main substantive findings are that income shocks are not very persistent, systematic differences in income growth rates are large, and individuals have substantial amounts of information about their future income prospects. Consequently, the amount of uninsurable lifetime income risk that households perceive is smaller than what is typically assumed in calibrated macroeconomic models with incomplete markets.

Keywords: Labor income risk, Indirect Inference, Heterogeneous Income Profiles, Persistence.

*For helpful comments and conversations, we thank Daron Acemoglu, Orazio Attanasio, Martin Browning, Raj Chetty, Lars Hansen, Yuichi Kitamura, Luigi Pistaferri, Jean-Marc Robin, and especially our discussants Victor Aguirregabiria, Richard Blundell, Stephane Bonhomme, Jonathan Parker, Sam Schulhofer-Wohl, Kjetil Storesletten, and Harald Uhlig, as well as seminar participants at Autonoma-Barcelona, Chicago FED, Iowa, Georgetown, HEC Lausanne, Princeton, Oxford, Seoul National, Sogang, Stanford, Tokyo, Toronto, UCL, Warwick, and the conference participants at the NBER Summer Institute, Banque de France/Bundesbank Spring 2007 Conference, 2007 SED Conference, 2007 Paris Workshop on Heterogeneity and Macrodynamics, IFS/UCL Workshop on Consumption Inequality and Income Dynamics, the 5th International PIER/IGIER Conference at Upenn, Advances in Theory Based Estimation Conference in Honor of Hansen and Singleton, 2007 ESSIM Conference, 2008 Research on Money and Markets Conference, JAE Conference on Distributional Dynamics at CEMFI-Madrid, Conference on Structural Models of the Labour Market at IFS, and 2008 AEA Winter meetings. Matthew Johnson provided excellent research assistance. Guvenen acknowledges financial support from the NSF under grant number SES-0649437. All remaining errors are our own.

†University of Minnesota, Federal Reserve Bank of Minneapolis, and NBER; guvenen@umn.edu; http://www.econ.umn.edu/~guvenen

‡Yale University; tony.smith@yale.edu; http://www.econ.yale.edu/smith.

1 Introduction

The goal of this paper is to elicit information about the nature of labor income risk from individuals' economic decisions (such as consumption-savings choice), which contain valuable information about the environment faced by individuals, including the future (income) risks they perceive.

To provide a framework for this discussion, consider the following process for log labor income of individual i with t years of labor market experience:

$$y_t^i = \underbrace{\left[a_0 + a_1 t + a_2 t^2 + a_3 Educ + ...\right]}_{\text{common life-cycle component}} \quad + \quad \underbrace{\left[\alpha^i + \beta^i t\right]}_{\text{profile heterogeneity}} \quad + \quad \underbrace{\left[z_t^i + \varepsilon_t^i\right]}_{\text{stochastic component}} \tag{1}$$

$$\text{where} \qquad z_t^i = \rho z_{t-1}^i + \eta_t^i, \qquad \text{and} \qquad \eta_t^i, \varepsilon_t^i \sim iid$$

The terms in the first bracket capture the life-cycle variation in labor income that is *common* to all individuals with given observable characteristics. The second component captures potential *individual-specific* differences in income growth rates (as well as in levels, which is less important). Such differences would be implied for example by a human capital model with heterogeneity in learning ability.[1] Finally, the terms in the last bracket represent the stochastic variation in income, which is written here as the sum of an AR(1) component and a purely transitory shock. This is a specification commonly used in the literature.

A vast empirical literature has estimated various versions of (1) in an attempt to answer the following two questions:

1. Do individuals differ *systematically* in their income growth rates? If such differences exist, are they quantitatively important? i.e., is $\sigma_\beta^2 \gg 0$?

2. How large and how persistent are income shocks? i.e., what is σ_η^2 and ρ?

Existing studies in the literature can be broadly categorized into two groups based on the conclusions they reach regarding these questions. The first group of papers *impose* $\sigma_\beta^2 \equiv 0$ based on outside evidence,[2] and with this restriction estimate ρ to be close to 1. We refer to this version of the process in (1) as the "*Restricted* Income Profiles" (*RIP*) model. The second group of papers do not impose any restrictions on (1) and find that ρ is significantly less than 1 and σ_β^2 is large. We refer to this version of (1) as the "*Heterogeneous* Income Profiles" (*HIP*) model. In other words,

[1] See for example, the classic paper by Ben-Porath (1967). For more recent examples of such a human capital model, see Guvenen and Kuruscu (2007, 2009), and Huggett, Ventura, and Yaron (2007).

[2] The outside evidence refers to a test proposed by MaCurdy (1982) in which he failed to reject the null of RIP against the alternative of HIP. Two recent papers, Baker (1997) and Guvenen (2009) argue that tests based on average autocovariances lack power against the alternative of a HIP process with an autoregressive component, and therefore, the lack of rejection of the RIP null does not provide evidence against the HIP model.

according to the RIP view, most of the rise in within-cohort income inequality over the life-cycle is due to large and persistent shocks, whereas in the HIP view, it is due to systematic differences in income growth rates. While overall we interpret the results of these studies, and especially those of the more recent papers, as more supportive of the HIP model, it is fair to say that this literature has not produced an unequivocal verdict.[3]

A key point to observe is that these existing studies do not utilize the information revealed by individuals' consumption-savings choice to distinguish between the HIP and RIP models.[4] But endogenous choices, such as consumption and savings, contain valuable information about the environment faced by individuals, including the future risks they perceive. Therefore, the main purpose of this paper is to use the restrictions imposed by the RIP and HIP processes on consumption data—in the context of a life-cycle model—to bring more evidence to bear on this important question. We elaborate further below on the advantages of focusing on consumption-savings choice (instead of using labor income data in isolation or using other endogenous choices, such as labor supply) for drawing inference about the labor income process.

In a sense, the two questions discussed so far only scratch the surface of what we mean by "the nature of income risk." This is because those two questions are statistical in nature, i.e., they relate to how the income process is viewed by the *econometrician* who studies past observations on individual income. But it is quite plausible that individuals may have more, or less, information about their income process than the econometrician at different points in their lifecycle, which raises two more questions:

3. If individuals indeed differ in their income growth rate as suggested by the HIP model, how much do individuals *know* about their β^i at *different points* in their life-cycle? In other words, what fraction of the heterogeneity in β^i constitutes "uncertainty" on the part of individuals as opposed to simply being some "known heterogeneity"?

4. What fraction of income movements measured by z_t^i and ε_t^i are "unexpected shocks" to income as opposed to being "anticipated changes" or simply "measurement error" in survey data?

These questions are inherently different than the first two in that they pertain to how *individuals* perceive their income process. As such, they cannot be answered using income data alone, but the answers can be teased out, again, from individuals' economic decisions. To give one example (to

[3] A short list of these studies includes MaCurdy (1982), Abowd and Card (1989), and Topel (1990), which find support for the RIP model; Lillard and Weiss (1979), Hause (1980), and especially the more recent studies such as Baker (1997), Haider (2001), and Guvenen (2009) which find support for the HIP model.

[4] Two recent papers do use consumption data but in a more limited fashion than this paper intends to do. In a recent paper, Huggett, Ventura, and Yaron (2007) study a version of the Ben-Porath model and make some use of consumption data to measure the relative importance of persistent income shocks versus heterogeneity in learning ability. Although the income process generated by their model does not exactly fit into the specification in equation (1) their results are informative. Second, Guvenen (2007) uses consumption data to investigate if a HIP model estimated from income data is consistent with some stylized consumption facts. While both of these papers are informative about the HIP versus RIP debate, they make limited use of consumption data, especially of the *dynamics* of consumption behavior.

question 4), consider a married couple who jointly decide that they will both work up to a certain age and then will have children at which time one of the spouses will quit his/her job to take care of the children. The ensuing large fall in household income will appear as a large permanent shock to the econometrician using labor income data alone, but consumption (and savings) data would reveal that this change has been anticipated.

Several papers have used consumption data and shed light on various properties of income processes (among others, Hall and Mishkin (1982), Deaton and Paxson (1994), Blundell and Preston (1998), and Blundell, Pistaferri, and Preston (2008)). This paper contributes to this literature in the following ways. First and foremost, existing studies consider only versions of the RIP model (i.e., they set $\sigma_\beta^2 \equiv 0$ at the outset), whereas a major goal of our paper is to distinguish between HIP and RIP models. Second, and furthermore, these studies also *impose* $\rho \equiv 1$, and only estimate the innovation variances. In other words, there is no existing study to our knowledge that uses consumption data and estimates ρ. Therefore, this paper will leave ρ unrestricted (even in the RIP version) and exploit consumption and income data jointly to pin down its value. Since many incomplete markets models are still calibrated using versions of the RIP process, the results of this exercise should be useful for calibrating those models. The third contribution of this paper will be in the method used for estimation—indirect inference—which is much less restrictive than, and has several important advantages over, the GMM approach used in previous work.[5]

1.1 Why Look at Consumption-Savings Choice?

Even if one is only interested in the first two questions raised above, using information revealed by intertemporal choices has important advantages. This is because one difficulty of using income data alone is that identification between HIP and RIP models partly depends on the behavior of the *higher*-order autocovariances of income.

To see this clearly, consider the case where the panel data set contains income observations on a single cohort over time. In this case, the second moments of the cross-sectional distribution for this cohort are given by:

$$\begin{aligned} var\left(y_t^i\right) &= \left[\sigma_\alpha^2 + 2\sigma_{\alpha\beta}t + \sigma_\beta^2 t^2\right] + var\left(z_t^i\right) + \sigma_\varepsilon^2 \\ cov\left(y_t^i, y_{t+n}^i\right) &= \left[\sigma_\alpha^2 + \sigma_{\alpha\beta}\left(2t+n\right) + \sigma_\beta^2 t\left(t+n\right)\right] + \rho^n var\left(z_t^i\right), \end{aligned} \tag{2}$$

where $t = 1, ..T$, and $n = 1, .., T - t$. There are two sources of identification between the RIP and HIP processes, which can be seen by inspecting these formulas. The first piece of information is provided by the change in the cross-sectional variance of income as the cohort ages (i.e., the

[5]Two important differences of the present paper from Guvenen (2007) is that that paper (i) only estimated $\widehat{\sigma}^2_{\beta|0}$ from consumption data, taking all other parameters as estimated from income data, and (ii) only used the rise in within-cohort consumption inequality as a moment condition. The present paper instead (i) brings consumption data to bear on the estimation of the *entire* vector of structural parameters, and (ii) does this by systematically focusing on the dynamic relationship between consumption and income movements.

diagonal elements of the variance-covariance matrix), which is shown on the first line of (2). The terms in the square bracket capture the effect of profile heterogeneity, which is a *convex* increasing function of age. The second term captures the effect of the AR(1) shock, which is a *concave* increasing function of age as long as $\rho < 1$. Thus, if the variance of income in the data increases in a convex fashion as the cohort gets older, this would be captured by the HIP terms (notice that the coefficient on t^2 is σ_β^2), whereas a non-convex shape would be captured by the presence of AR(1) shocks.

The second source of identification is provided by the autocovariances displayed in the second line. The covariance between ages t and $t+n$ is again composed of two parts. As before, the terms in the square bracket capture the effect of heterogeneous profiles and is a convex function of age. Moreover, the coefficients of the linear and quadratic terms depend both on t and n, which allows covariances to be decreasing, increasing or non-monotonic in n at each t. The second term captures the effect of the AR(1) shock, and notice that for a given t, it depends on the covariance lag n only through the geometric discounting term ρ^n. The strong prediction of this form is that, starting at age t, covariances should decay geometrically at the rate ρ, regardless of the initial age. Thus, in the RIP model (which only has the AR(1) component) covariances are restricted to decay at the same rate at every age, and cannot be non-monotonic in n.

Notice that for a cohort with 40 years of working life, there are only 40 variance terms, but many more—780 ($= (40 \times 41)/2 - 40$) to be precise—autocovariances, which provide crucial information for distinguishing between HIP and RIP processes. The main difficulty is that because of sample attrition, fewer and fewer individuals contribute to these higher autocovariances, raising important concerns about potential selectivity bias. To give a rough idea, if one uses labor income data from the Panel Study of Income Dynamics (PSID), and selects all individuals who are observed in the sample for 3 years or more (which is a typical sample selection criterion), the number of individuals contributing to the 20th autocovariance will be about $1/5$ of the number of individuals contributing to the 3rd autocovariance. To the extent that these individuals are not a completely random subsample of the original sample, covariances at different lags will have variation due to sample selection that can confound the identification between HIP and RIP models.

In contrast, because of its forward-looking nature, even short-run movements in consumption, and the immediate response of consumption to income innovations contain information about the perceived *long-run* behavior of the income process. Therefore even lower-order covariances of consumption would help in distinguishing HIP from RIP. (Notice that the dynamic aspect of the consumption-savings choice also distinguishes it from other decisions, such as labor supply, which are static in nature, unless one models intertemporally non-separable preferences in leisure.)

2 A Framework For Inferring Income Risk

2.1 Bayesian Learning about Income Profiles

Embedding the HIP process into a life-cycle model requires one take a stand on what individuals know about their own β^i. We follow Guvenen (2007) and assume that individuals enter the labor market with some prior belief about their β^i and then update their beliefs over time in a Bayesian fashion. Notice that the prior variance of this belief (denote by $\widehat{\sigma}^2_{\beta|0}$) measures how uncertain individuals are about their own β^i at time zero, addressing the third question raised in the introduction.

We now cast the learning process as a Kalman filtering problem which allows us to obtain recursive updating formulas for beliefs. Each individual knows her own α^i, observes her income and the transitory shock, y^i_t and ε^i_t, and must learn about $\mathbf{S}^i_t \equiv \left(\beta^i, z^i_t\right)$.[6] It is convenient to express the learning process as a Kalman filtering problem using the state-space representation. In this framework, the "state equation describes the evolution of the vector of state variables that is unobserved by the decision maker:

$$\underbrace{\begin{bmatrix} \beta^i \\ z^i_{t+1} \end{bmatrix}}_{\mathbf{S}^i_{t+1}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} \beta^i \\ z^i_t \end{bmatrix}}_{\mathbf{S}^i_t} + \underbrace{\begin{bmatrix} 0 \\ \eta^i_{t+1} \end{bmatrix}}_{\nu^i_{t+1}}$$

Even though the parameters of the income profile have no dynamics, including them into the state vector yields recursive updating formulas for beliefs using the Kalman filter. A second (observation) equation expresses the observable variable in the model—in this case, log income net of the fixed effect and transitory shock ($\tilde{y}^i_t$)—as a linear function of the underlying hidden state:

$$\tilde{y}^i_t \equiv y^i_t - \alpha^i - \varepsilon^i_t = \begin{bmatrix} t & 1 \end{bmatrix} \begin{bmatrix} \beta^i \\ z^i_t \end{bmatrix} = \mathbf{H}'_t \mathbf{S}^i_t,$$

We assume that both shocks have $i.i.d$ Normal distributions and are independent of each other, with $\mathbf{Q}$ and R denoting the covariance matrix of ν^i_t and the variance of ε^i_t respectively. To capture an individual's initial uncertainty, we model her prior belief over (β^i, z^i_1) by a multivariate Normal

[6] The assumption that α^i is observable is fairly innocuous here, because the uncertainty regarding this parameter is resolved very quickly even when the individual has substantial prior uncertainty, as shown in Guvenen (2007). Furthermore, the assumption that ε is observable is important but also seems plausible, since purely transitory shocks are likely to be easier to distinguish from persistent ones, whereas persistent shocks are more easily confused with the trend. We have also estimated a version of this model where ε was unobservable and found that the current specification fits the data better. One reason for the poorer fit of the specification with unobservable ε is that consumption responds "too much" to transitory shocks since the individual cannot tell it apart from persistent shocks. Finally, another advantage of the observable ε assumption is that it simplifies exposition: as will become clear below, in this case the consumption-savings model with a HIP process and learning nests as a special case the RIP process with no learning.

distribution with mean

$$\widehat{\mathbf{S}}^i_{1|0} \equiv (\widehat{\beta}^i_{1|0}, \widehat{z}^i_{1|0})$$

and variance-covariance matrix:

$$\mathbf{P}_{1|0} = \begin{bmatrix} \sigma^2_{\beta,0} & 0 \\ 0 & \sigma^2_{z,0} \end{bmatrix}$$

where we use the short-hand notation $\sigma^2_{\cdot,t}$ to denote $\sigma^2_{\cdot,t+1|t}$. After observing $\left(\tilde{y}^i_t, \widetilde{y}^i_{t-1}, ..., \widetilde{y}^i_1\right)$, the posterior belief about $\mathbf{S_t}^i$ is Normally distributed with a mean vector $\widehat{\mathbf{S}}^i_t$, and covariance matrix $\mathbf{P}_t$. Similarly, let $\widehat{\mathbf{S}}^i_{t+1|t}$ and $\mathbf{P}_{t+1|t}$ denote the one-period-ahead *forecasts* of these two variables respectively. Then, $\tilde{y}^i_{t+1}$ has a Normal distribution conditional on an individual's current beliefs:

$$\tilde{y}^i_{t+1} | \widehat{\mathbf{S}}^i_t \sim N\left(\mathbf{H}'_{t+1}\widehat{\mathbf{S}}^i_{t+1|t}, \mathbf{H}'_{t+1}\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}\right). \tag{3}$$

In this particular problem, the Kalman filtering equations can be manipulated to obtain some simple expressions. Define:

$$\begin{aligned} A_t &\equiv t\sigma^2_{\beta,t|t-1} + \sigma_{\beta z,t|t-1}, \\ B_t &\equiv t\sigma_{\beta z,t|t-1} + \sigma^2_{z,t|t-1}, \\ X_t &\equiv var_{t-1}\left(y^i_t\right) = A_t t + B_t \end{aligned}$$

Using the Kalman recursion formulas:

$$\begin{bmatrix} \widehat{\beta}^i_{t+1|t} \\ \widehat{z}^i_{t+1|t} \end{bmatrix} = \begin{bmatrix} \widehat{\beta}^i_{t|t-1} \\ \rho\widehat{z}^i_{t|t-1} \end{bmatrix} + \begin{bmatrix} A_t/X_t \\ \rho B_t/X_t \end{bmatrix} \left(\tilde{y}^i_t - \left(\widehat{\beta}^i_{t|t-1}t + \widehat{z}^i_{t|t-1}\right)\right)$$

Define the innovation to beliefs:

$$\widehat{\xi}_t = \tilde{y}^i_t - \left(\widehat{\beta}^i_{t|t-1}t + \widehat{z}^i_{t|t-1}\right)$$

Then we can rewrite:

$$\widehat{\beta}^i_{t+1|t} - \widehat{\beta}^i_{t|t-1} = (A_t/X_t)\,\widehat{\xi}_t \tag{4}$$
$$\widehat{z}^i_{t+1|t} - \rho\widehat{z}^i_{t|t-1} = (\rho B_t/X_t)\,\widehat{\xi}_t \tag{5}$$

An important point to note is that $\widehat{\xi}_t$ and (the true innovation to income) η^i_t do not need to have the same sign, a point that will play a crucial role below. Finally, the posterior variances evolve:

$$\sigma^2_{\beta,t+1|t} = \sigma^2_{\beta,t|t-1} - \frac{A_t^2}{X_t} \tag{6}$$

$$\sigma^2_{z,t+1|t} = \rho^2 \left[\sigma^2_{z,t|t-1} - \frac{B_t^2}{X_t}\right] \tag{7}$$

For a range of parameterizations A/X has an inverse U-shape over the life-cycle. Therefore, beliefs about β^i changes (and precision rises) slowly early on but become faster over time. In contrast, B/X declines monotonically. As shown in Guvenen (2007), optimal learning in this model has some interesting features. In particular, learning is very slow and the speed of learning has a non-monotonic pattern over the life-cycle (which is due to the fact that A/X has an inverse U-shape). If instead the prior uncertainty were to resolve quickly, consumption behavior after the first few years would not be informative about the prior uncertainty faced by individuals ($\widehat{\sigma}^2_{\beta|0}$).

Finally we discuss how an individual's prior belief about β^i is determined. Suppose that the distribution of income growth rates in the population is generated as $\beta^i = \beta^i_k + \beta^i_u$, where β^i_k and β^i_u are two random variables, independent of each other, with zero mean and variances of $\sigma^2_{\beta_k}$ and $\sigma^2_{\beta_u}$. Clearly then, $\sigma^2_\beta = \sigma^2_{\beta_k} + \sigma^2_{\beta_u}$. The key assumption we make is that individual i observes the realization of β^i_k, but not of β^i_u (hence the subscripts indicate *k*nown and *u*nknown, respectively). Under this assumption, the prior mean of individual i is $\widehat{\beta}^i_{1|0} = \beta^i_k$, and the prior variance is $\sigma^2_{\beta,0} = \sigma^2_{\beta_u}$. Further, to express the amount of prior uncertainty in relation to the heterogeneity in income growth rates, it is useful to write $\sigma_{\beta,0} = \lambda\sigma_\beta$, where λ is the fraction of population dispersion in growth rates that represents uncertainty on the part of individuals at the time they enter the labor market. Two polar cases deserve special attention. When $\lambda = 1$, individuals do not have any *private prior information* about their income growth rate (i.e., $\sigma^2_{\beta,0} = \sigma^2_\beta$ and $\widehat{\beta}^i_{1|0} = \overline{\beta}$ for all i, where $\overline{\beta}$ is the population average). At the other extreme, when $\lambda = 0$, each individual observes β^i completely and faces no prior uncertainty about its value.[7]

2.2 Consumption-Savings Decision

Each individual lives for T years and works for the first R ($< T$) years of her life, after which she retires. Individuals do not derive utility from leisure and, therefore, supply labor inelastically.[8] During the working life, the income process is given by the general (HIP) process specified in

[7]Notice that since each individual knows her α^i this already gives some information about her β^i as long as the two parameters are correlated. It is natural to think of λ as already incorporating this information. (One way to think about this is that β^i_k is captures all the correlation between α^i and β^i and $\beta^i_u \perp \alpha^i$.) It can be easily shown that there is an upper bound to λ (that depends on $\sigma^2_\alpha, \sigma^2_\beta$, and $\sigma_{\alpha\beta}$) that captures this minimum information obtained from α^i alone, and this upper bound is 1 when $\sigma_{\alpha\beta} = 0$.

[8]The labor supply choices of both the husband and wife appear to be important for drawing robust inference about the nature of income risk. Such extensions are conceptually feasible with indirect inference, although it increases computational costs. We therefore leave this extension for future research.

equation (1). Individuals can save at a constant interest rate r and can also borrow at the same rate subject to a lower limit specified below.

The relevant state variables for this dynamic problem are cash-on-hand (assets plus labor income), ω_t^i, and the vector of mean beliefs, $\widehat{\mathbf{S}}_t$. The dynamic programming problem of the individual can be written as:

$$V_t^i(\omega_t^i, \widehat{\mathbf{S}}_t^i; \alpha^i) = \max_{C_t^i, a_{t+1}^i} \left\{ U(C_t^i) + \delta E_t \left[V_{t+1}^i(\omega_{t+1}^i, \widehat{\mathbf{S}}_{t+1}^i; \alpha^i) \right] \right\}$$

$$s.t.$$

$$C_t^i + a_{t+1}^i = \omega_t \tag{8}$$

$$\omega_t = (1+r)\, a_t^i + Y_t^i \tag{9}$$

$$a_{t+1}^i \geq \underline{a}, \qquad \text{and Kalman recursions}$$

for $t = 1, ..., R-1$, where $Y_t^i \equiv e^{y_t^i}$ is the *level* of income, $\underline{a}$ is the borrowing limit, and V_t^i is the value function of a t year-old individual. The evolutions of the vector of beliefs and its covariance matrix are governed by the Kalman recursions given in equations (4, 5, 6, 7). Finally, the expectation is taken with respect to the conditional distribution of $\tilde{y}_{t+1}^i$ given by equation (3) and the distribution of the transitory shock, ε_{t+1} (i.e., a double integral).

During retirement, individuals receive annual pension payments from a retirement system that mimics the salient features of the US Social Security Administration's Old-Age Insurance Benefits System. Since there is no uncertainty (or learning) after retirement, the problem simplifies significantly:

$$V_t^i(\omega_t^i; Y) = \max_{c_t^i, a_{t+1}^i} \left[U(C_t^i) + \delta V_{t+1}^i(\omega_{t+1}^i; Y) \right] \tag{10}$$

$$s.t \qquad Y^i = \Phi\left(Y_R^i; \overline{Y}\right), \text{ and } eq.\ (8, 9)$$

for $t = R, ..., T$, with $V_{T+1} \equiv 0$; $\overline{Y}$ is the average labor earnings in the economy; and $\Phi\left(Y_R^i; \overline{Y}\right)$ is the retirement benefit function described more fully in Section 5.

2.3 Special Cases

The model described above encompasses a number of special cases of interest. Without any restrictions imposed, the framework is the general HIP model with Bayesian learning about income profiles. Another important benchmark is obtained when $\sigma_\beta \equiv 0$, in which case the model reduces to the standard RIP model, extensively studied in the literature. Yet, a third case of interest is when $\sigma_\beta > 0$ and $\lambda = 0$. In this case, individuals face a HIP process, but the only source of uncertainty arises from shocks as in the RIP model. This is an intermediate case between the HIP and RIP models.

3 An "Indirect Inference" Approach

In this section, we describe the indirect inference method used to estimate the parameters of the structural model laid out in the previous section. Essentially, indirect inference is a simulation-based method for estimating, or making inferences about, the parameters of economic models (Smith (1993), Gourieroux, Monfort, and Renault (1993)). It is most useful in estimating models for which the likelihood function (or any other criterion function that might form the basis of estimation) is analytically intractable or too difficult to evaluate, as is the case here: neither one of the consumption-savings models described above yields simple estimable equations that would allow a maximum likelihood or GMM estimation. In the general model with Bayesian learning, individuals' consumption choice depends on their beliefs, which is unobserved by the econometrician. Furthermore, even if we set $\sigma_\beta \equiv 0$, and consider the RIP model, the existence of borrowing constraints and CRRA utility makes the derivation of exact structural equations impossible. Aware of these difficulties, previous studies (which focused on the RIP model) made a number of simplifying assumptions, such as the absence of binding borrowing constraints, separability between consumption and leisure in the utility function, a simplified retirement structure, and so on, and employed approximations to the true structural equations in order to make GMM feasible.

Instead, the hallmark of indirect inference is the use of an "auxiliary model" to capture aspects of the data upon which to base the estimation. One key advantage of indirect inference over GMM is that this auxiliary model does *not* need to correspond to any valid moment condition of the structural model for the estimates of the structural parameters to be consistent. This allows significant flexibility in choosing an auxiliary model: it can be any sufficiently rich statistical model relating the model variables to each other as long as each structural parameter of the economic model has an independent effect on the likelihood of the auxiliary model.[9] This also allows one to incorporate many realistic features into the structural model without having to worry about whether or not one can directly derive the likelihood (or moment conditions for GMM) in the presence of these features.

While indirect inference shares a basic similarity to MSM (Method of Simulated Moments), it differs from MSM in its use of an "auxiliary model" to generate moment conditions. In particular, indirect inference allows one to think in terms of the dynamic structural relationships that characterize most economic models that are difficult to express as simple unconditional moments as is often done with MSM. We illustrate this in the description of the auxiliary model below.

3.1 Toward an Auxiliary Model

To understand the auxiliary model that will be used, it is useful to elaborate on the dependence of consumption choice on income shocks. As noted above, the key idea behind an auxiliary model

[9] In addition to some regularity conditions that the auxiliary model has to satisfy the precise specification of the auxiliary model will also matter for the efficiency of the estimator.

is that it should be an econometric model that is easy to estimate, yet one that captures the key statistical relations between the variables of interest in the model. Good candidates for an auxiliary model are provided by structural relationships that hold in models that are similar to the HIP and RIP models described above, and yet simple enough to allow the derivation of such relationships.

To this end, consider a simplified version of the HIP model, where we assume: (i) quadratic utility; (ii) $\delta\left(1+r^{f}\right)=1$, and (iii) no retirement. Further consider a simpler form of the HIP process:

$$Y_t^i=\alpha^i+\beta^i t+z_t^i, \tag{11}$$

where income, instead of its logarithm, is linear in the underlying components, and we set $\varepsilon_t^i \equiv 0$.[10] Under these assumptions, optimal consumption choice implies

$$\Delta C_t=\frac{1}{\varphi_t}\left[(1-\gamma)\sum_{s=0}^{T-t}\gamma^s\left(E_t-E_{t-1}\right)Y_{t+s}\right], \tag{12}$$

where $\gamma=1/\left(1+r^{f}\right)$ and $\varphi_t=\left(1-\gamma^{T-t+1}\right)$ is the annuitization factor. Substituting the simple HIP process in (11), we have:

$$E_t\left(Y_{t+s}^i\right)=\alpha^i+\widehat{\beta}_t^i\,(t+s)+\rho^s\widehat{z}_t$$

$$\left(E_t-E_{t-1}\right)Y_{t+s}^i=\left(\widehat{\beta}_t^i-\widehat{\beta}_{t-1}^i\right)(t+s)+\rho^s\widehat{\eta}_t^i$$

Substituting this last expression into (12) yields

$$\Delta C_t=\Phi_{t,T}^r\left(\widehat{\beta}_t^i-\widehat{\beta}_{t-1}^i\right)+\Psi_{T-t}^{r,\rho}\widehat{\eta}_t^i \tag{13}$$

where $\Phi_{t,T}^r$ and $\Psi_{T-t}^{r,\rho}$ are some known age-varying positive terms.[11] Moreover, $\Phi_{t,T}^r$ is a (slightly) convex increasing function of t, and $\Psi_{T-t}^{r,\rho}$ is constant and equal to 1 when $\rho=1$.

Remark. Before moving further, it is important to stress that equation (13) is obtained by fully solving the consumption-savings model and therefore requires (i) using the Euler equation, (ii) imposing the budget constraint, and (iii) taking a stand on a specific stochastic process for income. In this sense, the analysis here is in the spirit of Hall and Mishkin (1982) (and, more recently, Blundell, Pistaferri, and Preston (2008)) who derived the full consumption function (as we do here) rather than Hall (1978) who required only the Euler equation to hold. Therefore, by imposing stronger restrictions the current approach allows us to estimate the parameters of the

[10] Closed form solutions such as those below can still be derived in the presence of transitory shocks. We abstract from them here only for the clarity of exposition.

[11] $\Phi_{t,T}^r \equiv\left[\left(\frac{\gamma}{1-\gamma}\right)+\frac{t-(T+1)\gamma^{T-t+1}}{1-\gamma^{T-t+1}}\right]$ and $\Psi_{T-t}^{r,\rho}\equiv\frac{1-\gamma}{1-\gamma\rho}\left[\frac{\left(1-(\gamma\rho)^{T-t+1}\right)}{\left(1-\gamma^{T-t+1}\right)}\right]$

income process in addition to the preference parameters, which is all one can estimate in the Euler equation approach.

Continuing with the derivation, the Kalman filtering formulas above imply

$$\begin{aligned}\widehat{\beta}_t^i - \widehat{\beta}_{t-1}^i &= (A_t/X_t)\,\widehat{\xi}_t, \\ \widehat{z}_t^i - \rho\widehat{z}_{t-1}^i &= (B_t/X_t)\,\widehat{\xi}_t,\end{aligned} \tag{14}$$

which can be obtained easily from equations (4) and (5), but now $\widehat{\xi}_t$ needs to be reinterpreted as the *level* deviation: $Y_t^i - \left(\widehat{\beta}_{t|t-1}^i t + \widehat{z}_{t|t-1}^i\right)$. Plugging these expressions into (13), we get in the HIP model:

$$\Delta C_t = \left[\Phi_{t,T}^r\,(A_t/X_t) + \Psi_{T-t}^{r,\rho}\,(\rho^s B_t/X_t)\right] \times \widehat{\xi}_t \tag{15}$$

Instead in the RIP model, we have:

$$\Delta C_t = \Psi_{T-t}^{r,\rho} \times \eta_t^i \tag{16}$$

The last two equations underscore one of the key differences between the two frameworks: in the RIP model, only current η_t^i matters for consumption response, whereas in the HIP model the entire history of shocks matters through beliefs. As a result, two individuals hit by the same η_t^i may react differently depending on their history. Specifically, in the HIP model η_t^i and $\widehat{\xi}_t$ may have different signs. Therefore, an increase in income ($\Delta Y_t^i > 0$) may cause a fall in consumption ($\Delta C_t^i < 0$). In the RIP model, this will never happen.

An example of this case is shown in figure 1. This graph plots the income paths of two individuals, where we continue to assume $\varepsilon_t^i \equiv 0$ for simplicity. Individual 1 experiences a faster average income growth rate in the first five periods than individual 2, but observes the same rise in income between periods five and six. If these income paths are generated by a RIP process (and individuals correctly perceives them as such), then both individuals will adjust their consumption growth by exactly the same amount between periods five and six. Instead, if the truth is as in the HIP model, individual 1 will have formed a belief that her income growth rate is higher than that of individual 2, and was expecting her income to be closer to the trend line (shown by the dashed blue line). Therefore, even though her income increases, it is significantly below the trend ($\widehat{\xi}_t < 0$), which causes her to revise down her beliefs about her true β^i, and consequently her consumption level from equation (15). Specifically, we have:

> *Prediction 1:* The HIP model with Bayesian learning predicts that controlling for current income growth, consumption growth will be a *decreasing* function of average *past income growth rate.* Instead, the RIP model predicts no dependence on past income growth rate of this kind.

It is also possible to obtain a closed-form expression for the consumption (level) in the simplified version of the HIP model described above. One can easily see that the *level* of consumption contains

Figure 1: Distinguishing HIP from RIP (from Consumption Changes)

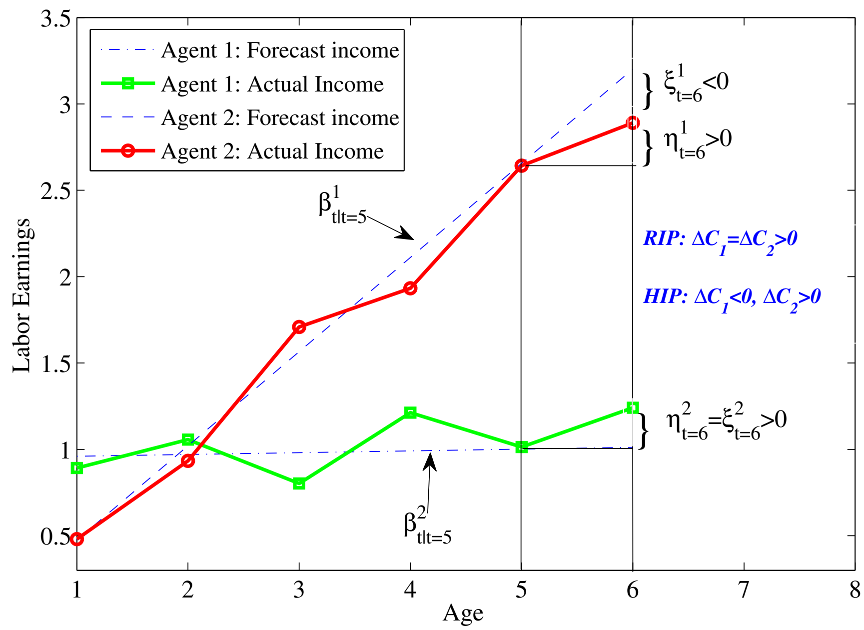


Figure 2: Distinguishing HIP from RIP (from Consumption Levels)

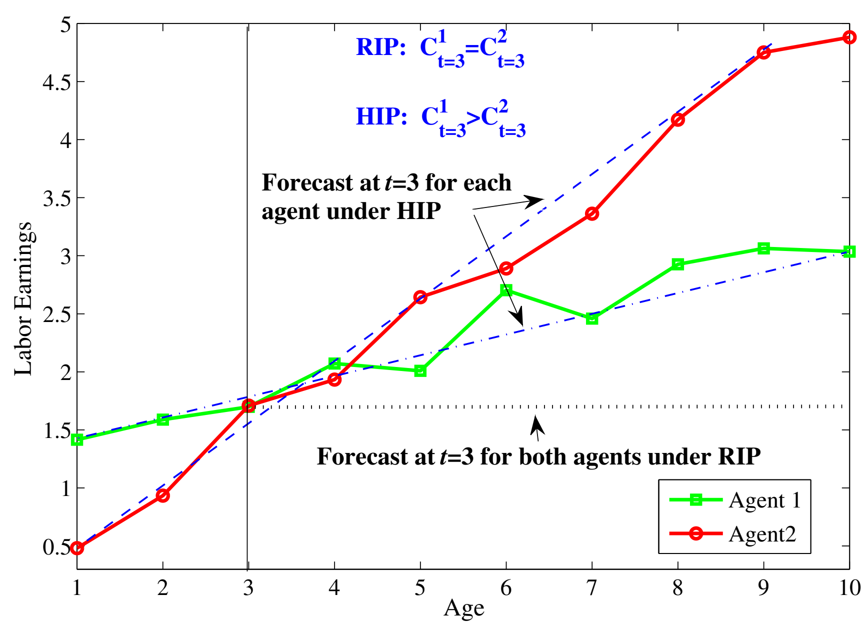
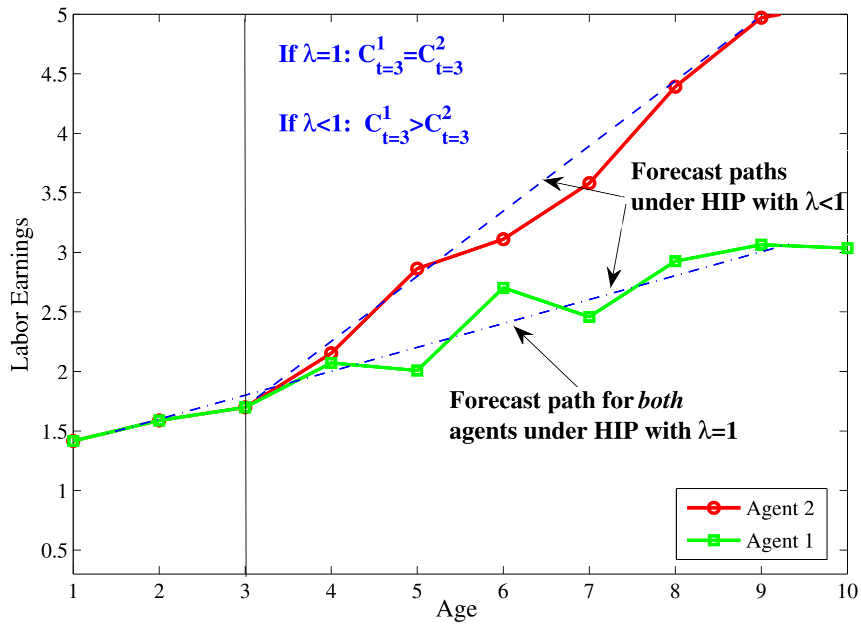


Figure 3: Determining the Amount of Prior Knowledge in HIP



information about whether individuals perceive their income process as HIP or RIP. An example of this is shown in figure 2. This example is most easily explained when income shocks are permanent ($\rho = 1$), which we assume for the moment. As before, individuals realize different income growth rates up to period 3. Under the RIP model, both individuals' forecast of their future income is the same as their current income (shown with the horizontal dashed lines). In contrast, with a HIP process, individual 1 will expect a higher income growth rate and therefore a much higher lifetime income than individual 2. Therefore, the first individual will have a higher consumption level than individual 2 at the same age, despite the fact that their current income levels are very similar. Therefore, we have:

> *Prediction 2:* The HIP model predicts that controlling for the current level of income *and* past average income level, an individual's current consumption level will be an increasing function of her past income growth *rate*. This implication is independent of whether or not individuals know their true income growth rate.

Finally, it is also easy to see that the level of consumption is also informative about how much prior information individuals have about their own β^i within the HIP framework (question 3 raised in the introduction). To see this, consider the next figure (3) which is a slight variation of the previous one. Here, both individuals are assumed to have observed the same path of income growth up to period 3 even though their true β^i are different. (This is possible since there are many stochastic shocks to the income process over time (coming from η_t), and the contribution of β^i to income is quite small). In this case, under the HIP model, if individuals have no private prior prior information abut their own true β^i (which will be the case when $\lambda = 0$) then both individuals should have the same consumption level. The more prior information each individual has about her true β^i the higher will be the consumption of the first individual compared to the second. Therefore, an auxiliary model can capture this relationship by focusing on the following dynamic relationship:

> *Prediction 3:* if $\lambda > 0$, then controlling for *past* income growth (as well as the current income level and past average consumption level) the consumption level of an individual will be increasing in her *future* income growth as well. This is because in this case the individual has more information about her true β^i than is known to the econometrician and what is revealed by her past income growth.

These three examples illustrate how one can use the structural relationships such as (15) and (16) that hold true exactly in a somewhat simplified version of the economic model of interest in order to come up with an auxiliary model. Indirect inference allows one to think in terms of these rich dynamic relationships instead of a set of moments (covariances, etc.). Below we are going to write a parsimonious auxiliary model that will capture these dynamic relationships to identify HIP from RIP and will also determine the degree of prior information (or equivalently, uncertainty) individuals face upon entering the labor market in the case of the HIP process.

3.2 A Parsimonious and Feasible Auxiliary Model

As shown above, the HIP model implies:

$$\Delta C_t = \Pi\left(\lambda, P_{t|t}, r, \rho, t, R, T\right) \times \left(Y_t^i - \left(\widehat{\beta}^i_{t|t-1} t + \widehat{z}^i_{t|t-1}\right)\right), \tag{17}$$

where $\Pi\left(\lambda, P_{t|t}, r, \rho, t, R, T\right) \equiv \Phi^r_{t,T}\left(A_t/X_t\right) + \Psi^{r,\rho}_{T-t}\left(B_t/X_t\right)$; the dependence of Π on λ and $P_{t|t}$, can be seen from the formulas for A_t and B_t. However, since $\widehat{\beta}^i_{t|t-1}$ and $\widehat{z}^i_{t|t-1}$ are unobserved by the econometrician (because they depend on all past income realizations as well as on each individuals' unobserved prior beliefs), this regression is not feasible as an auxiliary model. Moreover, this relationship was derived assuming a simplified HIP income process, quadratic utility, no borrowing constraints, and no retirement period, none of which is true in the life-cycle model we would like to estimate. Fortunately, as mentioned earlier, these issues do not represent a problem for the consistency of the estimates of the structural parameters that we are interested in.

We begin by approximating the relationship in (17) with the following feasible regression:

$$\begin{aligned} c_t = a_0 + a_1 y_t + a_2 y_{t-1} + a_3 y_{t-2} + a_4 y_{t+1} + a_5 y_{t+2} + a_6 \overline{y}_{1,t-3} + a_7 \overline{y}_{t+3,T} \\ + a_8 \Delta y_{1,t-3} + a_9 \Delta y_{t+3,T} + a_{10} c_{t-1} + a_{11} c_{t-2} + a_{12} c_{t+1} + a_{13} c_{t+2} + \epsilon_t \end{aligned}$$

where c_t is the logarithm of consumption; y denotes the logarithm of labor income; $\overline{y}_{a,b}$ denotes the average of log income from time a to b; and similarly $\Delta y_{a,b}$ denotes the average growth rate of log income from time a to b. Notice that we use the logarithm of variables rather than the level; since the utility function is CRRA and income is lognormal this seems to be a more natural specification. This regression captures the three predictions made by the HIP and RIP models discussed above by adding the past and future income growth rate as well as past and future income levels. Leads and lags of consumption are also added to capture the dynamics of consumption around the current date. To complete the auxiliary model, we add a second equation with y_t as the dependent variable, and use all the regressors above involving the leads and lags of income (of course current income is left out) as left hand side variables for a total of nine regressors in the second equation. Finally, we divide the population into two age groups—those between 25 and 38 years of age, and those between 39 and 55 years of age—and allow the coefficients of the auxiliary model to vary across the two groups.[12] For each age group, the auxiliary model has 23 regression coefficients (14 in the first equation and 9 in the second) and two residual variances (one for each equation in the auxiliary model)[13] for a total of 25 parameters. With two age groups, this yields a total of 50 reduced form parameters that determine the likelihood of the auxiliary model.

To implement the indirect inference estimator, we choose the values of the structural parameters so that the (approximate) likelihood of the observed data (as defined by the auxiliary model) is as large as possible. That is, given a set of structural parameters, we simulate data from the model, use this data to estimate the auxiliary model parameters, and evaluate the likelihood defined by the auxiliary model at these parameters. We then vary the structural parameters so as to maximize this likelihood. Viewed from another perspective, we are simply minimizing the difference between the (log) likelihood evaluated at two sets of auxiliary model parameters: the estimates in the observed data and the estimates in the simulated data (given a set of structural parameters). The advantage of this approach over other approaches to indirect inference (such as efficient method of moments or minimizing a quadratic form in the difference between the observed and simulated auxiliary model parameters) is that it does not require the estimation of an optimal weighting matrix. It is, however, less efficient asymptotically than the other two approaches, though this inefficiency is small when the auxiliary model is close to being correctly specified (and vanishes in the case of correct specification). The Monte Carlo analysis that we report later below indicates however that for the sample size used in this study, the present method yields excellent results and creates much less small sample bias (almost none) than alternative methods (such as minimizing a quadratic objective with an identity weighting matrix as often done in the literature).

[12] Although, the auxiliary model would correspond to the structural equation in (17) more closely if the coefficients were varying freely with age, this would increase the number of parameters in the auxiliary model substantially. Our experience is that the small sample performance of the estimator is better when the auxiliary model is more parsimonious, and therefore we opt for the specification here.

[13] Notice that the covariance between the residuals of the two regressions will be exactly zero, since y_t appears on the left hand side of the second equation and the right hand side of the first.

4 The Data

This section discusses the data used in the empirical analysis and describes how we construct the panel of imputed household consumption by combining data from the Panel Study of Income Dynamics (PSID) and the Consumer Expenditure Survey (CE). The unit of analysis in this paper is a married household—so both income and consumption are measured at the household level. The 1980-1992 panel from the CE survey we use in this paper is the same as the one used in Blundell, Pistaferri, and Preston (2008) and has been generously provided to us by the authors. For reasons that will become clear below, we combine this panel with the 1972-73 waves of the CE survey (which is available from the Bureau of Labor Statistics). For this earlier sample, we largely follow the selection criteria used in Blundell, Pistaferri, and Preston (2008) to make the two samples of the CE data consistent with each other. Turning to the PSID, we restrict attention to households that are in the core sample (i.e., we exclude households from the non-random SEO and Latino samples), whose head is between the ages of 25 and 55 (inclusive), has non-missing data on food expenditures and husband and spouse's labor income. A more complete description of the sample selection criteria we use for each data set along with further details of the imputation procedure are contained in Appendix A.

4.1 Constructing a Panel of (Imputed) Consumption

An important impediment to the previous efforts to use consumption data has been the lack of a sufficiently long panel on consumption expenditures. PSID has a long panel dimension but covers limited categories of consumption whereas the CE survey has detailed expenditures over a short period of time (four quarters). As a result, most previous work has either used food expenditures as a measure of non-durable consumption (available in PSID), or resorted to using repeated cross-sections from CE under additional assumptions.

In a recent paper, Blundell, Pistaferri, and Preston (2006) (hereafter, BPP) develop a structural imputation method, which imputes consumption expenditures in PSID using information from CE. The basic approach involves estimating a demand system for food consumption as a function of nondurable expenditures, a wide set of demographic variables, and relative prices as well as the interaction of nondurable expenditures with all these variables. The key condition is that all the variables in the demand system must be available in the CE data set, and all but non-durable expenditures must be available in PSID. One then estimates this demand system from CE, and as long as the demand system is monotonic in nondurable expenditures, one can invert it to obtain a panel of imputed consumption in the PSID.

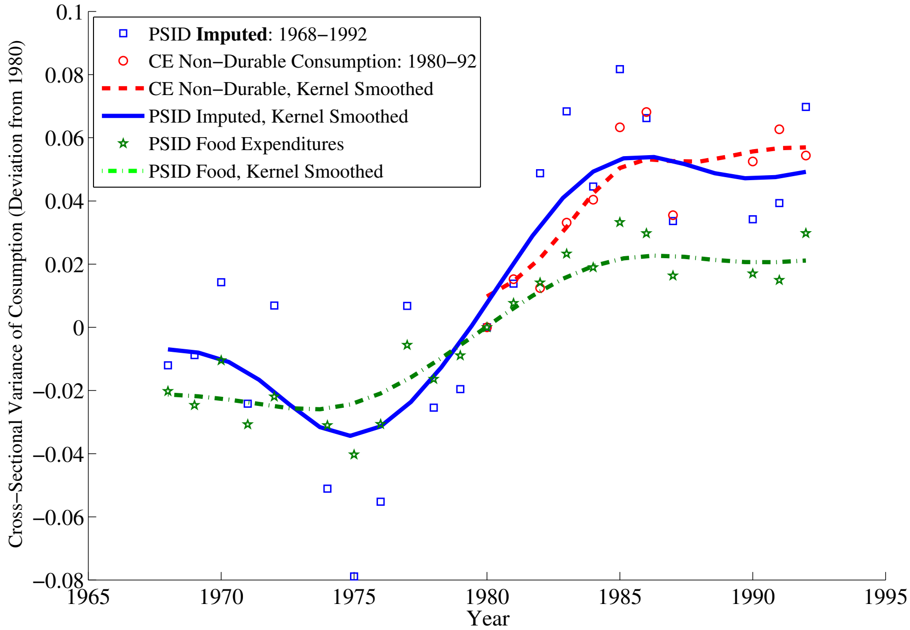
BPP implement this method to obtain imputed consumption in PSID for the period 1980 to 1992, and show that several statistics calculated using the imputed measure compare quite well to their counterparts from CE consumption data. In this paper, we modify and extend the method proposed by these authors as follows. First, these authors include time dummies interacted with

Table 1: Instrumental Variables Estimation of Demand for Food in the CE

Variable	Estimate	Variable	Estimate
$\ln(c)$	0.798^{***}	$\ln(c) \times I\{11\% \leq \Delta \log p_{fuel}\}$	0.00386^{*}
	(26.80)		(1.83)
$\ln(c) \times age \times I\{age < 37\}$	0.00036^{***}	$\ln(c) \times (year - 1980)$	-0.00057
	(3.38)		(-0.68)
$\ln(c) \times age \times I\{37 \leq age < 47\}$	0.00048^{***}	One child	0.149
	(5.45)		(1.16)
$\ln(c) \times age \times I\{47 \leq age < 56\}$	0.00042^{***}	Two children	0.564^{***}
	(5.75)		(3.98)
$\ln(c) \times age \times I\{56 \leq age\}$	0.00037^{***}	Three children+	1.203^{***}
	(6.08)		(8.23)
$\ln(c) \times$ High school dropout	-0.129^{***}	High school dropout	1.207^{***}
	(-7.57)		(7.61)
$\ln(c) \times$ High school graduate	-0.043^{***}	High school graduate	0.417^{***}
	(-2.78)		(2.90)
$\ln(c) \times$ One child	-0.014	Northeast	0.0587^{***}
	(-1.01)		(10.36)
$\ln(c) \times$ Two children	-0.055^{***}	Midwest	0.0293^{***}
	(-3.68)		(5.23)
$\ln(c) \times$ Three children+	-0.123^{***}	South	-0.0031
	(-7.92)		(-0.63)
$\ln(c) \times I\{5\% \leq \Delta \log p_{food} < 8\%\}$	0.00096	Family size	0.0509^{***}
	(1.01)		(16.20)
$\ln(c) \times I\{8\% \leq \Delta \log p_{food} < 11\%\}$	0.00858^{***}	$\ln p_{food}$	0.581^{**}
	(4.25)		(2.28)
$\ln(c) \times I\{11\% \leq \Delta \log p_{food}\}$	-0.00091	$\ln p_{fuel}$	-0.117
	(-0.39)		(-0.97)
$\ln(c) \times I\{5\% \leq \Delta \log p_{fuel} < 8\%\}$	0.00074	White	0.0824^{***}
	(0.66)		(11.38)
$\ln(c) \times I\{8\% \leq \Delta \log p_{fuel} < 11\%\}$	0.00091	Constant	-1.822^{***}
	(0.53)		(-2.65)
		Observations	21864

We pool the data from the 1972-73 waves of the CE with the 1980-92 waves. We instrument log food expenditures (and its interactions) with the cohort-education-year specific average of the log husband's and wife's hourly wage rates (and their interactions with age, education, and inflation dummies and a time trend). The t-statistics are contained in parentheses. The lowest value of Shea's partial R^2 for instrument relevance is 0.086, and the p-value of the F-test on the excluded instruments is smaller than 0.001 for all instruments.

Figure 4: Cross-sectional Variance of Log Consumption in CEX and Imputed PSID Data: 1968-1992.



nondurable expenditures in the demand system to allow for the budget elasticity of food demand to change over time, which they find to be important for the accuracy of the imputation procedure. However, CE is not available on a continuous basis before 1980, whereas we would like to use the entire length of PSID from 1968 to 1992, making the use of time dummies impossible. To circumvent this problem, we replace the time dummies with other variables that are available throughout our sample period—specifically, the interaction of nondurable expenditures with food and fuel inflation rate. The inclusion of these inflation variables is motivated by the observation that the pattern of time dummies estimated by BPP after 1980 is quite similar to the behavior of these inflation variables during the same period.

A second important element in our imputation is the use of CE data before 1980. In particular, CE data are also available in 1972 and 1973, and in fact these cross-sections contain a much larger number of households than the waves after 1980.[14] The data in this earlier period also appear to be of superior quality in certain respects compared to those from subsequent waves. For example, as shown by Slesnick (1992), when one aggregates several sub-components of consumption expenditures in the CE, they come significantly closer to their counterparts in the National Income and Product Accounts than the CE waves after 1980.[15] The use of this earlier data provides, in

[14]The sample size is around 9500 units in 1972-73 surveys, but range from 4000-6000 units in the waves after 1980. There are also some differences in the survey design in the earlier CE—such as the non-rotating nature of the sample in the 1972 and 1973 panels—but these differences do not appear consequential for our purposes. See Johnston and Shipp (1997) for a more detailed comparison of different waves of the CE survey over time.

[15]For example, in 1973 total expenditures measured by the CE is 90 percent of personal consumption expenditures

some sense, an anchor point for the procedure in the 1970s that improves the overall quality of imputation as discussed below. Finally, instead of controlling for life-cycle changes in the demand structure using a polynomial in age (as done by BPP), we use a piecewise linear function of age with four segments, which provides more flexibility. This simple change improves the life-cycle profiles of mean consumption and the variance of consumption rather significantly. With these modifications, we obtain an imputed consumption measure that has a fairly good fit to the corresponding statistics in the CE data.

Since food and non-food consumption are jointly determined, some of the right hand side variables in the demand system are endogenous. In addition, nondurable expenditures are likely to suffer from measurement error (as is the case in most survey data sets), which necessitates an instrumental variables approach. We instrument log nondurable expenditures (as well as its interaction with demographics and prices) with the cohort-year-education specific average of the log of the husband's hourly wage and the cohort-year-education specific average of the log of the wife's hourly wage (as well as their interaction with the demographics and prices). Table 1 reports the results from the estimation of the demand system using the CE data. Several terms that include the log of nondurable expenditures are significant as well as several of the demographic and price variables. Most of the estimated coefficients have the expected sign. We invert this equation to obtain the imputed measure of household non-durable consumption expenditures.

Figure 4 plots the cross-sectional variance of log consumption over time. BPP used this figure as the main target to evaluate the satisfactoriness of their imputation procedure. In the figure, the (red) circles mark the CE data whereas the (blue) squares show the imputed consumption in PSID. Similarly, the dashed (red) line and the solid (blue) line show the corresponding "smoothed" series obtained using the Nadaraya-Watson kernel regression (with a Gaussian kernel). The imputed consumption series tracks the CE data fairly well, showing an overall rise in consumption inequality of 6–7 log points between 1980 and 1986, followed by a drop from 1986 to 1987 and not much change after that date. The dash-dot line shows that if one simply were to use food expenditures in PSID instead, the overall pattern would remain largely intact, but the movements would be *quantitatively* muted compared to the data: the rise in consumption inequality would be understated by more than half by 1986 and by as much as two-thirds by 1991.

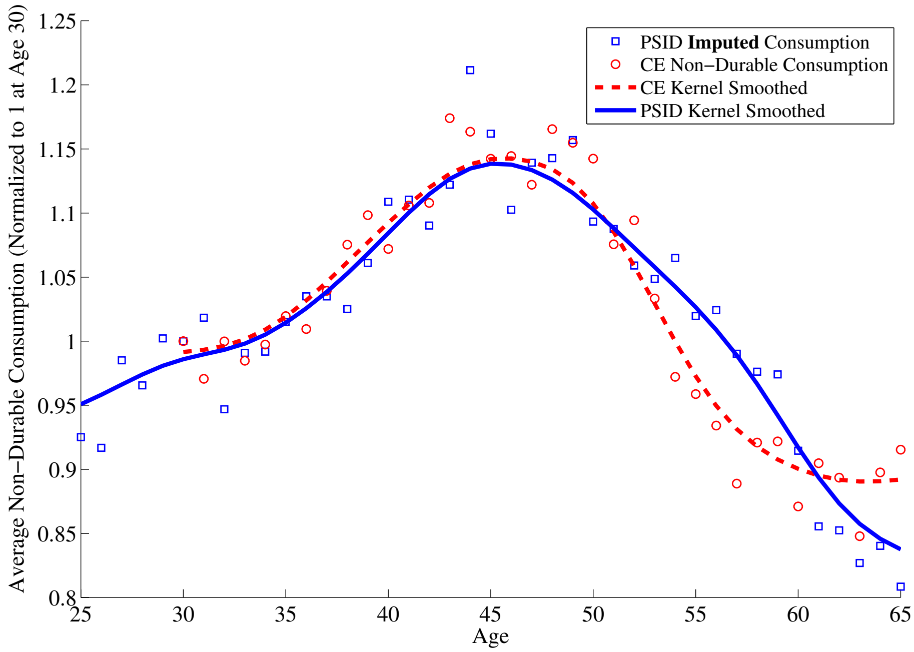
We next evaluate the quality of the imputation procedure in two other dimensions that are especially important for the estimation exercise. First, figure 5 plots the average life-cycle profile of log consumption implied by the CE data (marked with circles) as well as the counterpart generated by the imputed data (marked with squares).[16] Again, to reduce the noise in the data, the figure also plots kernel-smoothed versions of each series. The two graphs overlap remarkably well, especially up to early age 50.[17] Second, figure 6 plots the within-cohort variance of log consumption over the

as measured by NIPA, whereas this fraction is consistently below 80 percent after 1980 and drops to as low as 75 percent in 1987. Similarly, consumer services in the CE accounts for 93 percent of the same category in NIPA in 1973, but drops to only 66 percent in 1989.

[16]The lifecycle profiles are obtained by controlling for cohort effects as described in Guvenen (2009).

[17]If we do not use the 1972-73 CE in the imputation procedure the average profile of imputed consumption would

Figure 5: Life-cycle Profile of Average Consumption in CEX and Imputed PSID Data

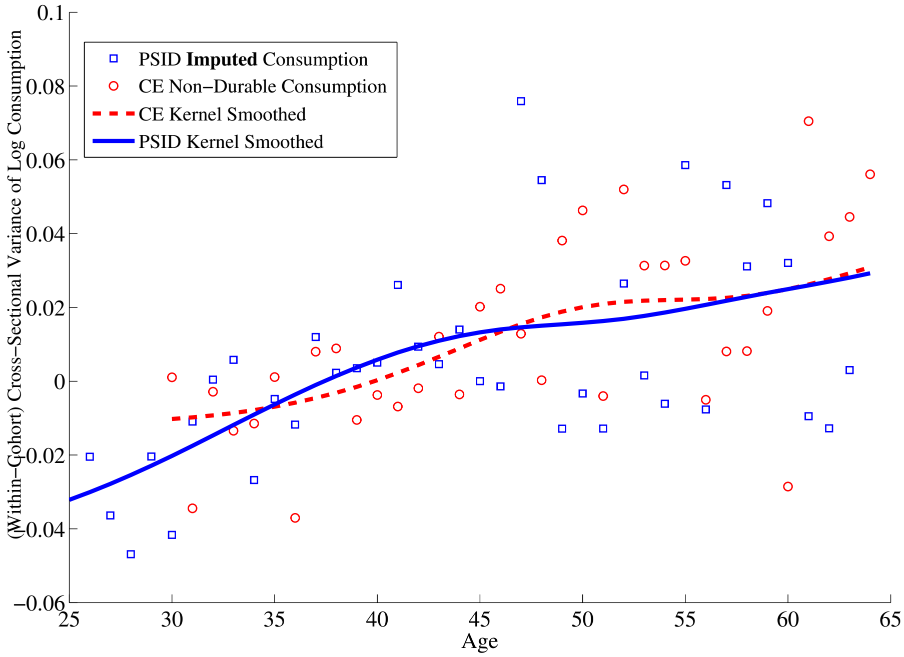


lifecycle along with the smoothed series. Both in the CE and with the imputed PSID data, the variance rises between age 25 and 65, although the total rise is rather small—about 5 log points. The finding of a small rise in within-cohort consumption inequality contrasts with earlier papers that have focused on the CE data over the period from 1980 to 1990, such as Deaton and Paxson (1994) and Guvenen (2007), but is consistent with more recent papers that have used samples extending to late 1990s (cf., Heathcote, Perri, and Violante (2009)). This finding will be important in understanding some of the estimates that we obtain in the structural estimation below.

Finally, it is also useful to provide some evidence on the quality of the imputation by testing its out-of-sample predictive ability at the household level. To this end, we split the CE sample used in the imputation above into two randomly drawn subsamples (each containing exactly half of the observations in each year). We use the first subsample to estimate the food demand system as above, which we then use to impute the non-durable consumption of the second subsample (control group). To eliminate sampling variation that results from the randomness of each subsample, we repeat this exercise 200 times. The results below refer to the average of these 200 replications. Comparing the actual non-durable expenditures of these households to that implied by the imputation is informative about the quality of the imputation. Figure 7 plots the actual consumption of the control group against the imputed one for each household (for the simulation with the median regression slope). The imputed consumption data forms a cloud that align very well with the 45-degree line. In fact, a linear regression of imputed consumption on the actual one

rise by 51 percent between ages 25 and 45 instead of the 22 percent rise in the baseline imputation and would therefore vastly overestimate the corresponding rise in the CE data shown in figure 5.

Figure 6: Life-cycle Profile of Consumption Variance in CEX and Imputed PSID Data



yields an average slope coefficient of 0.996 and a constant term of 0.25. The average R^2 of the regression is 0.67, implying that the imputed consumption has a correlation of 0.81 with the actual consumption at household-level.[18]

The fact that the slope coefficient is almost equal to 1 is important: a slope above 1 (with a positive intercept) would indicate that the imputation systematically overstates the variance of true consumption, which would in turn overstate the response of consumption to income shocks, thereby resulting in an overestimation of the size of income shocks. The opposite problem would arise if the slope coefficient was below 1. Furthermore, when a quadratic term is added to the regression of imputed consumption on actual consumption, it almost always comes out as insignificant. This implies that the imputation procedure does not result in systematic under- or over-prediction at different points in the distribution, which would again be problematic. (We have also repeated the same exercise by only using the 1980 to 1992 waves of the CEX. The results were very similar: the average slope was 1.036, the constant term was 0.12, and the R^2 was unchanged from before, at 0.67.)

As a final, and rather strict, test to detect whether systematic patterns exist in the imputation error, we regressed the difference between imputed and actual consumption for each individual (i.e., imputation error) on household characteristics including dummies for each age group, education dummies, family size, region dummies, number of children dummies and food and fuel prices. The median R^2 of this regression was 0.002 (and there was at most one variable that was significant

[18]Across simulations, the slope coefficient in the regression ranges from 0.978 to 1.020, and the R^2 ranges from 0.644 to 0.691.

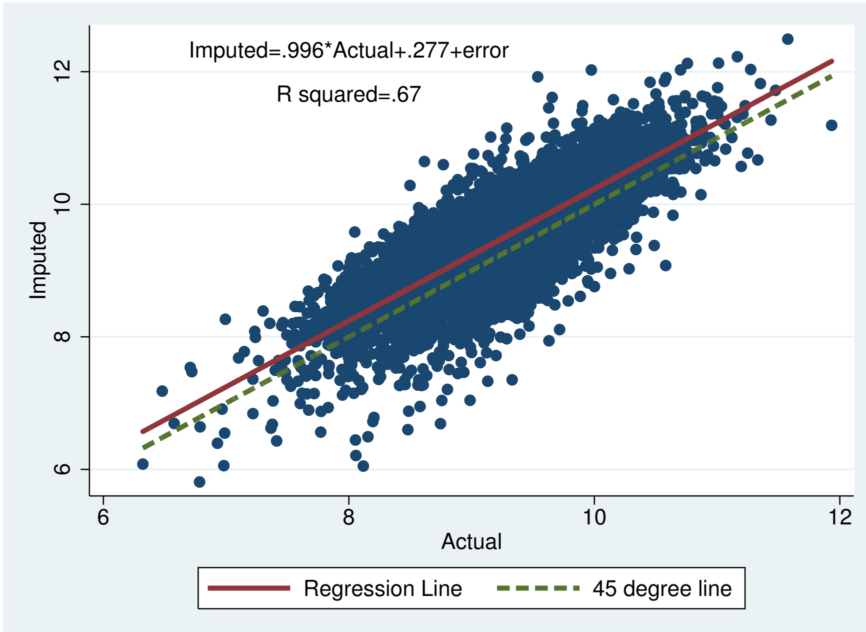


Figure 7: **Out of Sample Predictive Power of the Imputation Method in the CEX.** This plot is obtained by estimating the IV food demand system on a randomly chosen half of the CEX sample, and then imputing the consumption for the other half (control group). The figure plots the actual consumption of the control group versus their imputed consumption The average regression slope is 0.996, the average constant is 0.24, and average R^2 is 0.67 over 50 repetitions.

at 5 percent level in any given simulation) indicating no evidence of systematic imputation errors by demographic groups. Overall, we conclude that the imputation procedure works fairly well and does not result in any systematic over- or under-prediction of actual consumption, which is reassuring for the estimation exercise.

Measure of Household Labor Income. In PSID, households report their total taxable income which includes labor income, transfers and financial income of all the members in the household. The measure of labor income we use subtracts financial income from this measure, and therefore, includes the labor income of the head and spouse as well as several categories of transfer income (unemployment benefits, social security income, pension income, worker's compensation, welfare payments, child support, and financial help from relatives are the main components). PSID also reports estimated total taxes for all households until 1991. For 1992 and 1993 we use the National Bureau of Economic Research's (NBER) TAXSIM software to estimate taxes for each household. Since our income measure excludes asset income, for each year we regress the total tax payment on the asset income and labor income of each household to back out the labor portion of the taxes paid in each year. We then subtract this estimated labor income tax from household income above to obtain the household after-tax labor income measure used in the analysis below.

Converting the Data to Per-adult Equivalent Units. We adjust both the imputed consumption and income measures for demographic differences across households since such differences have no counterpart in our model. This is accomplished by regressing each variable on family size, an education dummy, a race dummy, a number of children dummy, a region dummy, a dummy indicating whether the head is employed, the number of earners in the household, a dummy indicating residence in a large city, and a set of cohort dummies.[19] We then use the residuals of these regressions—which are interpreted as consumption and income per-adult equivalent—in the analysis below. In the estimation, we use 2235 households and a total of 26441 household-year observations on labor income and household consumption.

5 Empirical Results

In this section, we apply the proposed methodology to the estimation of the consumption-savings model with Bayesian learning. To demonstrate the ability of this estimation method to uncover the true structural parameter vector in spite of (very) incomplete individual histories, substantial measurement error, potentially binding borrowing constraints, etc., we begin by conducting a Monte Carlo study using 150 "observed" data sets drawn from the true data generating process. We then present results with actual PSID data.

Some parameter values are set before the model is estimated. Specifically, the net annual interest rate is set to 3%. The number of years in the working life is set to 41, and the number of years in the retirement period is set to 15. Individuals have isoelastic utility with coefficient of relative risk aversion equal to 2. The borrowing constraint is set to 25% of the average annual labor income in the economy (which is consistent with sure repayment of debts at the end of life).[20] We next describe the specifics of the pension system.

Social Security System The pension system in the model mimics the US Social Security system, with one notable difference. In the actual US pension system, retirement income is tied to individuals' average labor earnings during the working years (denote it $\overline{Y}^i$).[21] However, adopting this exact structure here would add another state variable—$\overline{Y}^i$—to the dynamic problem above, increasing the already high computational burden of the estimation significantly. So, instead, we adopt the same functional form used in the US system for the $\Phi(\cdot)$ function, but instead of using $\overline{Y}^i$, we use the *predicted* average earnings given Y_R^i (the worker's earnings at the retirement age). This is done by first running the cross-sectional regression: $\overline{Y}^i = k_0 + k_1 Y_R^i$, and then using the predicted average earnings implied by this regression, which we denote by $\hat{Y}(Y_R^i)$. This structure

[19] Each cohort is defined by 5-year bands based on the birth year of each individual, e.g., those born between 1951 and 1955, 1956 and 1960, etc.

[20] In ongoing work, we are also estimating the relative risk aversion coefficient as well as the borrowing constraint. These results will be incorporated into the next revision of the paper.

[21] More precisely, the average is taken over the 35 working years with the highest earnings.

Table 2: Estimating the Full Consumption-Savings Model: Monte Carlo Results

	True value	Mean estimate	Std. Dev.
ρ	0.670	0.669	0.034
σ_η	0.190	0.191	0.009
σ_ε	0.150	0.150	0.015
$\sigma_\beta(\times 100)$	2.650	2.652	0.121
σ_α	0.490	0.493	0.031
λ	0.500	0.504	0.054
σ_{u^y}	0.200	0.199	0.003
σ_{u^c}	0.200	0.199	0.008

Note: Statistics are based on 150 replications

does not add a state variable but recognizes the empirical relationship between average earnings and pre-retirement earnings implied by each stochastic process. Now, let $\overline{Y}$ denote the economy-wide average lifetime labor income and let $\widetilde{Y}_R^i \equiv \hat{Y}(Y_R^i)/\overline{Y}$. Then the pension income is given by

$$Y^i = \Phi\left(Y_R^i; \overline{Y}\right) = \overline{Y} \times \begin{cases} 0.9\widetilde{Y}_R^i & \widetilde{Y}_R^i \leq 0.3 \\ 0.27 + 0.32(\widetilde{Y}_R^i - 0.3) & 0.3 < \widetilde{Y}_R^i \leq 2 \\ 0.81 + 0.15(\widetilde{Y}_R^i - 2) & 2\overline{Y} < \widetilde{Y}_R^i \leq 4.1 \\ 1.13 & 4.1\overline{Y} \leq \widetilde{Y}_R^i \end{cases}$$

5.1 Monte Carlo Analysis

We now conduct a Monte Carlo study using the setup above, which is the same setup that will be used in the estimation with the real data later below. So this Monte Carlo analysis should provide a good test of the performance of the proposed estimation methodology.

The missing observations in the Monte Carlo study are chosen to be exactly the same as in the observed data. We include only households with at least five observations between the ages of 25 and 55 (of the head), for a total of 2,235 individuals with an average of 12 observations on each (for a total of 26,441 household-year observations). We next add measurement error to both consumption and income:

$$y_t^{i,*} = y_t^i + u_t^{i,y},$$
$$c_t^{i,*} = c_t^i + \overline{u}^{i,c} + u_t^{i,c}$$

where $y_t^{i,*}$ and $c_t^{i,*}$ are measured income and consumption of household i, respectively, and $u_t^{i,y}$ and $u_t^{i,c}$ are i.i.d random variables with zero mean over time.[22] Notice that we also added a

[22]The assumption that the distribution of the measurement error $u_t^{i,c}$ is identical over time is somewhat problematic

second term to consumption, $\overline{u}^{i,c}$, which is an individual fixed measurement error with potentially non-zero mean in the cross-section. This fixed effect is needed for two reasons. First, and most importantly, recall that we regress both income and consumption on a full set of demographics to convert these variables into per-adult equivalent terms. However, one effect of this adjustment is that it introduces level differences between consumption and income, the magnitude of which varies by household. This fixed effect captures such differences.[23] Second, the model described above abstracts from initial wealth differences across households, which clearly exist in the data. These differences in wealth would also drive a household-specific wedge between the levels of income and consumption. The fixed effect is also a simple way to capture these differences in initial wealth levels. However, because all households in the simulated data have the same demographics and zero initial wealth, this fixed effect is redundant in the Monte Carlo analysis. Therefore, we set it to zero until we get to the estimation with real data below.

Incomplete histories are handled by "filling in" missing values in a reasonable way: basically, at each age that an individual has a valid income data point, we find the percentile ranking of this observation in the income distribution (at that age) in our sample. We then take the average of the percentile rankings for this individual over all the ages that she has a valid observation. Then for each missing income observation of this individual, we impute the income level corresponding to her average percentile ranking given the income distribution in our sample for that age. We apply the same procedure to fill in missing consumption data. We construct the past growth rate for age t in the auxiliary model by taking the difference between the latest valid observation before t and the first valid observation for the individual in data set and dividing this difference by the number of years between the two ages. The future growth rate at a given age is constructed analogously. If either variable cannot be constructed for a given age we use the average growth rate of that variable in the population instead. We use exactly the same procedure in both the simulated and observed data. The (approximate) likelihood, however, includes contributions only from those time periods in which the left-hand side variables are observed (i.e., not missing). Below, we consider alternative methods for filling in missing data and check the sensitivity of the results to the method used.

The results are contained in Table 2. The "true values" for the parameters are set to the estimates obtained using PSID income data alone. The initial values of the parameters are set

to the extent that this term contains the imputation error, because the imputation procedure itself would introduce time variation in the variance of this term (as BPP also observe). We currently do not consider time-variation in these variances, but will introduce it in the next version of the paper.

[23] An alternative way to understand this point is to observe that the level of household consumption and labor income have a certain relationship in the theoretical model, which is not preserved by the scaling introduced by the conversion to adult-equivalent units described above. For example, consider two households with the same income and consumption, but suppose that the first household has more children than the second, and both households have the same number of earners. Converting the variables to per-adult equivalent units will result in the first household having a lower consumption than the second one despite having the same income (since children consume but typically do not earn income). A similar issue arises between households with different number of earners given a certain level of total income and consumption. Since we do not explicitly account for such demographic differences in our model—which would complicate the analysis tremendously—we account for such differences in levels using the fixed effect as modeled here.

Table 3: Structural Estimation of the Consumption-Savings Model Using Real Data

Data:		
	Estimate	Std Error
ρ	0.756	0.029
σ_η	0.189	0.005
σ_ε	0.019	0.027
$\sigma_\beta(\times 100)$	1.785	0.392
σ_α	0.383	0.029
$cor_{\alpha\beta}$	-0.178	0.074
λ	0.191	0.124
σ_{u^y}	0.145	0.010
σ_{u^c}	0.355	0.002
$\sigma_{\overline{u}^c}$	0.434	0.009
$\mu_{\overline{u}^c}$	0.042	0.014
δ	0.958	0.003

randomly to ±20% of the true values. Each Monte Carlo run takes about 30 minutes on a state-of-the-art workstation. Clearly, the estimation method works well: bias is virtually absent and standard deviations are small. Although it is difficult, if not impossible, to prove identification in this setup, the results suggest strongly that local identification near the true parameter vector does indeed hold. These results are encouraging and suggest strongly that the proposed methodology is a feasible and practical method for estimating structural consumption-saving models with missing data, binding borrowing constraints, and multiple sources of heterogeneity.

5.2 Results Using PSID Data

We now estimate the lifecycle model using the PSID household after-tax labor income data and the imputed consumption data. In addition to the parameters in the Monte Carlo study, we now also estimate the mean and standard deviation of $\overline{u}^{i,c}$ (the consumption fixed measurement error).

Table 3 reports the results. The AR(1) process has an annual persistence of 0.79—which is estimated quite precisely—and an innovation standard deviation of about 19% (also estimated very precisely). Therefore, the joint dynamics of consumption and income data do not lend support to permanent shocks as a reasonable representation of the typical income shock. With consumption data, in principle, we can tell apart transitory shocks from measurement error in income, since consumption should respond to the former but not to the latter. In practice, however, because the response of consumption to transitory shocks is proportional to its annuitized value—which is small—this response is rather weak and identification is a problem empirically. In this framework, however, borrowing constraints are binding for a significant fraction of households—no less than 15 percent of households younger than 35 years of age. As a result, these households' consumption move one for one with transitory shocks allowing us to distinguish these shocks from pure measurement error. We estimate the standard deviation of transitory shocks to be only 2% and the standard deviation of measurement error in income to be about 14.5% annually.[24]

The standard deviation of income growth rates, a key magnitude of interest, is estimated to be 1.78 percent, which is substantial. To understand the economic significance of this estimate, note that assuming a 1 percent average income growth rate each year over the lifecycle, an individual who is one standard deviation above the mean will earn 2.99 times the median income and 4.09 times the income of an individual who is one standard deviation below the mean. Of course, not all this heterogeneity represents uncertainty on the part of the individuals, since each individual has some prior information about her true β^i by the time she enters the labor market. The parameter λ—which measures the degree of this prior uncertainty—is estimated to be 0.191, which reveals only a small amount of prior uncertainty regarding individuals' growth rate. Alternatively stated, individuals have significant amount of information about their future income prospects that is unobserved by the econometrician.

The classical measurement error in consumption has a standard deviation of 35.5 percent and includes the noise introduced by the imputation method. Furthermore, the fixed effect in measured consumption (that results from the conversion to per-adult equivalent terms) has a standard deviation of 43.4 percent (with a mean of 0.04) and both components are estimated very precisely.[25] Finally, the estimated value of the time discount factor is 0.958 with a very small standard error. Given the annual net interest rate of 3%, this estimate implies that individuals are impatient in the sense of Deaton (1991). Interestingly, this estimate is also remarkably close to the value estimated in Gourinchas and Parker (2002), in which the estimated value ranged between 0.957 and 0.960. One difference is that the estimate here has a much smaller standard error (0.003 here versus 0.010 to 0.018 in that paper), which may be due to the fact that auxiliary model may contain more precise information than the hump-shaped consumption profile targeted in that paper.

It is also useful to get a sense about how well the estimated model fits the data. First, table 4 displays the coefficients of the auxiliary model estimated with real data and simulated data. Recall that the auxiliary model has a total of 50 coefficients and the consumption model has 13 structural parameters. Panel A shows the coefficients of the income equation. A quick visual inspection shows that many of the coefficients implied by the structural model line up well with their empirical counterparts.[26] Panel B shows the coefficients for the consumption equation, which also shows a reasonable fit, but to a lesser extent than the income equation. This may not be

[24]The very small estimated value of σ_ε raises the question of whether this parameter cannot be disentangled from i.i.d measurement error causing it to drift toward the boundary. While this possibility cannot not be dismissed out of hand, recall that in the Monte Carlo study these two errors were identified very well (see table 2). Furthermore, when we re-estimated the model with real data by fixing σ_{u^y} at different values (0.07, 0.10, etc.) the estimate of σ_ε turned out to be higher (since the estimator attempts to fit the transitory movements in income), but the objective value was significantly higher than what was achieved by the unrestricted estimates presented in table 3.

[25]We found this component to be important for the overall estimation—failing to include this term results in implausible estimates for many parameter values, as the minimization routine struggles to make sense of the fixed level differences in the data that has no counterpart in the model.

[26]The overidentifying restrictions implied by the auxiliary model can be tested more formally using the test

Table 4: Coefficients of the Auxiliary Model: Simulated vs US Data

	y_{t-1}	y_{t-2}	y_{t+1}	y_{t+2}	$\overline{y}_{1,t-3}$	$\overline{y}_{t+3,T}$	$\Delta y_{1,t-3}$	$\Delta y_{t+3,T}$	c_{t-1}	c_{t-2}	c_{t+1}	c_{t+2}	y_t
Panel A: Income Equation													
Data	0.306	0.029	0.324	0.052	-0.192	0.396	0.166	-0.115					
Model	0.296	0.035	0.320	0.057	-0.171	0.382	0.212	-0.135					
Data	0.414	0.087	0.353	0.076	-0.070	0.132	0.216	-0.055					
Model	0.398	0.095	0.373	0.091	-0.057	0.084	0.172	-0.038					
Panel B: Consumption Equation													
Data	0.033	-0.057	-0.035	-0.030	-0.024	-0.020	0.092	-0.013	0.249	0.255	0.181	0.175	0.206
Model	0.036	-0.075	-0.021	-0.067	-0.045	0.012	0.017	-0.007	0.209	0.200	0.256	0.224	0.255
Data	0.059	-0.050	-0.020	-0.059	-0.060	0.005	0.142	0.000	0.268	0.254	0.180	0.189	0.184
Model	0.024	-0.021	-0.023	-0.046	0.009	-0.025	0.025	0.004	0.198	0.212	0.221	0.259	0.187
					Age group 1			Age group 2					
Variance of equation :					Inc.	Cons.		Inc.	Cons.				
Data					0.2184	0.3930		0.2348	0.3768				
Model					0.2223	0.3874		0.2317	0.3858				

surprising as this equation relies on all the machinery of the consumption-savings model and is not as straightforward as the income equation.

We next evaluate the fit of the model to the data along three dimensions that have received much attention in the incomplete markets literature. The first two are the change in the within-cohort variance of log income and log consumption. The well-known empirical fact that within-cohort dispersion of labor income rises significantly with age is displayed as the dashed line marked with diamonds in figure 8. The solid line marked with squares shows the corresponding graph implied by the estimated model, which shows a remarkably good fit to the data. Notice that the auxiliary model we use does not have terms that appear to capture the variance of income by age in an obvious fashion. But it seems that the rich structure contained in those two equations incorporate enough information to generate this good fit. Turning to the variance of log consumption, the line marked with triangles shows the empirical counterpart whereas the line with circles plots the model counterpart. While the rise in the variance implied by the model exceeds that in the data by about 7 log points, the total rise of 13 log points is rather small compared to earlier papers, such as in Storesletten, Telmer, and Yaron (2004) and Guvenen (2007) who generated rises as large as 35 log points. This is due to the fact that, as explained above, these earlier papers assumed a larger amount of income uncertainty to match the large rise in consumption inequality found by Deaton and Paxson (1994) who focused on the CE data from 1980 to 1990. The small rise in consumption inequality found here also explains why our structural estimation found a small estimate of λ, implying a smaller overall amount of uncertainty. Again, despite not having any terms that captures the rise in the variance of consumption explicitly the auxiliary model attempts to be consistent with the small rise, which in turn requires a small amount of income risk perceived by households.

We next turn to the average life cycle profile of consumption, which is, again, well-known to be hump-shaped over the life cycle. Figure 9 plots the average profile implied by the model against the empirical counterpart. The model captures the rise in average consumption from age 25 up to 45 quite nicely. While the model generates some fall in consumption from age 45 to 65, it does not capture the magnitude of the substantial rise observed in the data. This is probably due to the fact the estimation only uses income and consumption data up to age 55, which does not fully reflect the full drop in consumption after this age. Overall, we conclude that the model does a reasonably good job of matching some salient aspects of life cycle income and consumption patterns despite the fact that these do not appear as explicit moments in the estimation procedure.

Before closing this section it is useful to examine the robustness of these results to the method chosen for filling in missing data. This could be potentially important because more than two-thirds of the data in our sample is missing—and therefore filled in—compared to a fully balanced panel with the same number of individuals. As an alternative procedure, we consider a simpler

statistics developed in proposition 2 of Smith (1993). This requires the estimation of the covariance matrix of the auxiliary model coefficients, accounting for the misspecification of the auxiliary model. This can be done via simulations and will be undertaken as the next step in the paper.

Figure 8: Within-Cohort Evolution of Income and Consumption Inequality: Model vs US Data

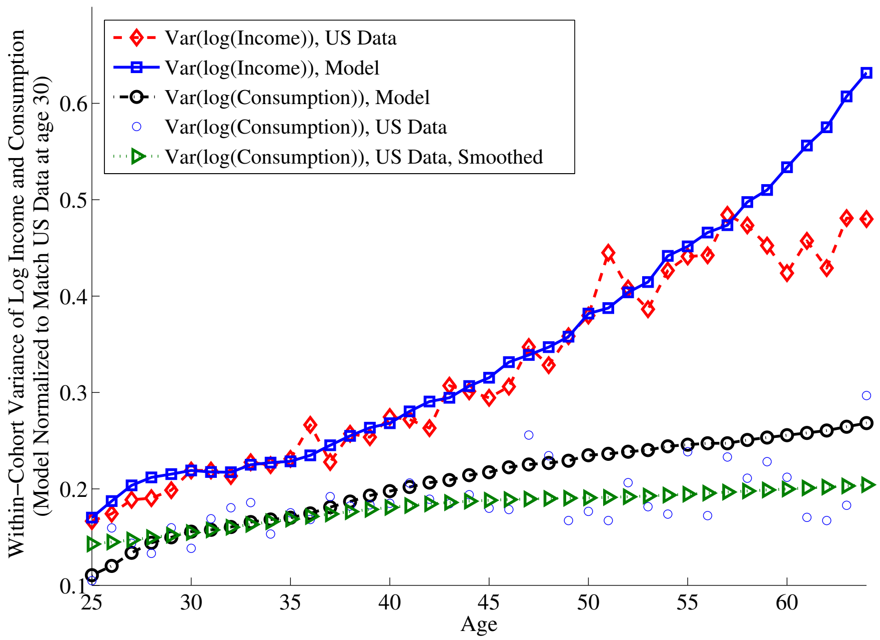
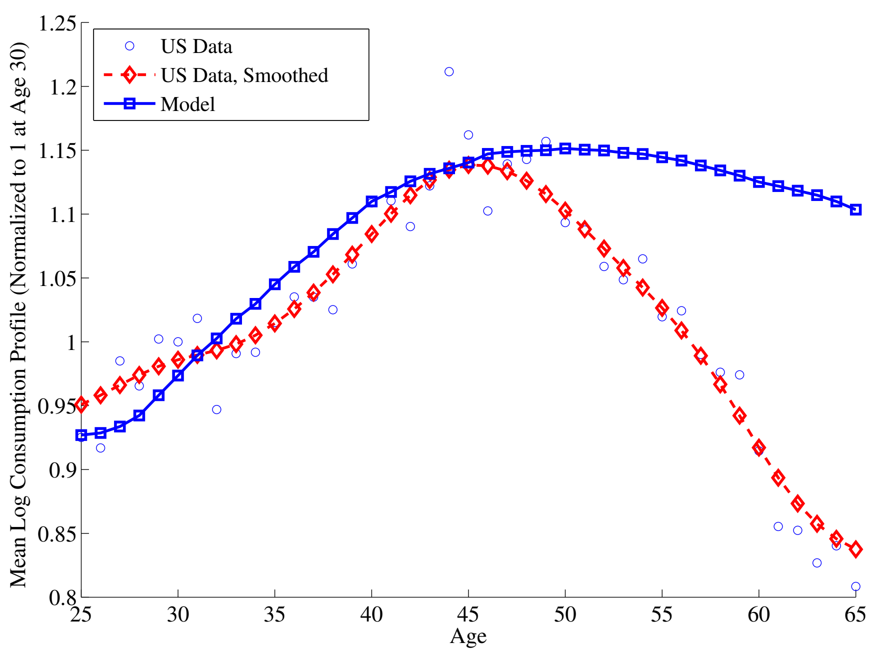


Figure 9: Mean Log Consumption Profile Over the Life Cycle: Model vs US Data



way to fill in missing data. In particular, for each individual we calculate the lifetime average of either log consumption or log income using available observations. If a consumption or income observation is missing in a given year, we simply replace the missing data with this average. We then use the data imputed in this fashion to construct all the regressors in the auxiliary model. As before, we only include non-imputed (ie, real) observations as left hand side variables. Our experimentation so far has found that the results are robust to this alternative method for filling in missing observations.[27]

6 Conclusions

The joint dynamics of consumption and labor income contains rich information that allows a sharper distinction between the RIP and HIP models. Monte Carlo results suggest that the indirect inference method works very well, even in the presence of frequently binding borrowing constraints, missing observations, retirement income, and so on, that make the auxiliary model a poor approximation to the structural relationships that need to hold in the model. On a more substantive level, we find that (i) income shocks have modest persistence, much less than a unit root, (ii) income growth rates display significant cross-sectional heterogeneity, (iii) individuals have much better information about their own income growth rates than what can be predicted by some observable variables available to the econometrician, and (iv) finally, despite significant prior information, there is also some prior uncertainty that affects consumption behavior throughout the lifecycle. Overall, our findings strongly suggest that the amount of uninsurable lifetime income risk that households perceive is smaller than what is typically assumed in calibrated macroeconomic models with incomplete markets.

A Appendix

A.1 CE data

A.1.1 1972-73 CE data

We create a measure of nondurable consumption expenditure by adding the expenditures on food, alcohol, tobacco, fuel and utilities, telephone, other services, laundry, clothing, transportation, personal goods, recreation, reading, gifts, and other goods. The original size of the 1972-73 CE is 19,975 households. We keep households in our sample if they are headed by a married male who is between 30 and 65 years old and have non-zero food and income reports. In Table 1 we report the number of households deleted from our sample during each sample selection requirement.

[27] More precisely, in the previous draft of this paper, we estimated a specification slightly different than what we consider here, where δ and $\mu_{\overline{u}^c}$ were not estimated. In that version we found that the alternative procedure for filling in missing data yielded results very similar to those obtained from the baseline method. Analogous estimation results for the current specification will be included in the revised draft shortly.

Table 5: CE Sample Selection

Criteria	Dropped	Remain
Initial Sample	—	19975
Male head	4470	15505
Age restriction	5200	10305
Non-zero income and food	709	9596
Married	874	8722
Non-missing education	213	8509

A.1.2 1980-92 CE

We merge the 1972-73 CE data with the 1980-92 data used in BPP. BPP use a similar sample selection as above. In addition, they exclude households with heads born before 1920 or after 1959. All nominal variables are expressed in constant 1982-84 dollars. Income is deflated using the CPI. Total food expenditures are deflated using the average food price series provided by the BLS. The inflation rates for food, fuel, alcohol, and transportation were determined by the corresponding price series provided by the BLS. We also drop households that have total real food consumption per adult equivalent less than $300. Here adult equivalent is defined as the square root of family size.

A.2 PSID data

A.2.1 Sample Cleaning

Our measure of total food consumption comes from summing the responses to the questions about food consumed at home and food consumed away from home in each year. (Except for 1968, where the survey only asked about total food expenditures). This gives us a total food expenditure variable in each survey wave except for 1972, 1987, and 1988, when no food expenditure questions were asked.

In the PSID the education variable is sometimes missing and sometimes inconsistent. To deal with this problem we take the highest education level that an individual ever reports and use it as the education variable for each year. Since the minimum age needed to be included in our sample is 25, this procedure does not introduce much bias to our estimated education variable.

A well-known feature of the age variable recorded in the PSID survey is that it does not necessarily increase by 1 from one year to the next. For example, an individual can report being 30 years old in 1970, 30 in 1971, and 32 in 1972. This may be perfectly correct from the respondents point of view, since the survey date may be before or after the respondent's birthday in any given year. We create a consistent age variable by taking the age reported in the first year that the individual appears as the head of a household and add one to this variable in each subsequent year.

The income variable we use is total after tax non-financial household income. The way we construct this variable varies across years in the PSID because of different questions asked and different variable definitions. From 1968-1974 we take total family money, subtract taxable income of the head and wife (which includes both asset and labor income), and add back head and wife annual labor income. The family money variable is defined as total taxable income and transfers of the head, wife, and others in the household. From 1975-1983 we take the family money variable and subtract the asset income of the head and the asset income of the wife. From 1975-1977 the asset income of the head is defined as the sum of the asset part of business income, the asset part of farming, and the asset part of rental income. From 1978-1982 the definition of the asset income of the head is the same, except for the addition of the asset part of gardening. From 1983-1991 the definition remains the same except dividend income is also added. For 1992 the definition remains the same except interest income and income from family trusts are added. From 1975-1983 wife asset income is listed as one variable. From 1984-1991 we generate wife's asset income as the sum of the wife's share of asset income and the wife's other asset income. For 1992 wife's asset income is the sum of wife's dividend income, interest income, family trust income, asset part of business income, and other asset income. From 1984-1992 to create the non-financial income variable we take family money and subtract head asset income, wife asset income, and asset income of other members of the household.

From the non-financial income variable we need to subtract taxes paid on non-financial income. For the years 1968-1990 we use the sum of the variables in the PSID that give the estimated federal tax liabilities of the head and wife and of others in the household. For 1975-1978 a variable is available that gives the amount of low income tax credit the household received. For these years the income tax credit is subtracted from the total amount of tax liability. We regress total tax liability on total labor income and its square and on total asset income and its square. We use these estimates to predict the total taxes paid on labor income.

For the years 1991 and 1992 we use the NBER TAXSIM software to estimate the total taxes paid by each household on labor income. We assume that the husband and wife file a joint tax return and that the number of dependents claimed equals the number of children in the household. We also use the annual property tax liability variable as an input to the TAXSIM software to account for property taxes being deducted from federal taxable income. Since the public release version of the PSID does not contain state identifiers, we do not use the TAXSIM software to estimate state taxes paid.

A.2.2 Sample Selection

We start with a possible sample of 67282 individuals interviewed between 1968-2005. To be in our final sample an individual must satisfy each of eight criteria in at least one year between 1968 and 1992. The number of individuals dropped at each stage in the sample selection is listed in Table 6.

Table 6: PSID Sample Selection

Criteria	Dropped	Remain
Initial sample	—	67282
Main sample	39906	27376
Continuously married	2805	24571
No major composition change	4	24567
Missing data	1032	23535
Outliers	71	23464
Topcoding	0	23464
Male and head of household	19232	4232
Age restriction	429	3803
Five observations or more	1568	2235

1. The individual must be from the original main PSID sample (not from the SEO or Latino subsamples).

2.We require that the individual be married and that the individual has not changed partners from the previous year.

3.We require that individuals had no significant changes in family composition. This means that they responded either "no change" or "change in family members other than the head or wife" to the question about family composition changes.

4. The individual must not have missing variables for the head or wife labor income. The education variable for the head must also not be missing (this education variable refers to the one created after the sample cleaning mentioned previously).

5. The individual must not have food or income observations that are outliers. An income outlier is defined as a change in real income relative to the previous year of greater than 500% or less than -80% or total income less than $1000. A food expenditure outlier is defined as real total household food expenditure greater than income or food expenditure per effective adult less than $300. Food expenditure per effective adult is defined as total household food expenditure divided by the square root of the number of members in the family.

6. We require that individuals have non-topcoded observations for the labor income of the head and wife and non-topcoded observations for total non-financial income.

7. The individual must be a male and the head of his household.

8. Household heads must be between 25 and 65 years old. (Only those between 25 and 55 are used in the main estimation in the paper.)

A.3 Additional Imputation Information

For the instrumental variables regression in the CE we create the cohort-education-year specific averages of the log of husband's and wife's hourly wage rates as follows: The cohorts are divided into 5 year cells by year of birth, starting with 1910 and ending with 1955. The education cells are divided into high school dropouts, high school graduates, and more than high school education. For each year (1972, 1973, and 1980–1992) and each cohort-education cell we calculate the mean of the log of hourly wages of household heads and wives. The four age dummies used in the interaction terms are: less than 37, between 37 and less than 47, between 47 and less than 56, and greater than or equal to 56. There are three inflation dummies: less than 5% inflation, between 5% and less than 11%, and greater than or equal to 11%. There are three children categories used in the interaction terms: one child, two children, and three or more children.

B Appendix

In this appendix we establish the asymptotic equivalence between "likelihood approach" to indirect inference employed in our estimation and the quadratic objective approach that is often used in the literature. The proof below can easily be extended to allow more general structural models (with a vector of exogenous variables, X_t, as well as more lags and leads of variable Y.

Consider the structural (ie., "true") model:

$$Y_t = f\left(Y_{t-1}, \beta\right) + \epsilon_t,$$

where $\epsilon_t \sim \text{iid}N\left(0, \sigma^2\right)$, σ^2 know, Y_0 given. Consider the auxiliary model: $Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \eta_t$, $\epsilon_t \sim$ iid $N\left(0, 1\right)$. The auxiliary model likelihood is

$$-\sum_{t=1}^{T} \left(Y_t - \gamma_0 - \gamma_1 Y_{t-1}\right)^2$$

$$\hat{h}_i\left(\beta\right) \equiv \underset{\gamma_0,\gamma_1}{\text{argmin}} \sum_{t=1}^{T} \left(Y_t^i\left(\beta\right) - \gamma_0 - \gamma_1 Y_{t-1}^i\left(\beta\right)\right)^2$$

where i denotes the i^{th}simulated data set, given β. Now define:

$$\hat{h}_M\left(\beta\right) \equiv \underset{\gamma_0,\gamma_1}{\text{argmin}} \sum_{i=1}^{M} \sum_{t=1}^{T} \left(Y_t^i\left(\beta\right) - \gamma_0 - \gamma_1 Y_{t-1}^i\left(\beta\right)\right)^2$$

as $M \to \infty$ (holding T fixed), $\hat{h}_M\left(\beta\right) \to h\left(\beta\right)$, where

$$h\left(\beta\right) \equiv \underset{\gamma_0,\gamma_1}{\text{argmin}} E \sum_{t=1}^{T} \left(Y_t^i\left(\beta\right) - \gamma_0 - \gamma_1 Y_{t-1}^i\left(\beta\right)\right)^2.$$

The approach in this paper is (assuming M is large):

$$\hat{\beta}_T = \min_{\beta} \sum_{t=1}^{T} \left(Y_t - \gamma_0(\beta) - \gamma_1(\beta) Y_{t-1}\right)^2,$$

where $\{Y_t\}_{t=0}^{T}$ is the observed data. The first order condition is:

$$\begin{aligned}
&\sum_t \left(Y_t - \gamma_0(\beta) - \gamma_1(\beta) Y_{t-1}\right) \gamma_0^{'}(\beta) + \\
&\sum_t \left(Y_t - \gamma_0(\beta) - \gamma_1(\beta) Y_{t-1}\right) \gamma_1^{'}(\beta) Y_{t-1} = 0
\end{aligned}$$

$$\begin{aligned}
=& - \gamma_0^{'}(\beta) \sum Y_t + \gamma_0(\beta) \gamma_0^{'}(\beta) T + \gamma_1(\beta) \gamma_0^{'}(\beta) \sum Y_{t-1} + \\
& - \gamma_1^{'}(\beta) \sum Y_t Y_{t-1} + \gamma_0(\beta) \gamma_1^{'}(\beta) \sum Y_{t-1} + \gamma_1(\beta) \gamma_1^{'}(\beta) \sum Y_{t-1}^2 \qquad (18)
\end{aligned}$$

Now, as an alternative, consider minimizing the following quadratic form:

$$\begin{bmatrix} \gamma_0(\beta) - \hat{\gamma}_0 \\ \gamma_1(\beta) - \hat{\gamma}_1 \end{bmatrix}^{'} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \gamma_0(\beta) - \hat{\gamma}_0 \\ \gamma_1(\beta) - \hat{\gamma}_1 \end{bmatrix},$$

where $\hat{\gamma}_T \equiv \underset{\gamma_0, \gamma_1}{\text{argmin}} \sum_{t=1}^{T} \left(Y_t - \gamma_0 - \gamma_1 Y_{t-1}\right)^2$. The F.O.C (with respect to β) is:

$$\begin{aligned}
& a_{11}\left(\gamma_0(\beta) - \hat{\gamma}_0\right) \gamma_0^{'}(\beta) + a_{12}\left(\gamma_0(\beta) - \hat{\gamma}_0\right) \gamma_1^{'}(\beta) + a_{12}\left(\gamma_1(\beta) - \hat{\gamma}_1\right) \gamma_0^{'}(\beta) + a_{22}\left(\gamma_1(\beta) - \hat{\gamma}_1\right) \gamma_1^{'}(\beta) = \\
& = \left(-a_{11}\hat{\gamma}_0 - a_{12}\hat{\gamma}_1\right) \gamma_0^{'}(\beta) - \left(a_{12}\hat{\gamma}_0 + a_{22}\hat{\gamma}_1\right) \gamma_1^{'}(\beta) + a_{11}\gamma_0(\beta) \gamma_0^{'}(\beta) + a_{12}\gamma_0(\beta) \gamma_1^{'}(\beta) \\
& \qquad + a_{12}\gamma_0^{'}(\beta) \gamma_1(\beta) + a_{22}\gamma_1(\beta) \gamma_1^{'}(\beta) = 0 \qquad (19)
\end{aligned}$$

We want to make this f.o.c. look like the one above, labelled (18). To do so, set:

$$a_{11} = T, a_{12} = \sum Y_{t-1}, a_{22} = \sum Y_{t-1}^2.$$

Then the last 4 terms in (19) match 4 of the 6 terms in (18). But what about the remaining 2 terms in each equation? One can show that these terms match up asymptotically, as the observed sample size T grows large. To see this:

$$
\begin{aligned}
&\text{plim}\left[\gamma_0'\left(\hat{\beta}_T\right)\left(T^{-1}\sum Y_t - T^{-1}T\hat{\gamma}_0 - \left(T^{-1}\sum Y_{t-1}\right)\hat{\gamma}_1\right)\right] = \\
=&\gamma_0'\left(\beta_0\right)\left(EY_t - \gamma_0\left(\beta_0\right) - \left(EY_{t-1}\right)\gamma_1\left(\beta_0\right)\right) \\
=&\gamma_0'\left(\beta_0\right)\times 0 = 0
\end{aligned}
$$

where β_0 is the "true" value of β (i.e. $\text{plim}\hat{\beta}_T = \beta_0$) because

$$EY_t - \gamma_0\left(\beta_0\right) - \left(EY_{t-1}\right)\gamma_1\left(\beta_0\right) = 0$$

is, asymptotically as $T \to \infty$, the f.o.c. that defines $\gamma_0\left(\beta_0\right)\left(= \text{plim}\hat{\gamma}_0\right)$. Similarly,

$$\text{plim}\left[\gamma_1'\left(\hat{\beta}_T\right)\left(T^{-1}\sum Y_tY_{t-1} - \left(T^{-1}\sum Y_{t-1}\right)\hat{\gamma}_0 - \left(T^{-1}\sum Y_{t-1}^2\right)\hat{\gamma}_1\right)\right]$$

$$= \gamma_1'\left(\hat{\beta}_0\right)\left(EY_tY_{t-1} - \left(EY_{t-1}\right)\gamma_0\left(\beta_0\right) - \left(EY_{t-1}^2\right)\gamma_1\left(\beta_0\right)\right) = 0,$$

again because the second term is (asymptotic) the f.o.c. that defines $\gamma_1\left(\beta_0\right) = \text{plim}\hat{\gamma}_1$. This shows that the two f.o.c.'s (18) and (19) are asymptotically equivalent.

To sum up, the LR approach—the approach we are currently using—is asymptotically equivalent, in this simplified case, to minimizing the following quadratic form:

$$\begin{bmatrix} \gamma_0\left(\beta\right) - \hat{\gamma}_0 \\ \gamma_1\left(\beta\right) - \hat{\gamma}_1 \end{bmatrix}' \begin{bmatrix} 1 & EY_{t-1} \\ EY_{t-1} & EY_{t-1}^2 \end{bmatrix} \begin{bmatrix} \gamma_0\left(\beta\right) - \hat{\gamma}_0 \\ \gamma_1\left(\beta\right) - \hat{\gamma}_1 \end{bmatrix}.$$

Note that the weighting matrix would be the optimal one if the auxiliary model were correctly specified. In this case,

$$T^{1/2}\begin{bmatrix} \hat{\gamma}_0 - \gamma_0\left(\beta\right) \\ \hat{\gamma}_1 - \gamma_1\left(\beta\right) \end{bmatrix} \to N\left(0, A\left(\beta_0\right)^{-1}\right)$$

where $A\left(\beta_0\right)$ is the weighting matrix above (which depends on the true value β_0). In general, however,

$$T^{1/2}\begin{bmatrix} \hat{\gamma}_0 - \gamma_0\left(\beta\right) \\ \hat{\gamma}_1 - \gamma_1\left(\beta\right) \end{bmatrix} \to N\left(0, A\left(\beta_0\right)^{-1} B\left(\beta_0\right) A\left(\beta_0\right)^{-1}\right)$$

where $B\left(\beta_0\right)$ is the expected value of the outer product of the gradient. In general, $A\left(\beta_0\right)+B\left(\beta_0\right) \neq 0$ because the auxiliary model is misspecified. Let $J = \begin{bmatrix} \gamma_0'\left(\beta\right) \\ \gamma_1'\left(\beta\right) \end{bmatrix}$, $A = A\left(\beta_0\right)$, and $B = B\left(\beta_0\right)$.

Then, using the (non-optimal) weighting matrix above, we get

$$T^{1/2}\left(\hat{\beta}_T-\beta_0\right)\rightarrow N\left(0,\left(J'AJ\right)^{-1}J'BJ\left(J'AJ\right)^{-1}\right)$$

Using the optimal weighting matrix, would yield

$$T^{1/2}\left(\hat{\beta}_T-\beta_0\right)\rightarrow N\left(0,\left(J'AB^{-1}AJ\right)^{-1}\right)$$

References

Abowd, J. M., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–45.

Baker, M. (1997): "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15(2), 338–375.

Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *The Journal of Political Economy*, 75(4), 352–365.

Blundell, R., L. Pistaferri, and I. Preston (2006): "Imputing consumption in the PSID using food demand estimates from the CEX," (W04/27).

——— (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5), 1887–1921.

Blundell, R., and I. Preston (1998): "Consumption Inequality And Income Uncertainty," *The Quarterly Journal of Economics*, 113(2), 603–640.

Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59(5), 1221–48.

Deaton, A., and C. Paxson (1994): "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102(3), 437–67.

Gourinchas, P.-O., and J. A. Parker (2002): "Consumption over the Life Cycle," *Econometrica*, 70(1), 47–89.

Guvenen, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, 97(3), 687–712.

——— (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12(1), 58–79.

Guvenen, F., and B. Kuruscu (2007): "Understanding the Evolution of the U.S. Wage Distribution: A Theoretical Analysis," NBER Working Papers 13096.

——— (2009): "A Quantitative Analysis of the Evolution of the U.S. Wage Distribution: 1970-2000," *NBER Macroeconomics Annual*.

Haider, S. J. (2001): "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991," *Journal of Labor Economics*, 19(4), 799–836.

Hall, R. E., and F. S. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50(2), 461–81.

Hause, J. C. (1980): "The Fine Structure of Earnings and the On-the-Job Training Hypothesis," *Econometrica*, 48(4), 1013–1029.

Huggett, M., G. Ventura, and A. Yaron (2007): "Sources of Lifetime Inequality," (13224).

Johnston, D., and S. Shipp (1997): "Trends in Inequality Using Consumption-Expenditures: The US from 1960 to 1993," *Review of Income and Wealth*, 43, 133–152.

Lillard, L. A., and Y. Weiss (1979): "Components of Variation in Panel Earnings Data: American Scientists 1960-70," *Econometrica*, 47(2), 437–454.

MaCurdy, T. E. (1982): "The use of time series processes to model the error structure of earnings in a longitudinal data analysis," *Journal of Econometrics*, 18(1), 83–114.

Slesnick, D. T. (1992): "Aggregate Consumption and Saving in the Postwar United States," *The Review of Economics and Statistics*, 74(4), 585–597.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Cyclical Dynamics in Idiosyncratic Labor Market Risk," *Journal of Political Economy*, 112(3), 695–717.

Topel, R. H. (1990): "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," (3294).