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

Fatih Guvenen[†]   Anthony Smith[‡]

August 3, 2010

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 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. 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). 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 substantially smaller than what is typically assumed in calibrated macroeconomic models with incomplete markets.

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

1 Introduction

The goal of this paper is to elicit information about the nature and amount of labor income risk from individuals' economic decisions. These choices—especially those that are dynamic in nature, such as the consumption-savings decision—contain valuable information about the economic environment that individuals face, including the future income risks they perceive. While these *perceptions* of risk are what ultimately determine individuals' behavior (and are, thus, central for economic and policy questions), they are only very imperfectly measured by actual income realizations, even in the absence of measurement problems. This is because, with income data alone, there is no way for an econometrician to identify which income movements represent surprises or "shocks" and which represent anticipated changes. This is where focusing on endogenous choices can help. The task of extracting such information from choice data faces numerous methodological hurdles, and part of the contribution of this paper is to make progress on tackling some of these technical issues.

To give a concrete example to the substantive questions we are interested in, consider Figure 1. It reproduces the well-known empirical fact that the within-cohort dispersion of labor earnings rises substantially over the life cycle (cf. Deaton and Paxson (1994), Phelan (1994), Guvenen (2009), among others).[1] There are (at least) two interpretations of this pattern, with potentially very different policy and welfare implications. The first possibility is that rising inequality may be due to the accumulation of large and persistent idiosyncratic shocks, which would imply that individuals face substantial income risk in the labor market. An alternative possibility is that rising income dispersion may be due to a *systematic* fanning out of earnings profiles over time. If this latter interpretation is correct, there are two further possibilities. If individuals can predict their own income growth rate when they enter the labor market (or they learn it quickly), the rise in inequality would mostly represent heterogeneity— and not any significant uncertainty for individuals. If the opposite is true—large initial uncertainty and slow learning—then the systematic fanning out can, in fact, represent significant uncertainty for individuals. The truth probably lies somewhere between these two polar cases. Therefore, a key empirical task is to *disentangle* the part of the rise in income inequality that represents *uncertainty/risk* from the component that is fixed/known *heterogeneity*. In turn, we would like to distinguish what part of risk comes from genuine income shocks (that can be detected in income data) and what part is due to "learning" and is, hence, in the perceptions of individuals (but not revealed in income data).

For clarity of discussion, let us consider the following process for the log labor income of individual i with t years of labor market experience:

$$y_t^i = \underbrace{g(t,\ \text{observables},...)}_{\text{common life cycle component}} + \underbrace{[\alpha^i + \beta^i t]}_{\text{profile heterogeneity}} + \underbrace{[z_t^i + \varepsilon_t^i]}_{\text{stochastic component}} \quad (1)$$

where $z_t^i = \rho z_{t-1}^i + \eta_t^i$, and η_t^i and ε_t^i are zero mean innovations that are i.i.d. over time and across individuals. The terms in the first set of brackets capture the life cycle variation in labor income that

---

[1]The data on household labor earnings are from the Panel Study of Income Dynamics (PSID).

Figure 1: Within-Cohort Inequality in Household Labor Earnings, PSID Data
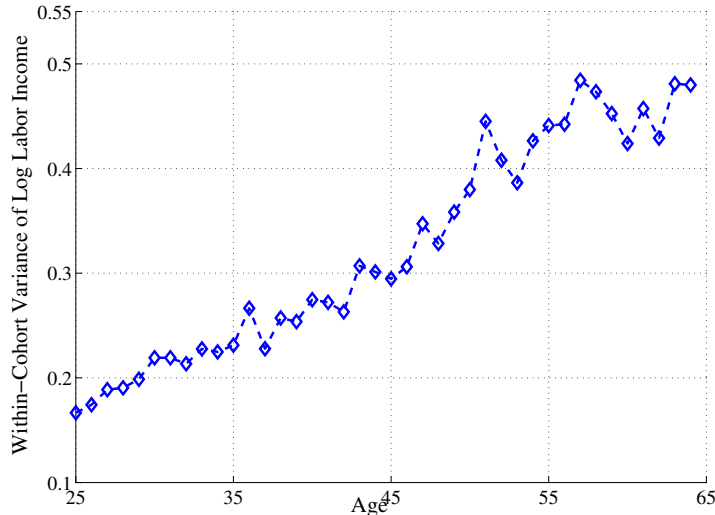
is *common* to all individuals with given observable characteristics. The second component captures potential *individual-specific* differences in income growth rates (as well as in levels). Such differences would be implied, for example, by a human capital model with heterogeneity in the ability to accumulate skills.[2] Finally, the terms in the last bracket represent the stochastic variation in income, which is modeled here as the sum of an AR(1) component and a purely transitory shock. This specification encompasses (or differs only in minor ways from) the processes estimated in the literature.[3]

A common assumption in the extant literature is to impose $\beta^i \equiv 0$, disallowing the possibility of individual-specific income growth rates. But in the absence of such differences, all the *rise* in income inequality will have to be attributed to income *shocks*, implying substantial uncertainty and risk (see (1)).[4] Put differently, this assumption rules out the possibility that individuals may be able to predict any aspect of their future (lifetime) income prospects. Such an assumption is difficult to defend on a priori grounds (in addition to being unnecessarily restrictive). Therefore, we relax this assumption in this paper and infer both how much individuals differ in their income growth rates as well as how much of these differences are known at different points in the life cycle.[5]

---

[2]See, for example, the classic paper by Ben-Porath (1967). For more recent examples of such a human capital model, see Guvenen and Kuruscu (2010, forthcoming).

[3]Among many others, see Lillard and Willis (1978), Hause (1980), MaCurdy (1982), Abowd and Card (1989), and more recently, Haider (2001), Meghir and Pistaferri (2004), Storesletten et al. (2004a), and Guvenen (2009). The process in (1) is richer than most of the specifications used to calibrate incomplete markets models, yet it still allows meaningful empirical identification. While one could postulate even more general processes (for example, allowing for separate permanent and persistent shocks components, or considering individual-specific quadratic terms, etc.), with the limited size of PSID samples, empirical identification would be problematic. Guvenen et al. (2010) do consider such richer specifications but have access to a substantially larger and cleaner data set from IRS administrative records. However, they do not have any consumption data, which is the focus of the current study.

[4]A short list of the earlier studies includes MaCurdy (1982), Abowd and Card (1989), and Topel (1990).

[5]Another question of this type is, What fraction of income movements measured by z_t^i and ε_t^i are unexpected shocks to income as opposed to being either anticipated changes or simply measurement error? While we do not study this

Briefly, we proceed as follows. We specify a life cycle consumption-savings model with CRRA utility, potentially binding borrowing constraints, and a realistic retirement pension system. Individuals face the income process in (1) but have imperfect information about their own β^i as well as the persistent shock, z_t^i. We model this learning process by assuming that individuals enter the labor market with a prior belief about their own growth rate and update these beliefs over time in a Bayesian fashion. We also allow for several types of measurement error as it is pervasive in micro income and consumption data. Then, we use indirect inference to estimate the structural parameters of this model, including the income dynamics parameters, the extent of individuals' prior information, and the tightness of borrowing constraints, among others.

Rather than selecting a set of *moments* upon which to base our estimation (as is typically done with the method of simulated moments), indirect inference allows us to focus on *a set of dynamic equations.* In this paper, we show through a Monte Carlo study that indirect inference yields excellent results when these dynamic equations are chosen to mimic the structural equations of a *simplified* (or perturbed) version of the full structural model that we wish to estimate. Specifically, if we assume quadratic utility and no borrowing constraints in the model above, we can derive exact structural equations that link consumption choice to innovations in the beliefs (about β^i and z_t^i). In Section 2.2 we use these equations to characterize the kinds of variation in consumption and income data that would be informative about different structural parameters (Propositions 1 and 2). But because this is a perturbed model, we *cannot* estimate these equations directly (e.g., via GMM) to obtain consistent estimates of our structural parameters. These equations *can*, however, form the basis of an "auxiliary model" by means of which indirect inference can deliver consistent estimates. Loosely speaking, the indirect inference estimator is obtained by choosing the values of the structural parameters so that the estimated model and the US data "look as similar as possible" when viewed through the "lens" of the auxiliary model.

The main substantive findings regarding the income process are that (i) income shocks are not very persistent ($\rho \approx 0.75$), (ii) systematic differences in income growth rates are large, and (iii) individuals have substantial amounts of information about their future income prospects. Consequently, the amount of lifetime income risk perceived by individuals upon entering the labor market is significantly smaller than what is typically assumed in calibrated macroeconomic models with incomplete markets.

The substantive question studied in this paper represents a particular example of a general situation that researchers commonly face: to elicit some information of interest from observed actions. For example, one might want to infer a female worker's perception of her market skills, future labor market attachment, and so on, by examining the timing of fertility decisions. Similarly, an operational measure of "consumer confidence" could be deduced by studying individuals' purchases of housing and other large durable goods in a structural economic model of the aggregate economy. Indirect inference can be very useful in these contexts and, in this paper, we draw some lessons on its implementation that are likely to be more general than the particular model studied here.

question here, the method that we employ is certainly flexible enough to study this question as well.

This paper is related to a small but important literature that uses panel data to study the transmission of income shocks to consumption when markets are incomplete. Important examples include Hall and Mishkin (1982), and more recently Blundell et al. (2008), and Krueger and Perri (2009). This paper contributes to this literature in the following ways. First, these studies rely on structural equations that explicitly link consumption to income innovations (to be typically estimated in a GMM framework), which imposes restrictions on the set of features one can allow in the structural model. Instead, indirect inference allows us to estimate a richer economic model and identify a number of additional parameters of interest. Second, one difficulty in this literature has been the lack of a sufficiently long panel of consumption expenditures on US households. Hall and Mishkin (1982) have used food expenditures as a measure of consumption, whereas Krueger and Perri (2009) use a rich panel data set on Italian households that contains a direct measure of consumption. Our paper is most closely related to Blundell et al. (2008), who developed a structural method for imputing nondurable consumption expenditures into the PSID by combining it with the Consumer Expenditure Survey (CE) data. Because the latter data set is available on a continuous basis only after 1980, their method has been tailored to generate imputed consumption for the PSID from 1980 to 1992. We extend their method and also use information in the (large) 1972–73 CE waves to obtain an imputed panel that covers 1968 to 1992. Finally, whereas these papers have imposed $\beta^i \equiv 0$ (and also $\rho \equiv 1$) at the outset, we estimate both the cross-sectional dispersion in β^i (as well as ρ) and the extent of individuals' knowledge about their own β^i using consumption dynamics, which is novel to this paper.[6] This exercise thus allows us to quantify how much of the *rise* in income inequality seen in Figure 1 represents risk versus known heterogeneity.

The model we estimate in this paper is based on the framework introduced in Guvenen (2007). There are at least two important differences between that work and the present paper. First, Guvenen (2007) took all the parameters of the income process as given (estimated in another paper from income data alone) and used the consumption data to calibrate the value of one parameter: the amount of prior information regarding β^i. In contrast, the present paper brings consumption data to bear on the estimation of the entire vector of structural parameters, which contains all the parameters of the income process, prior beliefs, preferences, borrowing constraints, as well as several types of measurement error. Second, that paper focused exclusively on the life cycle mean and variance profiles of consumption (using repeated cross sections); here, we use the joint dynamics of consumption and income using panel data to conduct a formal structural estimation.[7] This more formal analysis leads

---

[6] A related set of papers include Deaton and Paxson (1994), Blundell and Preston (1998), and more recently, Primiceri and van Rens (2009), which use structural models similar to those above but employ repeated cross sections of consumption and income, rather than a panel. These studies also impose $\beta^i \equiv 0$ and $\rho \equiv 1$. Primiceri and van Rens (2009) is a partial exception to this statement in that their framework potentially allows for systematic patterns in income growth, although this is conditional on $\rho \equiv 1$. They also do not estimate how much individuals know about their own growth rates.

[7] In the same spirit as Guvenen (2007), Huggett et al. (2007) use the life cycle profiles of income to pin down the parameters of the standard Ben-Porath model with cross-sectional differences in the ability to accumulate human capital, which gives rise to heterogeneous profiles. They make an a priori assumption that individuals know their income growth rate perfectly, rather than estimating prior information as we do here. However, the estimates in the present paper

us to conclude that the amount of prior information about β^i is even larger than what Guvenen (2007) gleaned from life cycle profiles: 88% of the cross-sectional variance of β^i is predictable by workers by the time they enter the labor market, compared to 62% in Guvenen (2007).[8]

From a methodological perspective, an important precursor to this paper is Gourinchas and Parker (2002), who estimate a life cycle consumption-savings model using the method of simulated moments. Their main focus is on whether such a model can explain the hump-shaped consumption profile over the life cycle. Further, they estimate the income process parameters from income data first and, in a second step, estimate risk aversion and the time discount factor from consumption data, whereas we estimate all parameters using both data sources jointly. Finally, in recent work, Bonhomme and Robin (2009, 2010) have proposed nonparametric and copula-based estimators for studying income dynamics. While in this paper we have concentrated on parametric classes of models (such as (1)) because of their dominant role in both empirical and quantitative work, the less restrictive methods proposed by these authors present an exciting avenue for future research.

The next section describes a linear-quadratic version of the consumption-savings problem that permits an analytical solution, thereby allowing us to theoretically characterize the information content in consumption choices. Section 3 then lays out the full model used in estimation. Sections 4 and 5 describe the data and the indirect inference methodology. Section 6 presents the results and several robustness exercises and extensions. Section 7 discusses the implications of the estimated model for calibrating economic models. Section 8 concludes.

2 A Framework for Inferring Income Risk

2.1 Bayesian Learning about Income Profiles

We begin by laying out a framework that allows for various possibilities regarding individuals' perceptions of their future income risk. For example, an individual is likely to have more information than the econometrician about his β^i at the time he enters the labor market and will update those beliefs as information is revealed in his income realizations. We model this process by assuming that an individual enters the labor market with some prior belief about his own β^i, which is then refined over time in a Bayesian fashion (following Guvenen (2007)). The prior variance of this belief distribution (denote by $\widehat{\sigma}^2_{\beta|0}$) measures how uncertain individuals are about their own β^i when they enter the labor market and is therefore a key parameter for determining the amount of perceived income risk.

Time 0: Prior Beliefs and Variance (Uncertainty). Imagine that for each individual, nature draws two random variables at time zero, β^i_k and β^i_u, which are independent of each other, with zero mean and variances of $\sigma^2_{\beta_k}$ and $\sigma^2_{\beta_u}$. The income growth rate is given by $\beta^i = \beta^i_k + \beta^i_u$, implying

---

indicate substantial amounts of prior information by individuals, partly vindicating their assumption.

[8]Our paper is also in the same spirit as Carneiro et al. (2003) and Cunha et al. (2005), who use the college choice to measure the amount of information youths have about their future income prospects. The latter paper finds that about 60% of the cross-sectional dispersion of labor earnings is predictable by age 18. Our estimate from full consumption dynamics over the life cycle—88% by age 25—is consistent with this finding. Both findings reveal substantial amounts of information held by young individuals about their labor market abilities.

$\sigma_\beta^2 = \sigma_{\beta_k}^2 + \sigma_{\beta_u}^2$. The key distinction between the two components is that individual i observes the realization of β_k^i, but not of β_u^i (hence the subscripts indicate *k*nown and *u*nknown, respectively). Then, the prior mean is $\widehat{\beta}_{1|0}^i = \beta_k^i$, and the prior variance is $\sigma_{\beta,0}^2 = \sigma_{\beta_u}^2$. To express the amount of prior uncertainty in relation to the heterogeneity in income growth rates, it is useful to define:

$$\lambda \equiv \frac{\sigma_{\beta,0}}{\sigma_\beta},$$

which is simply the fraction of the population dispersion of income 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., $\sigma_{\beta,0}^2 = \sigma_\beta^2$ and $\widehat{\beta}_{1|0}^i = \overline{\beta}$ 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.[9]

Updating Beliefs over the Life Cycle. We now cast the optimal learning process as a Kalman filtering problem, which yields convenient recursive updating formulas for beliefs. Each individual knows his own α^i, observes his income and the transitory shock, y_t^i and ε_t^i, and must learn about $\mathbf{S}_t^i \equiv \left(\beta^i, z_t^i\right)$.[10] The "state equation" describes the evolution of the vector of state variables that is unobserved by the decision maker:

$$\underbrace{\left[\begin{array}{c} \beta^i \\ z_{t+1}^i \end{array}\right]}_{\mathbf{S}_{t+1}^i} = \underbrace{\left[\begin{array}{cc} 1 & 0 \\ 0 & \rho \end{array}\right]}_{\mathbf{F}} \underbrace{\left[\begin{array}{c} \beta^i \\ z_t^i \end{array}\right]}_{\mathbf{S}_t^i} + \underbrace{\left[\begin{array}{c} 0 \\ \eta_{t+1}^i \end{array}\right]}_{\nu_{t+1}^i}.$$

A second (observation) equation expresses the observable variable in the model—in this case, log income net of the fixed effect and transitory shock ($\tilde{y}_t^i$)—as a linear function of the underlying hidden state:

$$\tilde{y}_t^i \equiv y_t^i - \alpha^i - \varepsilon_t^i = \left[\begin{array}{cc} t & 1 \end{array}\right] \left[\begin{array}{c} \beta^i \\ z_t^i \end{array}\right] = \mathbf{H}_t' \mathbf{S}_t^i.$$

---

[9]Notice that since each individual knows his α^i, this already gives some information about his 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_k^i captures all the correlation between α^i and β^i and $\beta_u^i \perp \alpha^i$. It can be easily shown that there is an upper bound to λ (which depends on $\sigma_\alpha^2, \sigma_\beta^2$, and $\sigma_{\alpha\beta}$) that captures this minimum information obtained from α^i alone, and this upper bound is 1 when $\sigma_{\alpha\beta} = 0$, and is 0 when $corr(\alpha, \beta) = \pm 1$.

[10]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). The assumption that ε is observable can be justified on the grounds that purely transitory shocks are likely to be easier to distinguish from persistent ones, because the latter are more easily confused with the trend. Furthermore, 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 (compared to the response implied by the data), 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 in Section 2.2, this assumption allows us to specify one general consumption-savings model that nests all the special cases we are interested in.

Both innovations have i.i.d. normal distributions and are independent of each other, with $\mathbf{Q}$ and R denoting the covariance matrix of ν_t^i and the variance of ε_t^i, respectively. Each individual's prior belief over (β^i, z_1^i) is represented by a multivariate normal distribution with mean $\widehat{\mathbf{S}}_{1|0}^i \equiv (\widehat{\beta}_{1|0}^i, \widehat{z}_{1|0}^i)$ and covariance matrix

$$P_{1|0} = \begin{bmatrix} \sigma_{\beta,0}^2 & 0 \\ 0 & \sigma_{z,0}^2 \end{bmatrix},$$

where we use the shorthand notation $\sigma_{\cdot,t}^2$ to denote $\sigma_{\cdot,t+1|t}^2$. After observing $(\widetilde{y}_t^i, \widetilde{y}_{t-1}^i, ..., \widetilde{y}_1^i)$, the posterior belief about $\mathbf{S}_t^i$ is Normally distributed with a mean vector $\widehat{\mathbf{S}}_t^i$ and covariance matrix $\mathbf{P}_t$. Define the *perceived innovation* to (log) income as

$$\widehat{\xi}_t^i \equiv \widetilde{y}_t^i - \mathbb{E}_{t-1}(\widetilde{y}_t^i) = \widetilde{y}_t^i - \left(\widehat{\beta}_{t|t-1}^i t + \widehat{z}_{t|t-1}^i\right),$$

which does not necessarily have the same sign as the true innovation to income, η_t^i—a point that will play a crucial role below. The recursive updating formulas are given by

$$\widehat{\mathbf{S}}_t^i = \widehat{\mathbf{S}}_{t|t-1}^i + \mathbf{K}_t \times \widehat{\xi}_t^i \tag{2}$$

$$\mathbf{P}_t = (I - \mathbf{K}_t \mathbf{H}_t') \times \mathbf{P}_{t|t-1}, \tag{3}$$

where $\widehat{\mathbf{S}}_{t|t-1}^i$ and $\mathbf{P}_{t|t-1}$ denote the one-period-ahead *forecasts* of these two variables and obtained from beliefs as $\widehat{\mathbf{S}}_{t|t-1}^i = \mathbf{F}\widehat{\mathbf{S}}_{t-1}^i$, and $\mathbf{P}_{t|t-1} = \mathbf{F}\mathbf{P}_{t-1}\mathbf{F}' + \mathbf{Q}$; and $\mathbf{K}_t \equiv \mathbf{P}_{t|t-1}\mathbf{H}_t \left[\mathbf{H}_t'\mathbf{P}_{t|t-1}\mathbf{H}_t\right]^{-1}$ is the (optimal) Kalman gain. Then, $\widetilde{y}_t^i$ has a Normal distribution conditional on an individual's beliefs at time $t-1$:

$$\widetilde{y}_t^i | \widehat{\mathbf{S}}_{t-1}^i \sim N\left(\mathbf{H}_t'\widehat{\mathbf{S}}_{t|t-1}^i, \mathbf{H}_t'\mathbf{P}_{t|t-1}\mathbf{H}_t\right). \tag{4}$$

For the problem at hand, (2) and (3) can be manipulated to obtain some simple expressions. First, (2) simplifies to

$$\widehat{\beta}_t^i - \widehat{\beta}_{t-1}^i = (A_t/X_t)\widehat{\xi}_t^i, \tag{5}$$

$$\widehat{z}_t^i - \rho\widehat{z}_{t-1}^i = (B_t/X_t)\widehat{\xi}_t^i, \tag{6}$$

where $A_t \equiv t\sigma_{\beta,t|t-1}^2 + \sigma_{\beta z,t|t-1}$, $B_t \equiv t\sigma_{\beta z,t|t-1} + \sigma_{z,t|t-1}^2$, and $X_t \equiv var_{t-1}\left(y_t^i\right) = A_t t + B_t$. Notice that A_t/X_t and B_t/X_t measure the fraction of the (one-step ahead) forecast variance that is due to the slope uncertainty and the persistent shock, respectively. Thus, a given $\widehat{\xi}_t^i$ is split according to the perceived share of variance attributed to each component. Second, (3) reduces to

$$\sigma_{\beta,t+1|t}^2 = \sigma_{\beta,t|t-1}^2 - \frac{A_t^2}{X_t} \tag{7}$$

$$\sigma_{z,t+1|t}^2 = \rho^2\left[\sigma_{z,t|t-1}^2 - \frac{B_t^2}{X_t}\right] + \sigma_\eta^2. \tag{8}$$

As shown by Guvenen (2007), an important feature of Bayesian learning in this framework is that beliefs about β^i change non-monotonically over the life cycle, owing to the inverse U-shape pattern followed by A^2/X. Consequently, the uncertainty regarding β^i can be very slow to resolve for a range. 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 ($\widehat{\sigma}^2_{\beta|0}$).

2.2 Understanding Identification: A Stylized Linear-Quadratic Framework

Before delving into the details of the full estimation, it is useful to provide a better understanding of the sources of identification. For example, if individuals indeed differ in their income growth rates (i.e., $\sigma_\beta \gg 0$), how would this fact be revealed in their consumption choice behavior? Similarly, can we detect the extent of an individual's prior uncertainty (that is, λ) about his β^i by observing the response of his consumption to income movements?

For this purpose, we begin with a very stylized life cycle model of the consumption-savings decision. Specifically, (i) individuals have quadratic utility over consumption, (ii) the time discount factor, δ, is the reciprocal of the gross interest rate, $1 + r$, (iii) there are no borrowing constraints, and (iv) there is no retirement. Finally, assume a simpler form of the income process in (1):

$$Y_t^i = \alpha^i + \beta^i t + z_t^i, \tag{9}$$

where the income *level* (instead of its logarithm) is linear in the underlying components, and we set $\varepsilon_t^i \equiv 0$.[11] Under these assumptions, the consumption-savings problem can be written as

$$V_t^i(\omega_t^i, \widehat{\beta}_t^i, \widehat{z}_t^i) = \max_{C_t^i, a_{t+1}^i} \left\{ -(C_t - C^*)^2 + \frac{1}{1+r} \mathbb{E}_t \left[V_{t+1}^i(\omega_{t+1}^i, \widehat{\beta}_{t+1}^i, \widehat{z}_{t+1}^i) \right] \right\}$$

$$\text{s.t.} \quad C_t^i + a_{t+1}^i = \omega_t^i, \tag{10}$$

$$\omega_t^i = (1 + r) a_t^i + Y_t^i, \quad \text{(Cash on hand)} \tag{11}$$

and the Kalman recursions (5), (6), (7), and (8).

Three Special Cases. The framework described above is a much simplified version of the full model we estimate in Section 5.2. It is, however, general enough that it encompasses three special cases of interest that we will refer to in the rest of the paper:

1. First, without any further restrictions imposed, the framework has a Heterogeneous Income Profiles process with Bayesian learning about individual income slopes—i.e., the HIP model.

2. A second important benchmark is obtained when $\sigma_\beta \equiv 0$, in which case there is no heterogeneity in profiles and no Bayesian learning. Thus, the framework reduces to the standard RIP (Restricted Income Profiles) model, extensively studied in the literature.

---

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

3. Finally, a third case of interest is when $\sigma_\beta > 0$ and $\lambda = 0$. In this case, individuals face a HIP process, but each individual knows his β^i at the time he enters the labor market. The only source of uncertainty in this case is the idiosyncratic shocks, as in the RIP model. This is an intermediate case between the HIP and RIP models in cases (1) and (2).

2.2.1 Information in Consumption Growth

Under these assumptions, optimal consumption choice satisfies

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

where $\gamma \equiv 1/(1+r)$ and $\varphi_t \equiv (1-\gamma)/\left(1-\gamma^{T-t+1}\right)$ is the annuitization factor. After substituting (9) in (12), some tedious but straightforward manipulations yield

$$\Delta C_t = \Phi(t;T,r) \left(\widehat{\beta}_t^i - \widehat{\beta}_{t-1}^i \right) + \Psi(t;T,\rho,r)(\widehat{z}_t^i - \rho \widehat{z}_{t-1}^i), \tag{13}$$

where Φ_t and Ψ_t are some age-dependent positive coefficients.[12] (In what follows, for clarity, we suppress all the arguments of Φ and Ψ except t, unless it creates confusion.) Finally, substituting (5) and (6) into (13) yields a key structural equation in this framework:

$$\Delta C_t = \Pi_t \times \widehat{\xi}_t^i, \tag{14}$$

where $\Pi_t \equiv \Phi_t \left(A_t/X_t \right) + \Psi_t \left(B_t/X_t \right)$ and $\widehat{\xi}_t^i$ is now reinterpreted as the innovation in the *level* of income—i.e., $Y_t^i - \left(\widehat{\beta}_{t|t-1}^i t + \widehat{z}_{t|t-1}^i \right)$. This equation basically says that consumption changes proportionally to the perceived innovation to income, which, as we shall see, may or may not have the same sign as η_t.

If we eliminate profile heterogeneity by setting $\sigma_\beta \equiv 0$ (and thereby also eliminate learning), the resulting (RIP) model implies

$$\Delta C_t = \Psi_t \times \eta_t^i. \tag{15}$$

The last two equations can be used to understand some of the key differences between the two frameworks.[13] When $\sigma_\beta \equiv 0$, only the current shock, 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. In the rest of this section, we present a

---

[12]Specifically, $\Phi(t;T,r) \equiv \left[\frac{\gamma}{(1-\gamma)} + \frac{t-(T+1)\gamma^{T-t+1}}{(1-\gamma^{T-t+1})} \right]$ and $\Psi(t;T,\rho,r) \equiv \frac{1-\gamma}{1-\gamma\rho} \left[\frac{\left(1-(\gamma\rho)^{T-t+1}\right)}{\left(1-\gamma^{T-t+1}\right)} \right]$.

[13]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) specifying a certain stochastic process for income. In this sense, the analysis here is in the spirit of Hall and Mishkin (1982) (and, more recently, Blundell et al. (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 income process in addition to the preference parameters, which is all one can estimate in the Euler equation approach.

Figure 2: Information about σ_β and λ in Consumption Changes
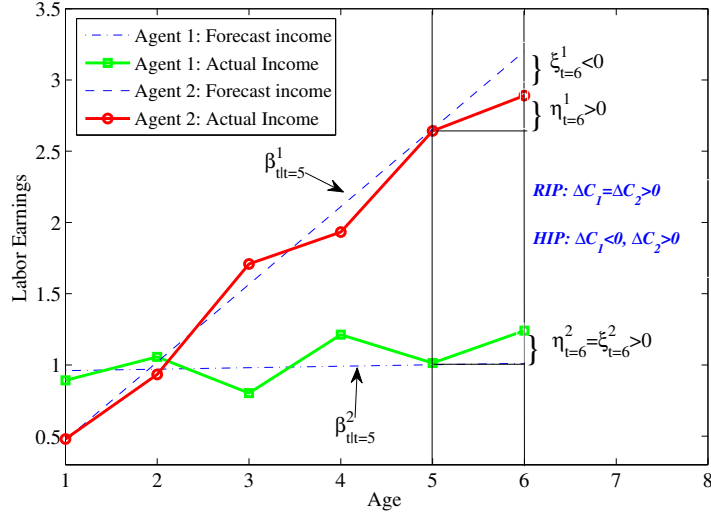
series of examples that help explain the intuition behind the identification of two key parameters—σ_β and λ. For this purpose, we specialize to the case where $\rho = 1$, which makes the exposition much simpler (although, the main conclusions we reach below hold regardless of the value of ρ). We also assume $\overline{\beta} = 0$, again, without loss of generality and for clarity of exposition.

Example 1. Consider Figure 2, which plots the income paths of two individuals up to period 6. Individual 1 experiences a faster average income growth rate in the first five periods than individual 2, but observes precisely the same rise in income between periods 5 and 6 ($\Delta Y_6^1 = \Delta Y_6^2$). If these income paths are generated by a process with $\sigma_\beta = 0$ (and thus $\beta^1 = \beta^2$), then the consumption choice of both individuals will satisfy equation (15), implying $\Delta C_6^1 = \Delta C_6^2 = \eta_6 > 0$. Instead, if the true income process has $\sigma_\beta > 0$ (HIP process), individual 1 will have formed a belief that his income growth rate is higher than that of individual 2, and forecast his income to be on the (dashed-blue) trend line. (Obviously, this remains true when $\lambda = 0$, and so when beliefs are perfectly accurate from the beginning.) Therefore, even though his income increases, it is below the expected trend ($\widehat{\xi}_6^1 < 0$), which causes him to revise down his beliefs about β^1 and, consequently, reduce his consumption level, from equation (14). In contrast, based on his past income growth (which is nearly zero), individual 2 is positively surprised to see any increase in his income between periods 5 and 6 ($\widehat{\xi}_6^2 > 0$) and will increase his consumption. Thus, we have $\Delta C_6^1 > 0$ and $\Delta C_6^2 < 0$ in response to the same income change for both agents.

The following proposition summarizes these implications in a more rigorous form. First, define $\Delta \overline{C}_t^i \equiv \mathbb{E}\left(\Delta C_t^i | \beta^i, \Delta Y_t^i\right)$. In words, $\Delta \overline{C}_t^i$ is the average consumption growth of an individual with slope parameter β^i and income growth ΔY_t^i between $t-1$ and t, where the expectation is taken over all possible histories up to $t-1$ (of initial signals and income realizations: $\{Y_1^i, Y_2^i, .., Y_{t-1}^i, \hat{\beta}_{1|0}^i\}$). In other words, $\Delta \overline{C}_t^i$ can be thought of as the consumption change of a *typical* individual with β^i and

11

who observed ΔY_t^i.

Proposition 1 [Information in Consumption Growth] *In the life cycle certainty-equivalent model with permanent shocks described above, consumption growth satisfies the following properties:*

(i) *Controlling for current income growth, consumption growth will, on average, be a* decreasing *function of an individual's β^i: i.e., $\frac{\partial \Delta \overline{C}_t^i}{\partial \beta^i} < 0$ for all t.*

(ii) *While this prediction is true for all values of λ (including when $\lambda = 0$ and, hence, when there is no learning), the relationship becomes stronger as λ rises: i.e., $\frac{\partial^2 \Delta \overline{C}_t^i}{\partial \beta^i \partial \lambda} < 0$ for all t).*

(iii) *Similarly (holding everything else constant), the response of consumption growth to income growth becomes stronger as λ increases: $\frac{\partial^2 \Delta \overline{C}_t^i}{\partial \Delta Y_t^i \partial \lambda} > 0$.*

Proof.   See Appendix C for all the omitted proofs and derivations.

To understand the empirical content of the proposition, note that even though β^i is not observed by the econometrician, in any given period, past and future income growth rates will be positively correlated with β^i. Therefore, the empirical relationship predicted by part (i) is that controlling for ΔY_t^i, consumption growth will be a decreasing function of the past income growth rate, which is observable by the econometrician. The second part of the proposition then implies that this negative dependence on the past income growth becomes stronger as individuals receive signals that are less informative at the beginning of life (i.e., a higher λ). Similarly, part (iii) implies that the response coefficient of consumption growth to income growth contains valuable information about the initial prior uncertainty faced by individuals. Loosely speaking, this is because when the initial signal is not very informative, optimal learning will result in a larger updating of beliefs about β^i in response to a given income realization, ΔY_t^i, which will in turn cause a larger change in consumption. Finally, it is easy to see that when $\sigma_\beta \equiv 0$, consumption growth will not depend on an individual's past or future income growth. As we shall see in the next section, we will use these observations to write an auxiliary model that captures the way in which consumption growth depends on past and future income growth rates, as well as how it responds to contemporaneous income growth, to infer the values of σ_β and λ as well as other parameters.

2.2.2   Information in Consumption Levels

We next turn to the information revealed in the levels of consumption and begin with the following useful lemma. (See Appendix C for the proofs.)

Lemma 1 *The consumption decision rule can be solved in closed form as a linear function of the state vector $\left(\omega_t^i, \hat{\beta}_t^i, \hat{z}_t^i \right)$:*

$$C_t^i = \varphi_t \omega_t^i + \gamma \Phi_{t+1} \hat{\beta}_t^i + \gamma \rho \Psi_{t+1} \hat{z}_t^i. \tag{16}$$

Figure 3: Information about σ_β in Consumption Levels
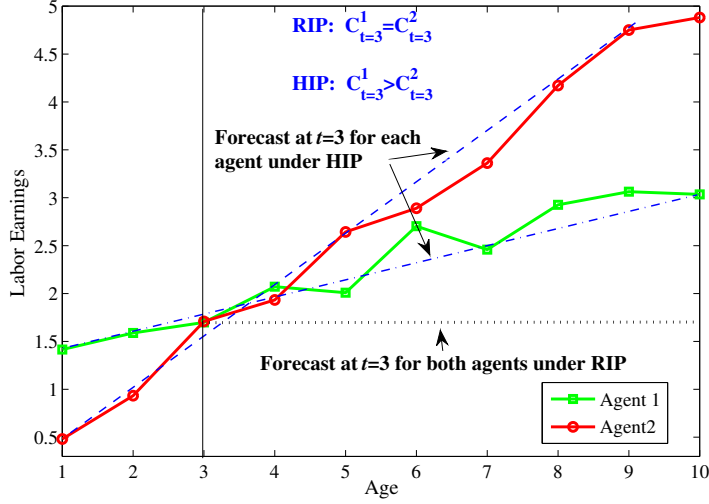
This expression clearly shows that, at every point in time, consumption choice reveals valuable information about individuals' perceived future income prospects as reflected in $(\hat{\beta}_t^i, \hat{z}_t^i)$. We now state the key result of this section and then present two specific examples to illustrate how this information can be used.

Proposition 2 [Information in Consumption Levels] *Controlling for an individual's current income and assets, the consumption level is an increasing function of the individual's beliefs about his income growth rate, $\widehat{\beta}_t^i$. This prediction holds true regardless of how much individuals know about their true income growth rate—i.e., for all $\lambda \in [0, 1]$. However, if $\sigma_\beta = 0$ consumption depends only on current income and assets.*

Proof. We present the proof in a way that is helpful for understanding the two examples that follow. Consider two individuals with $Y_t^1 = Y_t^2$, $A_t^1 = A_t^2$, and $\hat{\beta}_t^1 > \hat{\beta}_t^2$ as hypothesized in the proposition. Then we have

$$0 = Y_t^1 - Y_t^2 = \left(\hat{\beta}_t^1 - \hat{\beta}_t^2 \right) t + (\hat{z}_t^{\,1} - \hat{z}_t^{\,2}) \Rightarrow \left(\hat{\beta}_t^1 - \hat{\beta}_t^2 \right) t = -(\hat{z}_t^{\,1} - \hat{z}_t^{\,2}). \tag{17}$$

Taking the difference of the consumption level of each individual as given in equation (16) and using (17), we get

$$C_t^1 - C_t^2 = \gamma \Phi_{t+1} \left(\hat{\beta}_t^1 - \hat{\beta}_t^2 \right) + \gamma \rho \Psi_{t+1}(\hat{z}_t^{\,1} - \hat{z}_t^{\,2})$$
$$= \gamma \left(\hat{\beta}_t^1 - \hat{\beta}_t^2 \right) \left[\Phi_{t+1} - \rho t \Psi_{t+1} \right]. \tag{18}$$

Since $\hat{\beta}_t^1 - \hat{\beta}_t^2 > 0$ by hypothesis, $C_t^1 - C_t^2 > 0$ if and only if $\Phi_{t+1} - \rho t \Psi_{t+1} > 0$. The proof of Lemma A.1 in Appendix C establishes that $\Phi_{t+1} - t \Psi_{t+1} > 0$, which straightforwardly implies that

13

Figure 4: Information about Prior Uncertainty
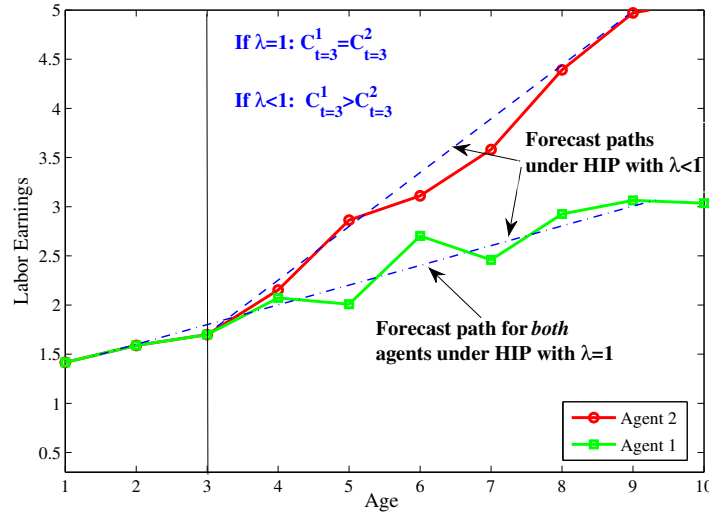
$\Phi_{t+1} - \rho t \Psi_{t+1} > 0$ as well, since $\rho \le 1$. Thus, controlling for current income and assets, an individual's consumption is higher if his (perceived) income growth prospect, $\hat{\beta}_t^i$, is higher. In contrast, when $\sigma_\beta = 0$, we have $\beta^1 = \hat{\beta}_t^1 = \beta^2 = \hat{\beta}_t^2 = \overline{\beta}$, which in turn implies from (18) that the two individuals will have the same consumption levels. ∎

To better understand the empirical content of the proposition, it is useful to study the following two examples. But first note that at any point in time, $\widehat{\beta_t^i}$ depends on two things: (i) an individual's initial belief $(\widehat{\beta_{1|0}^i})$ and (ii) the path of income realizations up to time t. Both examples consider two individuals with $\beta^1 > \beta^2$. In the first example, individuals enter the labor market with the same prior belief but experience different income paths consistent with their own β^i. The second example considers the opposite situation, with different priors but the same income history.

Example 2 (Past Income Growth Affects Current Consumption Level). Figure 3 plots a particular realization of income paths for two individuals. In this example, both individuals experience different growth rates up to period 3, but have $Y_3^1 = Y_3^2$. Now, if the true data generating process has $\sigma_\beta = 0$ (i.e., $\beta^1 = \beta^2$) and individuals perceive it as such, then both individuals' forecast of their future income would be the same: $\mathbb{E}_3(Y_{3+s}^1) = Y_3^1 = Y_3^2 = \mathbb{E}_3(Y_{3+s}^2)$, at all horizons $s \ge 0$ (shown in Figure 3 with the horizontal dotted line). If furthermore, both individuals started life with no wealth, it is easy to see that $C_3^1 = Y_3^1 = Y_3^2 = C_3^2$. In contrast, when $\sigma_\beta > 0$, individuals know that they can have different β^i's and will use the past income growth to form beliefs about their own β^i. Based on the high past income growth, individual 1 will expect a higher β^i and, therefore, a much higher lifetime income than individual 2. (And if $\lambda = 0$, then each will know his β^i with certainty from the beginning.) Therefore, the first individual will have a higher consumption *level* than individual 2 at the same age, despite having the same income level.

14

Example 3 (Dependence of Consumption Level on Future Income Growth Reveals Prior Information). One can turn the same argument around to see how the level of consumption can also be informative about the degree of private information, λ. To show this, we turn to Figure (4), which is a slight variation of Figure 3. Consider two individuals with $\beta^1 > \beta^2$, who nevertheless experience the same income realizations up to period 3. If, furthermore, $\lambda = 1$, then both individuals will have the same beliefs in period 3 and, therefore, will choose the same consumption level. If, on the other hand, individuals have some prior information (i.e., $\lambda < 1$), the individual that starts out with a higher prior belief ($\widehat{\beta}^1_{1|0} > \widehat{\beta}^2_{1|0}$) will also have a higher belief at time t ($\hat{\beta}^1_3 > \hat{\beta}^2_3$) and therefore have a higher consumption level. This implication stands in contrast to the RIP model, which predicts no such dependence on past income levels (beyond what is captured by current income and assets). Moreover, since $\widehat{\beta}^1_{1|0}$ is positively correlated with the true β^1, and both individuals in this example observed the same *past* income paths, the individual's prior belief will be correlated with his *future* future income growth. Thus, controlling for current income and assets, and the past income path, the correlation between current consumption and future income growth also reveals how much prior information the individual has. This is a useful prediction, as it shows how the observable variation in consumption and income can be used to infer the amount of prior information, which is unobservable.

These three examples illustrate how one can use the structural equations—such as (14), (15), and (16) that hold true exactly in a somewhat simplified version of the economic model to be estimated—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 (means, covariances, etc.). Below we shall write a parsimonious auxiliary model that captures these dynamic relationships to identify the key parameters of the income process.

The results of this section illustrate some important advantages of using the information revealed by intertemporal choices, such as consumption-savings, over using panel data on income alone. One difficulty of the latter approach is that identification between different income processes partly depends on the behavior of the *higher*-order autocovariances of income (Guvenen (2009) contains a detailed discussion of this point). 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, as can be seen from (14) and (16).

2.3   Indirect Inference versus GMM

The standard method for estimating consumption-savings models since Hall and Mishkin (1982) has been to derive explicit structural expressions that link observable variables (such as consumption and income) to unobservable variables (such as persistent and transitory income shocks). It is useful to contrast this approach with the indirect inference method we employ in this paper. To this end, consider the certainty-equivalent consumption-savings model described above, but in keeping with earlier work (Hall and Mishkin (1982); Blundell et al. (2008)), suppose that the income process is the sum of a permanent and a transitory shock, which implies $\Delta Y_t = \eta_t + \Delta \varepsilon_t$, where $\eta_t, \varepsilon_t \sim$ i.id. Here, it

can be shown that $\Delta C_t = \eta_t + \varphi_t \varepsilon_t$. These two equations can be jointly used to estimate the ratio of shock variances $(\sigma_\eta^2/\sigma_\varepsilon^2)$, which is a measure of the persistence, or durability, of shocks. To see how, consider this regression:

$$\Delta C_t = \pi \times \Delta Y_t + error, \qquad \text{where} \quad \pi \equiv \frac{1+\varphi_t(\sigma_\varepsilon^2/\sigma_\eta^2)}{1+2(\sigma_\varepsilon^2/\sigma_\eta^2)}. \tag{19}$$

Thus, $\sigma_\eta^2/\sigma_\varepsilon^2$ can be identified by estimating π from this regression. When income shocks are permanent (i.e., $\sigma_\varepsilon^2 = 0$), we get $\pi = 1$, and consumption moves in lockstep with income from (19). At the other extreme, when $\sigma_\eta^2 = 0$, $\pi = \varphi_t/2 \approx 0$, implying that consumption fluctuations will be much smoother than those in income. This identification strategy is commonly used in the literature, where a set of coefficients (such as π above) are estimated via GMM or a minimum-distance method. Notice, however, that even in this simple example, inference is feasible only because we have two exact expressions that link ΔC_t and ΔY_t to η_t and ε_t, which were used to derive the expression for π. It follows that the consistency of the estimates relies—potentially critically—on the validity of the assumptions we made above (quadratic utility, no borrowing constraints, etc.) to make the derivation of these two equations feasible.

To illustrate some of the potential difficulties with this approach, now consider the example shown in the left panel of Figure 5. The line marked with squares is income, and the dashed line marked with circles is consumption. (Ignore the dashed-dot line for now.) The main observation that jumps out of this picture is that the individual's consumption and income move almost one-for-one during most of his life (from 25 to 57). Thus, using (19) for inference would lead one to conclude that income shocks are nearly permanent. Indeed, a GMM estimation using the coefficient above (i.e., $\sigma(\Delta C_t, \Delta Y_t)/\sigma^2(\Delta Y_t)$) as the only moment yields $\sigma_\eta^2/\sigma_\varepsilon^2 = 27.2$—that is, income shocks are almost completely permanent, when in fact the true ratio used in the simulation was $\sigma_\eta^2/\sigma_\varepsilon^2 = 2$!

To understand the source of this substantial bias, now consider the dash-dot line, which plots the asset position (scaled to fit in the figure): the individual is right up against his constraint up to the mid-50s, after which time savings for retirement starts to kick off. But borrowing constraints had to be ignored to derive (19), which turns out to be critical for this individual. In other words, the consumption of this individual moves one-for-one with his income because he is not able to borrow any further and hence has no choice but to consume his income—not because income shocks are permanent and the individual chooses to accommodate these shocks fully. Although this is clearly an extreme case example, it sounds a cautionary note that this common assumption has the potential to bias inference if a nonnegligible fraction of households are borrowing constrained.

A second example is shown in the right panel of Figure 5. A particular income path is plotted here along with the consumption path that would be chosen by an individual when his risk aversion is, respectively, 3 (circles) and 0.3 (diamonds).[14] As could be expected, when the individual has higher risk aversion, the consumption path is much smoother than when he is more risk tolerant. Now assume that the econometrician observes the consumption path of the risk tolerant individual,

---

[14]Although assets have not been plotted to save space, the individual is never constrained in either case.

Figure 5: Inferring Persistence of Shocks from the Comovement of Income and Consumption Using GMM Moment Conditions
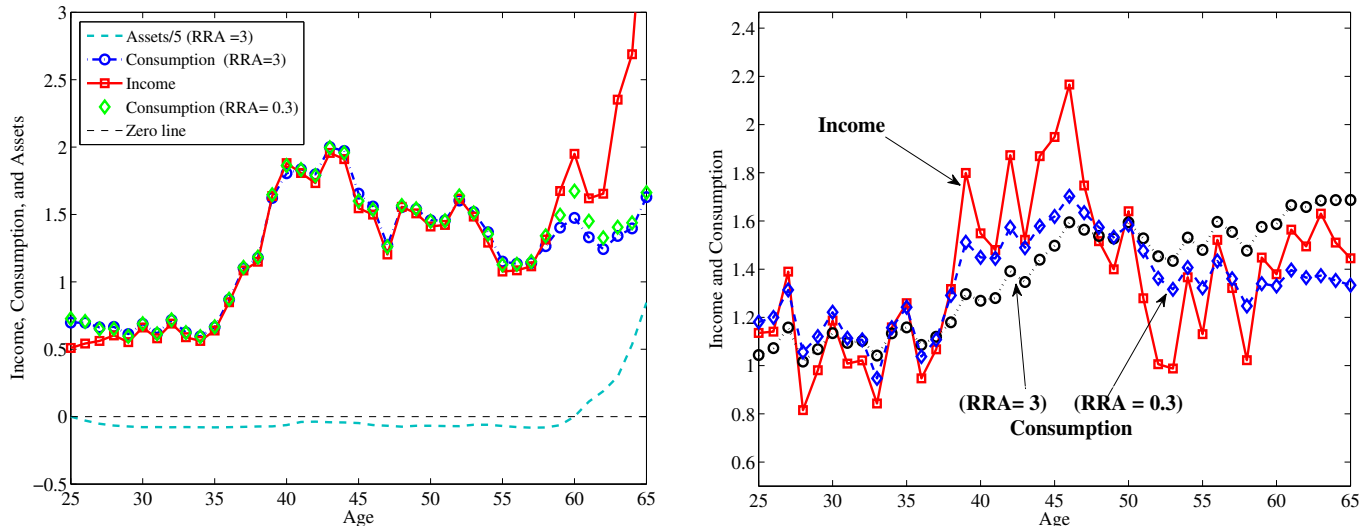
but since risk aversion is not observable, he assumes a value of 3.0 as a reasonable figure. (When equation (19) above is derived using quadratic utility, an assumption about the value of risk aversion is made automatically.) In this case, observing the strong response of consumption to income, the econometrician would be led to overestimate the persistence of income shocks. Of course, the opposite case would arise if the econometrician assumes a risk aversion lower than the true value.

The indirect inference approach we use in this paper allows us to relax several of these assumptions, model and estimate the tightness of the borrowing constraints and the time discount factor, and explore the effects of different risk aversion parameters on the estimates of income dynamics parameters.

3   The Full Consumption-Savings Model

Each individual lives for T years and works for the first R $(< T)$ years of his life, after which he retires. Individuals do not derive utility from leisure and, therefore, supply labor inelastically.[15] During the working life, the income process is given by the stochastic process specified in equation (1). As before, individuals can borrow and lend a constant interest rate r subject to an age-dependent lower limit as specified below. This model is essentially the same as the one analyzed in Guvenen (2007) (i.e., with the exception of the richer specification of the borrowing constraint and computation of retirement pension described below.)

The relevant state variables for this dynamic problem are cash-on-hand (assets plus labor income),

---

[15]While endogenous labor supply would be a valuable extension and is technically feasible, it raises a number of difficult issues to deal with (such as the extensive margin of female labor supply, household utility function, etc.) and would make the problem computationally much more challenging than it already is. We therefore leave this extension for future research.

ω_t^i, and the vector of mean beliefs, $\widehat{\mathbf{S}}_t = \left(\widehat{\beta}_t^i, \widehat{z}_t^i\right)$. Therefore, the dynamic program is

$$V_t^i(\omega_t^i, \widehat{\beta}_t^i, \widehat{z}_t^i; \alpha^i) = \max_{C_t^i, a_{t+1}^i} \left\{ \frac{(C_t^i)^{1-\phi}}{1-\phi} + \delta \mathbb{E}_t \left[V_{t+1}^i(\omega_{t+1}^i, \widehat{\beta}_{t+1}^i, \widehat{z}_{t+1}^i; \alpha^i) \right] \right\}$$

$$\text{s.t.} \qquad C_t^i + a_{t+1}^i = \omega_t^i$$

$$\omega_t^i = (1+r)\, a_t^i + Y_t^i$$

$$Y_t^i = \underline{Y} + \exp(\overline{y}_t + \alpha^i + \beta^i t + z_t^i + \varepsilon_t^i) \tag{20}$$

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

for $t = 1, ..., R-1$, where V_t^i is the value function of a t-year-old individual; $\underline{a}_t$ is an age-dependent borrowing limit, which will be specified in a moment; and $\overline{y}_t$ is the population average of log income at age t (calculated from the PSID sample). The evolutions of the vector of beliefs and its covariance matrix are governed by the Kalman recursions (5, 6, 7, 8). Finally, the expectation is taken with respect to the conditional distribution of $\widetilde{y}_{t+1}^i$ given in (4).[16]

Notice that the expression for labor income now includes an additional term, $\underline{Y} > 0$, whose purpose is to make sure that even the lowest income realization is never too close to zero, which can happen since all the terms in y_t^i are normally distributed and, thus, have no lower bound. But the possibility of income hitting zero would make borrowing effectively impossible (since the individual would have no funds to pay back its debt), which does not seem realistic given the sizeable uncollateralized borrowing by households observed in the data.

Borrowing Constraints. As discussed above, the tightness of the borrowing constraints can have a potentially large impact on the estimates of the income process parameters. Therefore, rather than picking a (n arbitrary) value for $\underline{a}_t$ beforehand, we estimate the borrowing limit along with the rest of the structural parameters. Our starting point is the *natural* borrowing limit, which is essentially the *loosest* limit that still guarantees full repayment by the last period (T) even if the individual gets the lowest income realization in every period. Here, this limit would be $\underline{a}_t = \sum_{\tau=1}^{T-t} \gamma^\tau \min(Y_\tau) = \underline{Y} \frac{1-\gamma^{T-t}}{1-\gamma}$. Although this is a conceptually clean and useful benchmark, it has the somewhat questionable implication that individuals face a looser constraint when young than when old, which is the opposite of what we seem to observe in real life. Of course, the relaxation of borrowing constraints for many households with age probably has to do with a longer credit history becoming available to lenders, which effectively reduces the risk of lending. To capture these concerns in a simple fashion (without introducing complications of default and credit rating), we assume that banks use a potentially higher interest rate to discount individuals' future labor income during working years in calculating their borrowing limit, but simply apply the risk-free rate for discounting retirement income. That is, we

---

[16]As mentioned in footnote 10, we found that whether or not ε_t is assumed to be observable has a negligible effect on the empirical results. But when ε_t is observable the expectations operator in this dynamic program needs to be computed as a double integral, which slows down computations. Thus, we report all the results below assuming that ε_t is *not* observable.

define

$$\underline{a}_t \equiv \sum_{\tau=1}^{R-t} (\psi\gamma)^\tau \min(Y_\tau) + \sum_{\tau=R-t+1}^{T-t} \gamma^\tau \min(Y_\tau) = \underline{Y}\gamma \left[\frac{\psi\left(1-(\gamma\psi)^{R-t}\right)}{1-(\gamma\psi)} + \gamma^{R-t}\frac{1-\gamma^{T-R}}{1-\gamma} \right],$$

where $\psi \in [0,1]$ measures the tightness of the borrowing limit. When $\psi = 0$, no borrowing is allowed against future labor income; when $\psi = 1$, individuals can borrow up to the natural limit. This specification generates borrowing limits that become looser (tighter) with age when ψ is sufficiently low (close to 1). The tightness parameter ψ will be estimated in the empirical analysis.

Retirement Period. 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:

$$V_t^i(\omega_t^i; Y) = \max_{c_t^i, a_{t+1}^i} \left[\frac{(C_t^i)^{1-\phi}}{1-\phi} + \delta V_{t+1}^i(\omega_{t+1}^i; Y) \right] \tag{21}$$

$$\text{s.t.} \quad Y^i = \Phi\left(Y_R^i; \overline{Y}\right), \text{ and eq. (10, 11)}$$

for $t = R, ..., T$, with $V_{T+1} \equiv 0$.

Social Security System. The pension system in the model—captured by the function Φ—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 $\overline{Y}^i$).[17] Adopting this exact structure here, however, would add another state variable—$\overline{Y}^i$—to the dynamic problem above, increasing the already high computational burden of the estimation. 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 the worker's earnings at the retirement age (Y_R^i). This is accomplished 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 does not add a state variable but recognizes the empirical relationship between average earnings and the earnings at retirement age 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 function is given by

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

---

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

4   The Data

This section discusses the data used in the empirical analysis and describes how we construct the panel of imputed household consumption. 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 obtained from Blundell et al. (2008). 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 et al. (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 nationally representative core sample, whose head is between the ages of 25 and 55 (inclusive), and has nonmissing 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, is contained in Appendix A.

4.1   Constructing a Panel of (Imputed) Consumption

The 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 nondurable consumption (available in PSID), or resorted to using repeated cross-sections from CE under additional assumptions.

In a recent paper, Blundell et al. (2006) (hereafter, BPP) develop a structural method that imputes consumption expenditures for PSID households using information from the CE survey. 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. To deal with the endogeneity of food and nonfood expenditures as well as measurement error in these variables, the estimation is carried out with an instrumental variables regression. The key condition for the imputation procedure to work is that all the variables in the demand system must be available in the CE data set, and all but nondurable expenditures must be available in the PSID. One then estimates this demand system using the CE data, 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 the PSID for the period 1980 to 1992, and show that several statistics calculated using the imputed measure compare quite well with their counterparts from the CE data.

Our Imputation Procedure.   In this paper, we modify and extend the method proposed BPP to cover the period 1968 to 1992. Here we provide a brief overview of our method and a discussion of the quality of the imputation. Further details are in Appendix A.3.

First, BPP include time dummies interacted with 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 data is not available on a continuous basis

before 1980, whereas we would like to use the entire length of the PSID (going back to 1968), making the use of time dummies impossible. To circumvent this problem, we replace the time dummies with the food and fuel inflation rate, which 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.[18] The data in this earlier period also appear to be of superior quality in certain respects compared with those from subsequent waves.[19] The use of these earlier data provides, in 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 provides a good fit to the corresponding statistics in the CE data. Here, we summarize the most relevant statistics. Further details are contained in Appendix A.3.

We begin with two dimensions of consumption data that are crucial for our estimation exercise. First, the left panel of Figure 6 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).[20] To reduce the noise in the data, the figure also plots the corresponding "smoothed" series obtained by a Nadaraya-Watson kernel regression, with a Gaussian kernel. The two graphs overlap remarkably well, especially up to early age 50.[21] Second, the right panel plots the *within-cohort variance* of log consumption over the life cycle 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 studied the CE data over the period from 1980 to 1990 (such as Deaton and Paxson (1994) and Storesletten et al. (2004b)) but is consistent with more recent papers that have used samples extending to the late 1990s (cf. Heathcote et al. (2010)). This

---

[18]The sample size is around 9,500 units in 1972–73 surveys, but ranges from 4,000-6,000 units in the waves after 1980. There are also some differences in the survey design in the earlier CE—such as the nonrotating 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.

[19]Slesnick (1992) shows that when one aggregates several subcomponents of consumption expenditures in the CE, they come significantly closer to their counterparts in the National Income and Product Accounts (NIPA) than the CE waves after 1980. For example, in 1973 total expenditures measured by the CE is 90% of personal consumption expenditures as measured by NIPA, whereas this fraction is consistently below 80% after 1980 and drops to as low as 75% in 1987. Similarly, consumer services in the CE accounts for 93% of the same category in NIPA in 1973 but drops to only 66% in 1989.

[20]The life cycle profiles are obtained by controlling for cohort effects as described in Guvenen (2009).

[21]If we do not use the 1972–73 CE in the imputation procedure, the average profile of imputed consumption would rise by 51% between ages 25 and 45 instead of the 22% rise in the baseline imputation and would therefore vastly overestimate the corresponding rise in the CE data shown in Figure 6.

Figure 6: Mean and Variance Profile of Log Consumption over the Life Cycle
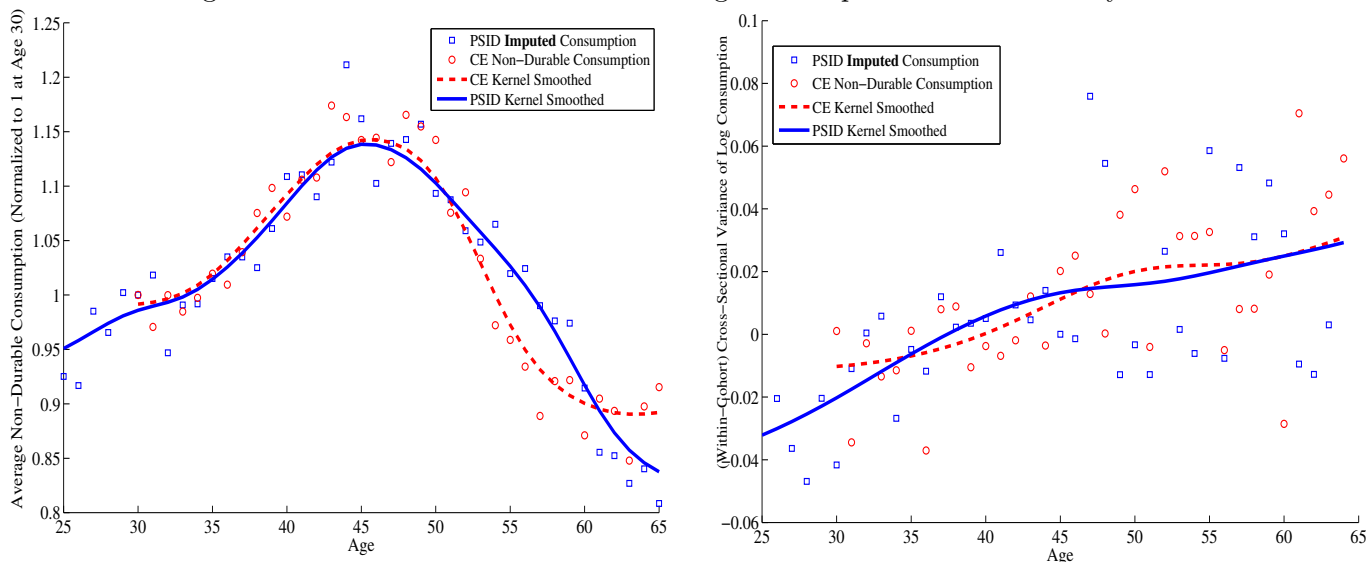
finding will be important in understanding some of the estimates (especially, λ) that we obtain in the structural estimation below.

Another useful exercise is to test the *out-of-sample* predictive ability of the imputation procedure. To do this, we split the CE sample used in the imputation above into two randomly drawn subsamples (each containing exactly half of the observations in each survey year). We use the first subsample to estimate the food demand system as above, which we then use to impute the nondurable consumption of the second subsample (control group).[22] 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 form a cloud that aligns very well with the 45-degree line. In fact, a linear regression of imputed consumption on the actual one 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.[23] 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.

As a final, and rather strict, test to detect whether systematic patterns exist in the imputation error, we regressed it 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 at the 5% level in any given simulation), indicating no evidence of systematic imputation errors by

---

[22]To control for the randomness of each subsample, we repeat this exercise 200 times.

[23]The results in the text refer to the average of these 200 replications. 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.
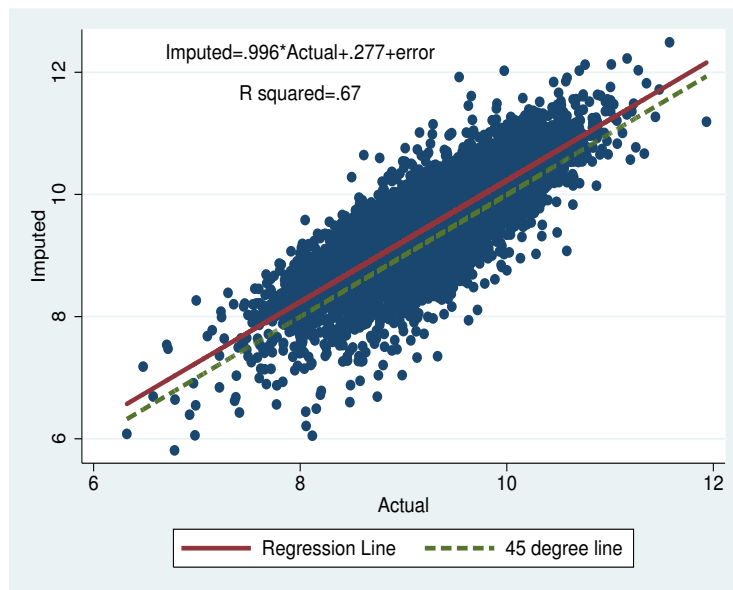
Figure 7: **Out of Sample Predictive Power of the Imputation Method in the CEX.** This plot is obtained by estimating the IV food demand system on a randomly chosen half of the CEX sample, and then imputing the consumption for the other half (control group). The figure plots the actual consumption of 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 200 repetitions.

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.

Measure of Household Labor Income. In the 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, financial help from relatives, and so on). We then subtract the labor portion of income taxes paid by each household, which is estimated using the procedure described in Appendix A.2.

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, a race dummy, a region dummy, a dummy indicating whether the head is employed, a dummy indicating residence in a large city, and a set of cohort dummies.[24] We then use the residuals of these regressions—which are interpreted as consumption and income per-adult equivalent—in the analysis below.

---

[24]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.

5 Econometric Methodology: Indirect Inference Estimation

We now 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 the parameters of economic models, first introduced by Smith (1990, 1993) and later extended by Gourieroux et al. (1993) and Gallant and Tauchen (2006). It is most useful in estimating models for which the likelihood function (or any other criterion function to be maximized) is analytically intractable or too difficult to evaluate, as is the case here. For example, in the general model with Bayesian learning, individuals' consumption choices depend on their beliefs, which is unobserved by the econometrician, making a direct estimation impractical. 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 (e.g., no borrowing constraints, quadratic utility, a simplified retirement structure, and so on), and/or 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 over GMM is that this auxiliary model does *not* need to correspond to any valid moment condition of the economic model for the structural estimates to be consistent. This allows significant flexibility in choosing an auxiliary model: it can be any statistical model relating the model variables to each other *as long as* each structural parameter has an independent effect on at least one (reduced form) parameter of the auxiliary model.[25] This 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 the 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 (such as (14) and (16)) that are difficult to express as simple unconditional moments, as is often done with MSM. We illustrate this point now.

5.1 A Parsimonious and Feasible Auxiliary Model

In light of propositions 1 and 2, it would seem that equations (14) and/or (16) would be ideal candidates to form the basis of an auxiliary model. For example, the response coefficient in equation (14) is $\Pi_t (\lambda, \sigma_\beta, \sigma_{\alpha\beta}, \sigma_\eta, r, \rho; R, T)$, which depends on several key variables that we wish to estimate.[26] However, notice that $\widehat{\beta}^i_{t|t-1}$ and $\widehat{z}^i_{t|t-1}$ that appear in these equations are unobserved by the econometrician, which makes those equations infeasible as an auxiliary model. But as discussed in Section

---

[25]Technically, this last condition corresponds to the "full rank" requirement of the binding function. For more details, see Smith (1993) and Gourieroux et al. (1993). While the full rank condition is sufficient for the consistency of the estimates, the precise specification of the auxiliary model does matter for the efficiency of the estimator.

[26]The dependence of Π on $\lambda, \sigma_\beta, \sigma_{\alpha\beta}$, and σ_η can be seen from the formulas for A_t and B_t.

2.2, beliefs are updated over the life cycle using past income realizations, and prior beliefs are likely to be correlated with future income realizations, which leads us to the following feasible regression:[27]

$$c_t = \mathbf{a}'\mathbf{X}_{c,t} + \epsilon_t^c = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t+1} + a_4 y_{t+2} + a_5 \overline{y}_{1,t-3}$$
$$a_6 \overline{y}_{t+3,R} + a_7 \Delta y_{1,t-3} + a_8 \Delta y_{t+3,R} + a_9 c_{t-1} + a_{10} c_{t-2} + a_{11} c_{t+1} + a_{12} c_{t+2} + \epsilon_t^c, \tag{22}$$

where c_t and y_t are the log of consumption and income, respectively; $\Delta y_{\tau_1,\tau_2}$ and $\overline{y}_{\tau_1,\tau_2}$ are, respectively, the average of the growth rate and the average of the level of log income from time τ_1 to τ_2; and $\mathbf{a}$ and $\mathbf{X}_{c,t}$ denote the vectors of coefficients and regressors. The use of logged variables in this regression seems natural given that the utility function is CRRA and income is log-normal. This regression captures the predictions made by the HIP and RIP models discussed above, by adding past and future income growth rates 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 income regressors above as left-hand-side variables:

$$y_t = \mathbf{b}'\mathbf{X}_{y,t} + \epsilon_t^y = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + b_3 y_{t+1} + b_4 y_{t+2} + b_5 \overline{y}_{1,t-3} + b_6 \overline{y}_{t+3,R}$$
$$+ b_7 \Delta y_{1,t-3} + b_8 \Delta y_{t+3,R} + \epsilon_t^y. \tag{23}$$

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.[28] For each age group, the auxiliary model has 22 regression coefficients (13 in the first equation and 9 in the second), two residual variances, and the correlation between the two residuals 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.

5.2 Empirical Preliminaries

Preset Values. Individuals work for $R = 41$ years, and the retirement duration is 15 years ($T = 80$). The price of the one-period discount bond is set to 0.95, implying an interest rate of $r = 1/0.95 - 1 \approx 5.26\%$. The *common* life cycle profile of log income ($g(\cdot)$ in (1)) is captured by feeding into the model the empirical profile computed from our PSID estimation sample. The (potential) lower bound of income, $\underline{Y}$, is set to be 5% of average income in this economy. As we discuss in further detail below, we have found it very challenging to separately identify the coefficient of relative risk aversion, δ, and the time discount factor, ϕ, despite substantial efforts (expanding the auxiliary model, adding

---

[27]Moreover, the expressions for consumption in Section 2.2 were derived under several restrictive assumptions, none of which is satisfied in the full model we wish to estimate. Fortunately, these issues do not represent a problem for the consistency of the estimates of the structural parameters that we are interested in.

[28]Although the auxiliary model would correspond to the structural equations in (14) and (16) more closely if the coefficients were varying freely with age, this would increase the number of parameters in the auxiliary model substantially. We have experimented with having one or three age groups but found the small sample performance of the estimator to be significantly better with the specification adopted here.

some new moments, and so on). Therefore, in our benchmark case, we proceed by fixing ϕ at 2 and estimating δ. Later, we will conduct detailed sensitivity analyses with respect to the values of the parameters preset in this section (including ϕ, r, and $\underline{Y}$).

Measurement Error. We add measurement error to simulated consumption and income data:

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

where $y_t^{i,*}$ and $c_t^{i,*}$ are measured variables of household i, respectively, and $u_t^{i,y}$ and $u_t^{i,c}$ are zero-mean random variables that are independent over time, with standard deviations of σ_y and σ_c.[29] Notice that we have also added a second term to consumption, $\overline{u}^{i,c}$, which is an individual fixed measurement error with potentially nonzero mean in the cross section and standard deviation σ_{c_0}. This fixed effect is needed for two reasons. First, and most importantly, recall that we regress both income and consumption on a set of demographics to convert these variables into per-adult equivalent terms. One effect of this adjustment is to introduce level differences between consumption and income, the magnitudes of which vary across households. This fixed effect captures such differences.[30] Second, the model described above abstracts from initial wealth differences across households. 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.

Missing Observations. In the observed data set, 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). With more than half of the observations missing compared to a fully balanced panel, one question is, how does one run the regressions in (22) and (23)? For missing values of regressors, we simply use values that are constructed or "filled-in" using a reasonable procedure.[31] We only use actual observations on the

---

[29]The variable $u_t^{i,c}$ contains the imputation error which is heteroskedastic owing to the mechanics of the imputation procedure (a point also observed by BPP). In our estimation we will impose stationarity in variances to deal with the computational burden, but further work is needed on the effects of this assumption.

[30]An alternative way to understand this point is to observe that the levels 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 a 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.

[31]Basically, at each age that a household 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 household over all the ages that it has a valid observation. Then for each missing income observation of this household, we impute the income level corresponding to its 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 growth

left hand side of regressions. As will become clear in the Monte Carlo analysis, a strength of the indirect inference method is that the particular filling-in method is not critical for the estimation as long as the same procedure is applied consistently to real and simulated data (which is what we do). As an extreme example, if we simply fill in all missing values with zeros, the estimates would still be consistent, as we show in the Monte Carlo analysis.

Matching the Wealth-to-Income Ratio. The auxiliary model specified above does not explicitly target the amount of savings and wealth generated by the estimated model. One goal of this paper, however, is to provide estimates of income processes (together with combinations of time discount factor, risk aversion, and borrowing constraints) that can be used for calibrating life cycle models to be used in quantitative macroeconomic analysis. For this purpose, it is important to make sure that the estimated model yields a reasonable amount of savings. Furthermore, households' wealth holdings contain some useful information that can shed light on (some of) the estimated parameters. Unfortunately, detailed information about asset holdings in the PSID is available only intermittently (and in fact, only in 1984 and 1989 in our sample period), which makes it difficult to use some dynamic equations, such as (22) and (23), as additional auxiliary model regressions. As a compromise, we add one static moment condition—the median wealth-to-income ratio (denoted by WY henceforth)—to the objective function that the indirect inference procedure is minimizing, to be described in a moment. Specifically, we add $10 \times (WY_{PSID} - WY_{Model})^2$ to the objective function in (25). This additional moment also makes pinning down δ much easier in the estimation.

To calculate the empirical target, WY_{PSID}, we use the wealth supplement of PSID, available in 1984 and 1989.[32] We use households in our estimation sample that were also present in the PSID in 1984 and/or 1989 (so that we can calculate a measure of wealth for them). The median value of WY is 0.99 in 1984 and 1.17 in 1989, averaging 1.08, which we take as our empirical target.[33]

5.3 Implementation: A Gaussian Objective Function

Loosely speaking, the indirect inference estimator is obtained by choosing the values of the structural parameters so that the estimated model and the US data "look as similar as possible" when viewed through the lens of the auxiliary model. More concretely, define

$$\epsilon_t^{i,Data} \equiv \left[c_t^{i,Data} - \mathbf{a}' \mathbf{X}_{c,t}^{i,Data}, \; y_t^{i,Data} - \mathbf{b}' \mathbf{X}_{y,t}^{i,Data} \right]$$

rate variables differently: the past growth rate for age t in the auxiliary model is computed by taking the difference between the latest valid observation before t and the first valid observation for the individual in the data set and dividing this difference by the number of years between the two points. 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 over the life cycle instead.

[32]Our wealth measure includes cash and demand deposits; time and saving deposits, CDs and money market accounts; stocks, bonds, and mutual fund holdings (including in IRAs); cash surrender value of life insurance policies; net equity in unincorporated businesses; and net equity in owner-occupied housing and other real estate. From the sum of these assets we subtract consumer debt (credit card debt, student and auto loans, etc.). The income measure is total household labor and asset income in that year.

[33]When we include all households up to age 65, the corresponding ratios are 1.19 and 1.43.

to be the residuals of estimated equations (22) and (23), which is understood to equal zero when data for household i in year t is missing. The superscript *Data* specifies the data source used in the regression, which is either the PSID or the structural model (indicated by SIM). The objective function we use is

$$\mathscr{L}(\mathbf{a}, \mathbf{b}, \Sigma, Data) = |\Sigma|^{-J/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{2235} \sum_{t=1968}^{1993} \boldsymbol{\epsilon}_t^{i,Data} \Sigma^{-1} (\boldsymbol{\epsilon}_t^{i,Data})' \right), \tag{24}$$

where J is the total number of household-year observations used in the regressions (26,411 in the baseline estimation). Although the objective function is in the form of a multivariate Gaussian density, it is not, strictly speaking, the likelihood of the auxiliary model regressions (22) and (23). This is because, these equations have as regressors both the past *and* the future values of endogenous variables, which makes it impossible to obtain the proper likelihood by conditioning on past observations (or the future separately). Thus, to avoid a confusion of terminology, and for lack of a better term, we shall refer to $\mathscr{L}$ as a "Gaussian objective function."

To implement the estimator, we first maximize the objective function in (24) using real data (i.e., from the PSID) to obtain a set of reduced-form parameters, denoted by $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}}, \widehat{\Sigma})$. Next, we follow a similar procedure using simulated data. The vector of structural parameters that we estimate is

$$\boldsymbol{\theta} \equiv (\sigma_\alpha, \sigma_\beta, corr_{\alpha\beta}, \rho, \sigma_\eta, \sigma_\varepsilon; \lambda, \sigma_y, \mu_c, \sigma_c, \sigma_{c_0}; \delta, \psi).$$

For a given $\boldsymbol{\theta}$, simulate a data set from the structural model that matches exactly the number of observations and missing data pattern found in the PSID data set, and estimate (22) and (23), which yields $\tilde{\mathbf{a}}_1$, $\tilde{\mathbf{b}}_1$, and $\widetilde{\Sigma}_1$. Now, using a fresh sequence of random draws (for all the stochastic elements in the structural model), repeat the same procedure to obtain $\tilde{\mathbf{a}}_2$, $\tilde{\mathbf{b}}_2$, and $\widetilde{\Sigma}_2$. Repeat this N^{SIM} times and construct the following averages: $\tilde{\mathbf{a}} = \frac{1}{N^{SIM}} \sum_{n=1}^{N^{SIM}} \tilde{\mathbf{a}}_n$, and analogously for $\tilde{\mathbf{b}}$ and $\widetilde{\Sigma}$. Then, we use these averaged parameter values—estimated from *simulated* data—to evaluate the objective function (24) using the observed (PSID*)* data: $\mathscr{L}(\tilde{\mathbf{a}}(\boldsymbol{\theta}, SIM), \tilde{\mathbf{b}}(\boldsymbol{\theta}, SIM), \tilde{\Sigma}(\boldsymbol{\theta}, SIM), PSID)$. If the simulated data looks *exactly* like the PSID data—in the sense that the estimated auxiliary model parameters for the two data sets are identical—then the two objective values would be identical; otherwise the Gaussian objective function will always be higher when evaluated at $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}}, \widehat{\Sigma})$ than $\left(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\Sigma}\right)$, because the latter does not maximize (24) with PSID data (but instead with simulated data). Finally, the indirect inference estimator is defined as

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left[\mathscr{L}(\widehat{\mathbf{a}}, \widehat{\mathbf{b}}, \widehat{\Sigma}, PSID) - \mathscr{L}(\tilde{\mathbf{a}}(\boldsymbol{\theta}, SIM), \tilde{\mathbf{b}}(\boldsymbol{\theta}, SIM), \tilde{\Sigma}(\boldsymbol{\theta}, SIM), PSID) \right]. \tag{25}$$

In effect, our indirect inference estimator maximizes the Gaussian objective function associated with the auxiliary model subject to the "cross-equation" restrictions that the structural model imposes on its parameters. An important advantage of this estimator is that it obviates the need to estimate an optimal weighting matrix; obtaining precise estimates of such matrices is often difficult. Instead, our estimator uses an *implicit* weighting matrix that is close to optimal (to the extent that the

auxiliary model is close to being correctly specified) and delivers very good small sample results. In particular, in Appendix C.2, we show that this estimator is asymptotically equivalent to one that minimizes a quadratic form in the difference between the auxiliary model parameters calculated using the observed and simulated data, with the weighting matrix being the optimal one if the auxiliary model were actually correctly specified. This weighting matrix is not optimal here (since the auxiliary model is not an exact "reduced form" for the structural model), but our Monte Carlo analysis in the next section demonstrates that we obtain excellent results, with little bias and small standard errors.

5.4   Monte Carlo Analysis

To investigate the ability of the proposed estimation method to uncover the true structural parameter vector with the specified auxiliary model, we conducted a detailed Monte Carlo study. To save space, here we briefly summarize the main findings, and present a more detailed set of results in Appendix B. As discussed in the appendix, we experimented with several different auxiliary models and found that the one described previously delivered the best results in terms of bias and precision of estimates.

Two main conclusions emerge from those analyses. One, the estimation method works very well: bias is virtually absent for most parameters and is very small for the remaining few. Standard deviations are very small, indicating that all the parameters, with the exception of $corr_{\alpha\beta}$, can be pinned down fairly precisely. Two, the income process parameters can be estimated using income data alone without any noticeable bias. However, under some plausible parameter combinations, the precision of the estimates of some key variables is significantly higher when estimated from income data alone (e.g., the standard error on σ_β goes from 0.17 to 0.27, and for $corr_{\alpha\beta}$ it goes from 0.15 to 0.29). This reduced precision makes it harder to separate whether the rise in income inequality is coming from σ_β or from σ_α through the strong correlation.

We conclude that although it is difficult, if not impossible, to prove identification in this general setup, this Monte Carlo analysis suggests 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 consumption-saving models with widely missing data, binding borrowing constraints, and multiple sources of heterogeneity and randomness.

6   Estimation Results

We now estimate the life cycle model using the PSID household after-tax labor income data and the imputed consumption data. Table I reports the results. Standard errors are obtained via parametric bootstrap. This is a very time-consuming process (taking more than three months for each configuration on a state-of-the-art workstation), so we only report standard errors for the benchmark estimates (columns 1 and 2).

Parameters of the Income Process: $\rho, \sigma_\eta, \sigma_y \sigma_\alpha, \sigma_\beta, \text{corr}_{\alpha,\beta}$.   The first column reports the results from our benchmark estimation (where the values of the pre-set parameters have been described above). The AR(1) process has an estimated annual persistence of 0.755 and an innovation standard

Table I: Estimating the Full Consumption-Savings Model

	Baseline	Income data only	Low interest rate $\gamma = 0.97$	Alternative filling-in method	Doubling minimum income $\underline{Y} = 0.10$	Use data up to age 65	No prior uncertainty $\lambda = 0$	Maximum prior uncertainty $\lambda = \lambda^{\max}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income Processes Parameters (can be identified with income data alone)								
σ_α	0.284	0.298	0.284	0.220	0.293	0.228	0.268	0.248
	(0.027)	(0.038)						
σ_β	1.852	1.343	1.856	1.916	1.886	1.088	1.756	1.04
	(0.188)	(0.271)						
$corr_{\alpha\beta}$	-0.162	0.558	-0.164	0.003	-0.166	-0.161	-0.086	0.751
	(0.151)	(0.289)						
ρ	0.754	0.783	0.755	0.760	0.759	0.801	0.777	0.806
	(0.025)	(0.022)						
σ_η	0.196	0.200	0.196	0.200	0.208	0.200	0.195	0.196
	(0.005)	(0.005)						
σ_ε	0.004	—	0.005	0.006	0.007	0.003	0.006	0.010
	(0.021)							
Economic Model Parameters (need consumption data)								
λ	0.345	—	0.380	0.327	0.374	0.520	0.0 (fixed)	0.656^a
	(0.074)	—						
δ	0.950	—	0.964	0.950	0.951	0.943	0.954	0.951
	(0.001)	—						
ψ	0.874	—	0.790	0.921	0.757	0.992	0.761	0.895
	(0.083)	—						
Measurement Error and Transitory Shocks (need consumption data)								
σ_y	0.147	0.147	0.147	0.145	0.156	0.152	0.148	0.151
	(0.007)	(0.005)						
σ_c	0.356	—	0.356	0.356	0.356	0.356	0.355	0.356
	(0.002)	—						
σ_{c_0}	0.428	—	0.429	0.414	0.427	0.432	0.424	0.433
	(0.009)	—						
max % constrained..	12.7%	—	16.1%	10.2%	12.5%	9.1%	14.1%	13.2%
.. at age	30	—	30	31	30	35	30	33
$\underline{a_0}$ / mean income	0.33	—	0.35	0.44	0.41	0.93	0.21	0.37
$\underline{a_{55}}$ / mean income	0.53	—	0.60	0.59	0.87	0.73	0.44	0.55

^{a}This value of λ represents the highest value feasible (i.e., highest prior uncertainty) given that households know their α^i, which is correlated with β^i and thus contains information about it.

deviation of 19.6%, both estimated precisely. Below, we present a wide range of sensitivity analyses that show that these estimates are remarkably robust. These results, hence, lead us to conclude that the joint dynamics of consumption and income data do not lend support to permanent shocks as a reasonable representation of the typical income shock. Having said that, these results should not be interpreted as evidence against the existence of permanent shocks—as clearly there are such shocks. Rather, they indicate that the *typical* shock received by a *typical* household is better represented as having a modest persistence.

Turning to the fixed heterogeneity across households, first, the cross-sectional standard deviation of the fixed (level) effects in income, σ_α, is 0.284, implying that a household who is one (two) standard deviation above the mean earns 76% (211%) more than another household who is one (two) standard deviation below the mean. Second, the standard deviation of income *growth rates*—a key parameter of interest—is estimated to be 1.84%, which is substantial. For example, by age 55, an individual with a β^i one (two-) standard deviations above the mean will earn 1.77 times (3.16 times) the median income. Moreover, both σ_α and σ_β are estimated precisely, with t-statistics of approximately 10. For comparison, studies that use income data alone report t-statistics in the range of 4 to 5 for σ_β (see, e.g., Baker (1997), Haider (2001), and Guvenen (2009)). Third, the correlation between these two kinds of fixed heterogeneity ($corr_{\alpha\beta}$) is small but negative, –0.16, though it cannot be statistically distinguished from zero. We can, however, easily reject a strong correlation of either sign.

Systematic differences across individuals in income growth rates and a negative correlation of this growth rate with initial income are both predicted by plausible variants of the human capital model. Thus, these relationships have received significant attention since the early days of the human capital literature (see, e.g., Mincer (1974) and Hause (1980)). One implication of a negative correlation is that individuals with low initial income will catch up and pass those with high initial income (at what is referred to as the "crossover" or "overtaking" point), a phenomenon these early studies document. Our results bring fresh corroborating evidence by jointly utilizing consumption and income data.

Estimation Using Only Income Data. Column 2 reports the estimates of the income process obtained by using only the income regression (23). Compared to the benchmark estimates, the main noticeable differences are in the estimates of σ_β and $corr_{\alpha\beta}$. The remaining parameters are very similar across the two cases, suggesting that they are pinned down very well with income data, and that there is no conflicting information in consumption data that pulls these estimates away from those obtained using income data alone. As for σ_β, it goes down fairly significantly compared to the benchmark case. An important pitfall to avoid is to view σ_β as the only determinant of the rise in income inequality. To show this, Figure 8 plots the variance of log income computed using the benchmark estimated model (marked with squares), as well as the one using the present estimates from income data only (marked with plus signs). As seen here, if anything, income inequality increases slightly more under the latter estimates, because now $corr_{\alpha\beta} = 0.55$, to be compared with –0.16 in the benchmark case. Finally, notice that the standard errors of these two parameters are significantly higher when estimated with income data alone, suggesting that the two parameters are not identified very precisely with income

data alone. Consumption data is especially informative about these two parameters.

Parameters Identified with Consumption Data: $\boldsymbol{\lambda, \delta, \psi}$. Of course, not all of the heterogeneity in income growth rates represents *uncertainty*, because each individual has some prior information about his true β^i by the time he enters the labor market. The parameter λ—which measures the degree of this prior uncertainty—is estimated to be 0.345, which reveals only a small amount of prior uncertainty regarding individuals' growth rate. Using the notation developed in Section 2.1, the predictable (by the households) component of income growth represents a substantial fraction of the total dispersion in the population: $\sigma_{\beta_k}^2/\sigma_\beta^2 = 1 - \lambda^2 = 0.880$. Although small, the remaining uncertainty is still economically nonnegligible, as we show in the next section.

The estimated time discount factor is 0.9496, which implies $\delta\gamma \approx 1$. However, because average income is growing over the life cycle, the average individual is impatient—in the sense of Deaton (1991)—according to these estimates. Although the standard error on δ is extremely small (0.001), this is *conditional* on the fixed value of ϕ: unfortunately, the estimate of δ is quite sensitive to the preset value of risk aversion. To determine whether if this sensitivity could undermine our results about income risk, in Section 6.3 we conduct a detailed sensitivity investigation.

Finally, ψ, which measures the tightness of the borrowing constraint, is estimated to be about 0.874. This estimate implies that the effective interest rate applied to discount households' future labor income for determining the borrowing limit is $(1+r)/\psi - 1 \approx 20.5\%$, which is a premium of 15.24% over the risk-free interest rate. Another useful measure is the maximum debt that a household is allowed to carry as a fraction of average income: $\underline{a}_t/\mathbb{E}(Y^i)$. For a household whose head is 25 years old, this limit is 33% of (the economy-wide) average income, and it rises to 53% at age 55.[34] Notice that these estimates are consistent with a relaxation of borrowing constraints as households get older. Finally, in the estimated model, between 6.5% and 12.7% of households are up against their borrowing constraint between ages 30 and 40. The percentage who are constrained falls to about 5% or less beyond age 40.

Measurement Errors and Transitory Shocks: $\boldsymbol{\sigma_\varepsilon, \sigma_y, \sigma_c, \sigma_{c_0}, \mu_c}$. With consumption data, in principle, we can tell transitory shocks apart from i.i.d. 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 nonnegligible fraction of households. As a result, these households' consumption would move one-for-one with transitory shocks, allowing us to distinguish these shocks from pure measurement error. We estimate the standard deviation of the transitory measurement error in income to be about 14.7% annually, whereas true i.i.d. shocks to income have a very small

---

[34]Notice that the standard error on the estimate of ψ is rather large (0.096). The implied 95% confidence intervals for the borrowing constraints are wide at age 25 (ranging from 16% to 58% of average income), but narrow by age 55, becoming 40% to 65% of average income.

standard deviation—less than 1% annually.[35] Finally, the transitory measurement error in consumption has a standard deviation of 35.5% and includes the noise introduced by the imputation method. Furthermore, the fixed effect in measured consumption has a standard deviation of 42.8%, and both components are estimated with extremely high precision.[36]

6.1 Goodness of Fit: Benchmark Estimated Model

6.1.1 Auxiliary Model

To investigate the fit of the model, we begin with the auxiliary model coefficients estimated from PSID data and those generated by the benchmark estimates of the structural model. Table II displays the (50) coefficients of interest. Throughout the table, a boldface font indicates that the coefficient is highly significant (a *t-statistic* of 5.0 or more).[37] Similarly, superscripts (**) and (*) indicate significance at 1% and 5% levels, respectively. Starting at the first row—which reports the data coefficients for the income equation of the young group—all regressors, except for the last two, are very highly significant. The next row displays the corresponding coefficients obtained from the simulated model. For each coefficient, we conduct a hypothesis test that the model and data coefficients are equal to each other. If the test is rejected with a t-statistic greater than 5, this is indicated with the superscript († † †). Similarly, rejections at the 1% and 5% levels are indicated with superscripts (††) and (†), respectively.

In the income equation, the leads and lags of the income variable, as well as the past and future average income, are almost always very significant (with t-statistics typically greater than 30). The coefficients from the simulated data match their data counterparts well: for the 11 coefficients in this regression that are very significant (in bold), one cannot reject equality between the model- and data-implied coefficients at 5% level or higher. Further, for only 3 out of the 18 coefficients in this regression can one reject that the model and data coefficients are statistically different. Overall, the estimated structural model does a fairly good job of matching the auxiliary model coefficients in the income equation.

The consumption regression (22) is more challenging in many ways: The left-hand-side variable is now the output of a complex structural model that filters through the income shocks into the consumption choice. Moreover, consumption data contain not only measurement error (as did income data) but also imputation error. Inspecting the coefficients, we see that the lagged income term is

---

[35]The very small estimated value of σ_ε raises the question of whether this parameter cannot be disentangled from 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 A.4). Another point to keep in mind is that our model does not allow for any partial insurance (that is, informal smoothing opportunities) beyond self-insurance. To the extent that there is such partial insurance for transitory shocks (as in Blundell et al. (2008)), our estimate of σ_ε represents the size of the shock that gets *transmitted* to consumption, which may be small.

[36]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. In light of this, we have explored if relaxing the restriction $\mu_c = 0$ that we imposed in the benchmark estimation could matter for the parameter estimates. With μ_c unrestricted, we estimated it at 0.04. However, no other parameter estimate had a noticeable change from the baseline, so we decided to restrict it to zero in the rest of the paper.

[37]We compute the standard errors using a nonparametric bootstrap (sampling individuals with replacement).

Table II: Coefficients of the Auxiliary Model: Benchmark Estimated Model vs. US Data

Panel A: Income Equation

	constant	y_{t-1}	y_{t-2}	y_{t+1}	y_{t+2}	$\bar{y}_{1,t-3}$	$\bar{y}_{t+3,T}$	$\Delta y_{1,t-3}$	$\Delta y_{t+3,T}$	c_{t-1}	c_{t-2}	c_{t+1}	c_{t+2}
Young Group:													
(1) Data	**-0.036**[a]	**0.346**	**0.360**	**0.077**	**0.097**	**-0.098**	**0.150**	0.037	-0.022				
(2) Model	-0.026††[b]	0.337	0.357	0.093	0.102	-0.081	0.124	-0.016††	-0.021				
Middle-Age Group:													
(3) Data	0.006**	**0.418**	**0.358**	**0.111**	**0.093**	-0.027	0.043**	0.031	0.028				
(4) Model	-0.001††	0.401	0.375	0.122	0.105	-0.038	0.022	0.012	0.045				

Panel B: Consumption Equation

	constant	y_{t-1}	y_{t-2}	y_{t+1}	y_{t+2}	$\bar{y}_{1,t-3}$	$\bar{y}_{t+3,T}$	$\Delta y_{1,t-3}$	$\Delta y_{t+3,T}$	c_{t-1}	c_{t-2}	c_{t+1}	c_{t+2}
Young Group:													
(5) Data	-0.007	**0.108**	0.042*	-0.023	-0.005	-0.045*	-0.017	0.030	-0.002	**0.248**	**0.262**	**0.178**	**0.175**
(6) Model	-0.021††	0.105	0.079†	-0.022	-0.039	-0.079	0.025	0.001	0.013	0.208††	0.205†††	0.248†††	0.229†††
Middle-Age Group:													
(7) Data	-0.004	**0.136**	0.046**	-0.014	-0.040*	-0.082**	0.012	0.030	0.028	**0.270**	**0.260**	**0.177**	**0.187**
(8) Model	0.007††	0.103	0.041	0.009	-0.044	0.050†††	-0.042††	0.027	0.041	0.197†††	0.211††	0.224††	0.257†††

Panel C: Residual Variances and Correlations

	Young Group			Middle-Age Group		
	$\sigma^2(\epsilon_t^y)$	$\sigma^2(\epsilon_t^c)$	$\rho(\epsilon_t^y,\epsilon_t^c)$	$\sigma^2(\epsilon_t^y)$	$\sigma^2(\epsilon_t^c)$	$\rho(\epsilon_t^y,\epsilon_t^c)$
(9) Data	**0.222**	**0.396**	**0.117**	**0.235**	**0.379**	**0.114**
(10) Model	0.226	0.392	0.144††	0.232	0.389	0.120

[a] A bold font indicates that the estimated parameter has a t-statistic greater than 5.0; superscript (**) indicates a t-statistic greater than 2.57 (significant at 1%), and (*) indicates a t-statistic greater than 1.96 (significant at 5%). The largest t-statistic for parameter estimates is 34.5, and for the residual variances it is 67.81.

[b] Parameters estimated using simulated data are marked with one, two, three (†) signs in the superscript if the t-statistic for the hypothesis test of equality between simulated and data coefficients is greater than 1.96, 2.57, and 5.0, respectively. The largest t-statistic for these hypothesis tests is 6.56.

very highly significant (for both age groups), and the model matches them quite well with no evidence of a statistical rejection. Another set of coefficients that are also very significant are the leads and lags of consumption. Here, we can reject the equality of model coefficients to their empirical counterparts fairly easily. It is clear that the estimation procedure tries to bring these coefficients as close together as possible (as evidenced by the fact that the model coefficients are of the right order of magnitude and always have the correct sign), but it seems that the model lacks a key ingredient that drives a wedge between the simulated and data coefficients. One conjecture is that some sort of temporal dependence in the utility function, such as habit formation, could help resolve this problem by increasing the correlation of consumption at lower lags compared to the benchmark model. Nevertheless, exploring alternative utility function specifications is beyond the scope of this paper and is left for future work. Among the remaining coefficients, there are five that are significant at the 5% level or higher, and for two of these, the model coefficients do not match up, whereas for the other three they do.

Panel C displays the variance-covariance matrix of the regression residuals, whose elements are very highly significant, with the t-statistics exceeding 60 for variances and 10 for correlations. Consequently, the Gaussian objective function puts significant weight on these terms. Out of the six elements in panel C, the estimated model fails to match only one of them (the correlation of residuals across equations for the young group).

Finally, we conduct a formal model specification test. A natural test statistic is the difference between the data (PSID) and estimated model objective values, whose distribution is easily obtained via bootstrap.[38] Under the null hypothesis that the estimated model is the true data generating process, this test statistic has a mean of 2.29, with a *maximum* value—across the 100 replications—of 4.03. Computing the same statistic using the benchmark estimated model and PSID data yields a value of 26.37, indicating a strong rejection, despite the fact that, as shown in Table II, the auxiliary model coefficients in the actual and simulated data sets are quantitative close to each other.

To sum up the findings so far, the estimated structural model matches several very significant coefficients of the auxiliary model quite well, but also falls short in matching the coefficients on lagged and future consumption, resulting in a statistical rejection. For most practical purposes, though, the more relevant question is: How well does the estimated model replicate some economically interesting moments? This question is of special interest here, because the mapping between the auxiliary model coefficients and the economically relevant moments is not transparent. We turn to this task in the next section.

[38]More specifically, we first generate a simulated data set from the structural model by setting the parameter values to those obtained in the actual benchmark estimation. Call this the "real" data set. We evaluate the Gaussian objective function ($\mathscr{L}$) using this real data set. Then, using a new set of seeds for the random number generators, we simulate a new data set and estimate the parameters that must have generated this newly simulated data. We reevaluate $\mathscr{L}$ using these estimated parameters and the real data simulated in the first step. We repeat this second step a large number of times, which gives us a probability distribution for the test statistic under the null hypothesis that the "real" data are generated from the estimated structural model.

6.1.2 Income and Consumption over the Life Cycle: Model vs. Data

We evaluate the fit of the estimated model to the data along three dimensions that are important benchmarks in the incomplete-markets literature. The first two are the evolution of the within-cohort variance of log income and log consumption over the life cycle. The third one is the average life cycle profile of log consumption. (The average life cycle profile of income is matched by design.) However, there are two distinct ways to perform the model-data comparison, each with its own valid interpretation. The difference arises from the fact that the data used in estimation contain (potential) cohort effects in variances.[39] As a result, when we plot the life cycle inequality profiles of consumption and income computed from the data used in estimation, these figures do not correspond to the well-known graphs reported in the literature (such as in Deaton and Paxson (1994), Storesletten et al. (2004a), and Guvenen (2009), who remove these cohort effects before constructing their figures). Nevertheless, for testing the fit of the model to the data—as an econometrician would do—these uncleaned data variances *are* the right empirical counterpart, since they are computed directly from the data used in the estimation. Furthermore, for a valid comparison, the statistics must be constructed for a sample size that matches the number of individuals in the PSID as well as the missing data pattern observed. Because these three graphs are not directly targeted in the estimation, they provide useful overidentifying restriction tests of the model.

A second perspective is that of a quantitative economist who is interested in using the structural estimates for calibrating a life cycle model and wants to know how the model-generated inequality figures will compare to the stylized features of life cycle income and consumption behavior (i.e., those in a stationary environment). For this purpose, the relevant comparison is to statistics from data *cleaned* of cohort effects. Furthermore, it now makes more sense to construct the model statistics using a large number of simulated individuals. Given these two valid perspectives depending on a researcher's goals, we conduct the analysis in both ways and discuss when the difference matters.

First, Figure 8 plots the variance of log income and consumption using our PSID estimation sample. The income variance (marked with diamonds) increases by only 13 log points over the life cycle, less than half the rise of 25–35 log points or more reported in several studies (see, e.g., Storesletten et al. (2004a) and Heathcote et al. (2010)). As noted before, the difference arises from the lack of adjustment here for cohort effects, a point that we discuss at greater length in Section 7. As for the model counterpart, the simulated statistics do not match the level of their data counterparts exactly. For example, at age 25, the income variance is 1.6 log points lower in the model than in the data, which is a relatively minor difference, whereas the consumption variance is 12.7 log points higher, which is a much bigger difference.

To facilitate the comparison of *trends* over the life cycle, we shift the model variance graphs vertically so as to match the data variances at age 30. Now looking at the evolution of income

---

[39]That is, even though we take out cohort effects from the *levels* of consumption and income, there are likely to be cohort-specific differences in the *variances* of these variables, which are not removed because there is no easy way to do so. Although these cohort effects could be introduced into the structural model (e.g., by making σ_α to be cohort-specific), this would introduce at least a dozen additional parameters into the estimation.

Figure 8: Life Cycle Inequality: PSID Estimation Sample vs Estimated Structural Model
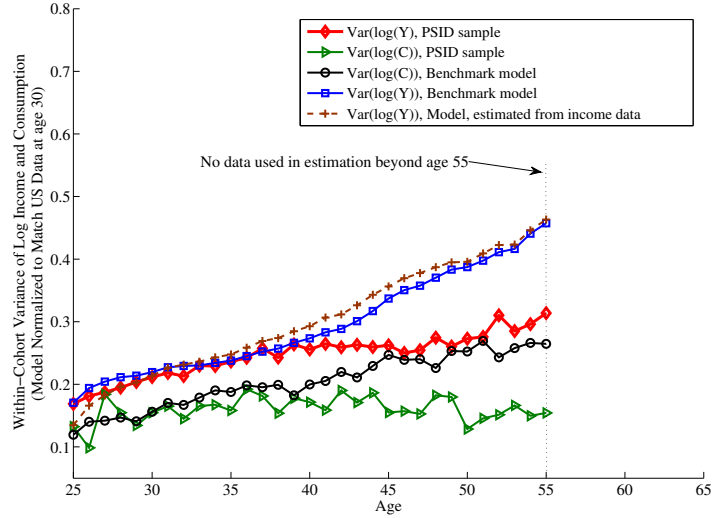

inequality, the model agrees with the trend in the data quite well up to age 40. But inequality continues to grow in the model, whereas it tapers off in the data. Interestingly, the variance of log consumption figures reveal a similar pattern: consumption inequality in the data rises slightly up to age 40, matched quite well by the model. But inequality continues to grow past that age in the model, whereas it stagnates and falls slightly in the data.[40]

Finally, Figure 9 plots the average life cycle profile of consumption, calculated from our PSID estimation sample (line marked with diamonds). Because in the estimation we are not using household data beyond age 55, the graph after this age should not be used for testing the goodness of fit (but will be relevant for the calibration comparison later). The estimated model tracks the empirical variance until about age 45, after which point average consumption in the data falls significantly, whereas the model counterpart simply tapers off.

Overall, the model does a good job of matching some salient aspects of life cycle income and consumption patterns before age 45, despite the fact that these do not appear as explicit moments in the estimation procedure. The fit of the model is less satisfactory after age 45, partly because we use data only up to age 55 (more on this later).

[40]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 et al. (2004a) and Guvenen (2007) who generated rises as large as 35 log points. This is because, 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 capture 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.

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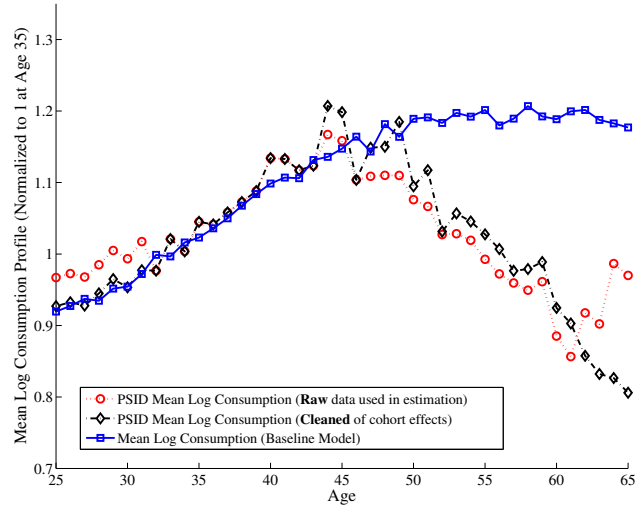

6.2 Robustness

Despite many appealing features of simulation-based structural estimation methods, a potentially important drawback is that the outcome of the estimation procedure depends *potentially* on all the assumptions made on the structural model. In contrast, GMM has the advantage that any feature of the model that does not appear explicitly in the moment conditions has no relevance for the estimation results. Consequently, it is essential to examine whether the parameter estimates we obtain are sensitive to the key features of the model that have been fixed in advance. We now vary some of these choices we made and examine their impact on the estimates.

Lower Interest Rate. In our benchmark, we interpreted the risk-free asset as corresponding to a broad set of assets available to households, which motivated our relatively high choice of $r = 5.26\%$. Another perspective is that such an asset can be thought of as a government bond, so that a lower return may be more appropriate. To explore the sensitivity of our results, we reestimate the model, setting $r = 3.1\%$ (i.e., $\gamma = 0.97$). As seen in column 3, this change has virtually *no* effect on the estimates of the income process as well as measurement error parameters. As for the economic model parameters, λ increases—slightly—from 0.345 to 0.38, whereas δ increases significantly (from 0.95 to 0.964) and borrowing constraints become tighter—ψ falls significantly, from 0.87 to 0.79—both presumably to make the model match the wealth-to-income ratio moment better. Overall, the most important parameters about the income process as well as λ seem very robust to reasonable changes in the interest rate.

Alternative Filling-in Method. We now examine the robustness of the estimation results to the method chosen for filling in missing data (column 4). This could be potentially important because more than half of the values in our sample are missing—and therefore filled in—compared to a fully

Figure 10: Inequality: PSID vs Structural Model Estimated Up to Age 65
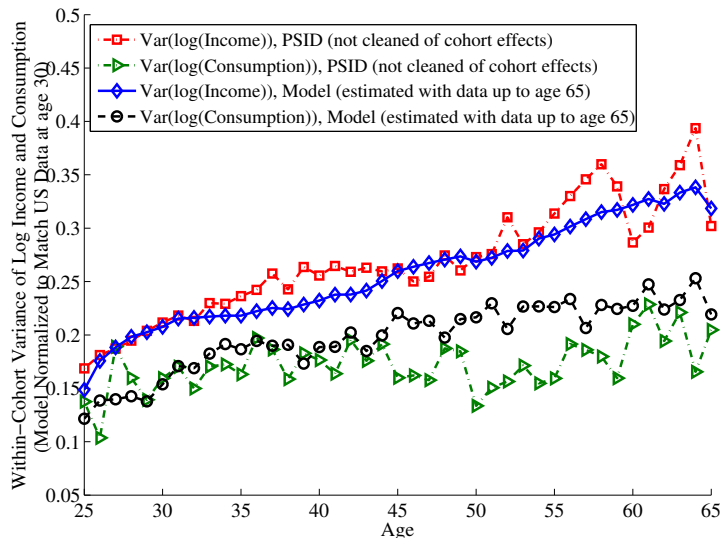
balanced panel. As an alternative procedure, we consider a much simpler filling-in method: for each individual we calculate the lifetime average of either log consumption or log income using available observations. If a consumption or income observation is missing in a given year, we simply replace the missing data with this average. We then use this filled-in data to construct all the missing right-hand-side variables in the regressions. As seen here, with the exception of σ_α, the estimates are largely unchanged from the benchmark case.

Higher Minimum Income. We now investigate the sensitivity to the choice of minimum income $\underline{Y}$, by doubling its magnitude to 10% of median income. The effects on the estimates are very mild, with the only noteworthy changes being a rise in λ, from 0.345 to 0.375, and a fall in ψ from 0.874 to 0.756. However, because $\underline{Y}$ has been doubled, the borrowing constraint is actually looser than before: $\underline{a}_0$ rises to 41% of average income, compared with 33% in the benchmark case.

Using All Available Data Up to Age 65. In the estimation so far, we have restricted our sample to ages 55 or younger, for several reasons.[41] Still, it is useful to examine how the results would change if the entire sample (up to age 65) were used in estimation. As seen in column 6 of Table I, some parameters change very little, whereas other important ones do change. For example, σ_β falls to 1.08 and λ simultaneously rises to 0.51, implying that the amount of prior uncertainty about growth rates (as measured by the prior standard deviation) falls slightly from $1.85 \times 0.345 = 0.638$ to $1.08 \times 0.51 = 0.55$. Figure 10 plots the fit of the estimated model to the PSID data, now extended

---

[41]One is that our assumption of linearity for the individual-specific trend is more likely to be accurate for households before this age, as widening income inequality slows down near retirement. Second, labor hours inequality increases near the retirement age, which weakens the link between wage and income inequality. Given that we are abstracting away from labor supply choice here, it seems more appropriate to restrict attention to the earlier period. Finally, the number of individuals in our sample goes down quickly at older ages, increasing the noise and reducing the usefulness of data from this group.

up to age 65. Now both the variance of log income and consumption graphs fit much better to the data. It seems that the deviations of the variance of log income and consumption figures from their data counterpart were "tolerable" as viewed through the auxiliary model, when data up to age 55 were used in estimation. But the inequality profiles implied by $\sigma_\beta = 1.85\%$ deviate farther from the data after age 55 at an increasing rate. This leads the estimator to reduce σ_β as well as $\lambda\sigma_\beta$, which then results in a better fit for both graphs. It is interesting to see that even though these moments have not been used in the estimation explicitly, matching the auxiliary model coefficients somehow ensures that the estimated model does a reasonable job of matching these economically important figures.

Fixing Prior Uncertainty: Two Polar Cases.   One of the key findings of our benchmark estimation is the low estimated value of λ, indicating a small degree of prior uncertainty. Given that solving a consumption-savings model with Bayesian learning (as we do here) is more complicated than solving without, it is natural to ask if simply setting $\lambda = 0$ (thereby eliminating learning completely) distorts the implications of the model significantly. A second polar case that deserves attention is one where λ is as high as theoretically possible. This is the case where the only prior information available to the individual is α^i (which is informative as long as $corr_{\alpha\beta} \neq 0$). We denote the corresponding degree of maximum prior uncertainty as $\lambda^{\max}$. When $\lambda = \lambda^{\max}$, we are as close as one can get in this framework to the default assumption in the existing literature with RIP processes, which holds that all of the rise in income inequality represents risk.

Columns (7) and (8) of Table I report the estimation results for these two cases. The first point to note is that the estimates with $\lambda = 0$ are broadly in line with the benchmark case, with a few exceptions: δ is higher (0.954 here vs. 0.950 in the baseline), σ_β is somewhat smaller (1.77% vs. 1.85% in the baseline), and perhaps most importantly, $\psi = 0.76$ vs. 0.87 in the baseline. It seems that the estimation compensates for the reduction in risk coming from a lower λ by raising the time discount factor and tightening the borrowing constraint. But, overall, these estimates are quite consistent with the benchmark case.

Turning to column 8, the picture is quite different: now σ_β is down to 1.04% from 1.85%, and $corr_{\alpha\beta}$ is 0.75, up from –0.16. The value of $\lambda^{\max}$ implied by this correlation is 0.656. It is interesting to note that the amount of prior uncertainty—as measured by $\lambda\sigma_\beta$—in this case is $0.656 \times 1.04\% = 0.68$, which is very close to the value in the benchmark case: $0.345 \times 1.85\% = 0.64$! Thus, it seems that the estimation procedure tries to undo the restriction $\lambda = \lambda^{\max}$ (which overstates the amount of uncertainty), by reducing σ_β, in turn keeping the total amount of uncertainty unchanged. This finding suggests that the amount of prior uncertainty is actually quite well pinned down and that the consumption regression in the auxiliary model contains useful information to estimate it.

Overall, we interpret these results as providing some justification for a researcher to abstract away from Bayesian learning and calibrate the income dynamics parameters of his/her economic model with the parameters estimated in column 7. The other polar case seems harder to justify and does not have any computational advantages over the benchmark estimated model.

6.3 On Risk Aversion, Time Preference Rate, and Borrowing Limit

Although it would be valuable per se if we could separately identify δ and ϕ, this is not the central aim of this paper. But the assumptions we make regarding these parameters (e.g., whether they are fixed, estimated, etc.) still matter, because they could affect the inference regarding income risk—potentially seriously—in turn jeopardizing the main goal of our investigation (as shown in Section 2.3). Therefore, we begin by conducting a series of sensitivity experiments to understand the effects of preference parameters on the overall estimation results. To this end, we first reestimate the benchmark specification, but now fix the risk aversion at, respectively, 1 (column 1 of Table III) and 3 (column 2). A quick glance across these two columns and column 1 of Table I reveals two findings. First, δ and ϕ move strongly in opposite directions: δ goes up to 0.9526 when ϕ is reduced to 1, and goes down to 0.9416 when ϕ is increased to 3. Second, and fortunately, the remaining—11—parameters are *virtually unchanged* from the benchmark case, a quite striking finding. We now discuss these two sets of results in turn.

6.3.1 Are δ and ϕ Separately Identified?

The strength of the opposite movement in δ and ϕ (as we vary the latter) is remarkable. In fact, the correlation between the two estimates is worth reporting: –0.97! However, a correlation with three data points (one for each of the three values for risk aversion) is, obviously, not as informative as one would like, so we conducted a simple Monte Carlo study where we fixed all the parameter values except δ and ϕ, which are estimated. Across 200 repetitions, the correlation between the estimates of the two parameters was –0.88. The results suggest that only a particular combination of the two parameters is identified, but that there is insufficient information to disentangle the two. In unreported results, we have tried adding additional regressors into equations (22) and (23), as well as adding new regressions suggested by theory, such as the second order moments of consumption growth or levels (computed in various ways) to capture precautionary savings demand, which could be informative about ϕ. In every case, we found the same strongly negative correlation.

Given how strongly this result manifests itself in our framework, we turn to another paper, Gourinchas and Parker (2002), which estimates the same two parameters also jointly using income and consumption data. There are several important differences between our paper and theirs, leaving open the possibility that there could be different sources of identification in that paper that perhaps could overcome the difficulty we face. These authors report 18 different estimates of δ and ϕ across four classes of experiments: (i) their baseline estimation using robust and optimal weighting matrices (two results), (ii) estimates for five education groups, (iii) estimates for four occupational groups, and finally (iv) estimation results for seven different robustness exercises. In most cases, the standard errors are small, indicating that both parameters are estimated quite precisely. (Furthermore, notice that there is no particular reason for the estimates of δ and ϕ to be correlated across these 18 pairs if each parameter is precisely identified.) Nevertheless, the correlation between the 18 pairs of estimates turns out to be –0.989. Moreover, this result is not driven by a few outliers. For example, the five es-

Table III: Sensitivity of Structural Estimates to Restrictions on Economic Model Parameters

Preset Parameters↓	Role of Preference Parameters						Borrowing Limit	
	Low RRA	High RRA	δ fixed	δ fixed	δ estim.	δ fixed	$\psi = 1$	$\underline{a}_t \equiv 0$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ϕ (risk aversion)	1	3	1	1	1	1	2	2
δ (time discount factor)	estim.	estim	0.94	0.94	estim.	0.953	estim.	estim.
Weight on WY moment	10.0	10.0	1.0	0.0	0.0	0.0	10.0	10.0
Income Processes Parameters (can be identified with income data alone):								
σ_α	0.283	0.282	0.339	0.326	0.332	0.281	0.333	0.272
σ_β	1.838	1.841	3.997	2.165	2.093	1.850	2.129	1.747
$corr_{\alpha\beta}$	-0.161	-0.161	0.660	-0.243	-0.162	-0.161	-0.139	-0.101
ρ	0.756	0.756	0.821	0.724	0.738	0.760	0.750	0.765
σ_η	0.196	0.196	0.238	0.194	0.196	0.195	0.195	0.194
σ_ε	0.005	0.005	0.007	0.001	0.0124	0.005	0.029	0.005
Economic Model Parameters (need consumption data)								
λ	0.368	0.330	0.998	0.001	0.035	0.343	0.360	0.283
δ	0.953	0.942	0.94*	0.94*	0.938	0.953*	0.951	0.949
ψ	0.877	0.871	0.002	0.997	0.998	0.923	0.99*	$\underline{a}_t \equiv 0$
Measurement Error and Transitory Shocks (need consumption data)								
σ_y	0.147	0.147	0.129	0.147	0.146	0.147	0.144	0.149
σ_c	0.356	0.356	0.355	0.353	0.354	0.356	0.356	0.355
σ_{c_0}	0.427	0.429	0.514	0.420	0.412	0.426	0.433	0.426
Max % constrained	13.9%	10.5%	52.2%	21.7%	14.3%	12.2%	7.2%	38.1%
Max constr. age	31	33	25	35	35	33	35	25
Wealth-to-income ratio	1.08	1.08	0.94	1.03	0.15	1.03	1.08	1.07
$\underline{a}_0$ / mean income	0.33	0.32	0.07	0.90	0.90	0.44	0.83	0.0
$\underline{a}_{55}$ / mean income	0.53	0.53	0.31	0.73	0.73	0.59	0.71	0.0

[a]This estimated value is essentially at the upper bound that we have imposed on this parameter during estimation.

timates for different educational groups have a correlation of –0.978, whereas the correlation is –0.975 for the different occupation groups and –0.999 across the robustness results. (Throwing out two pairs from the last case that have very low estimates of δ yields the lowest correlation we obtain: –0.953.)

Based on these findings, we conjecture that the lack of identification between δ and ϕ may be endemic to the estimation of these parameters from consumption-savings models with fixed interest rates. We view these (admittedly negative) results as providing a challenge for future work to find ways to pin these two key parameters down precisely.

6.3.2 Does It Matter for the Estimates of Income Risk That ϕ Is Not Separately Identified?

As noted above, the core issue for this paper is the estimation of income risk. On this front, the news is more encouraging: all the parameters relating to income risk are robust to variations in risk aversion, which is reassuring. One question these results bring up is the following: Can we simply fix a reasonable combination of (ϕ, δ) (say, based on values commonly used in the literature) and estimate only the income process parameters, or is it important to estimate at least one of these two parameters as we have done so far?

To answer this question, we fix $\phi = 1$ and $\delta = 0.94$, and estimate the remaining parameters (reported in column 3 of Table III). These estimates are dramatically different from the benchmark values and appear very implausible. For example, σ_β is now 3.997 (which is essentially at the upper bound we imposed for computational reasons) and $corr_{\alpha\beta}$ is 0.66, implying an enormous rise in the variance of log income over the life cycle that is many times what is observed in the data; λ is now 0.687, which is at almost its highest theoretical value and implies that households perceive 2/3 of this *overestimated* rise in income inequality as risk/uncertainty. Moreover, 52% of households now appear borrowing constrained. By any measure, these estimates are quite extreme. To make things worse, we should note that here we are reporting the results of the estimation when the weight on the WY moment is reduced from 10.0 in the benchmark case to 1.0. If it were not for this change, the estimates would be even more extreme, with the remaining parameters also getting stuck at their bounds ($\rho = 0.9999$, etc.) Overall, this experiment illustrates how quickly things can go wrong if proper care is not applied. Thus, the specific values of (ϕ, δ) do not seem to matter for the other estimates *only* if we estimate at least one of those parameters.

One reaction could be that perhaps these extreme outcomes are the results of imposing the WY moment: because δ is fixed, the model cannot adjust this parameter to easily match the value of WY observed in the data and instead substantially increases both the amount of income risk (as roughly measured by $\lambda\sigma_\beta$) and the tightness of the borrowing constraint ($\psi = 0.02$). Even with these dramatic adjustments the model still undershoots the wealth-to-income ratio: 0.94 in the estimated model vs. 1.08 in the data. Thus, as important as this moment may be for a proper calibration, perhaps we can obtain more plausible estimates if we reestimate the model by dropping that moment. This exercise is carried out in column 4. Indeed, some of the parameter estimates look more reasonable now: σ_β is 2.167% and the correlation is –0.24, which is not substantially different from the benchmark. Other

estimates of the income process are also plausible and close to their benchmark values. Unfortunately, though, the parameters of the economic model now look quite suspect: $\lambda = 0.001$ and $\psi = 0.997$, both having moved from one bound to the other. Furthermore, the value of WY, which has not been imposed as a moment, is now –0.03 as compared to 1.08 in the data! A final thought is that perhaps in addition to eliminating the WY moment, we should also not fix δ and instead estimate it. The results are displayed in column 5, and the results are barely changed from the previous column.

These negative results lead to another important question. It seems that the WY moment is very important for properly estimating some parameters (such as λ and ψ, among others). But previously we spent significant efforts discussing how the auxiliary model regressions were important for pinning down these parameters. Is it possible that this emphasis was misplaced and the identification of several important parameters is coming mainly from the WY moment? It turns out the answer is no. What is happening instead is that a proper value of δ is essential for the estimation exercise and the WY moment simply ensures that δ is pinned down at a reasonable value given the other parameters of the model.[42] This can be seen as follows. In column 6, we reestimate the same model as in column 5—that is, without the WY moment, *but* we fix δ at its estimated value (0.953) when ϕ was set to 1 (in column 1), when the WY moment was used and δ was estimated. Notice that now we will not use the WY moment but only rely on the auxiliary model regressions. The estimates in column 6 of Table III are very similar to those in column 1, and all appear very reasonable. This confirms our conjecture that the WY moment's main role is to pin down the appropriate value of δ, and once that is achieved all other parameters are pinned down by the auxiliary model. (The small qualification to this statement is that ψ is 0.927 in column 5 instead of 0.88 in the benchmark case, which suggests that the WY moment perhaps also contains some information about the borrowing constraint. This would not be surprising.)

To summarize, we find that: (i) fixing either δ or ϕ and estimating the other is perfectly fine for properly estimating all the remaining parameters of the structural model, but (ii) fixing both δ and ϕ simultaneously creates severe biases. The main role of the WY moment appears to be to pin down a plausible value of δ that is consistent with the ϕ chosen, but has otherwise very little impact on the estimates of remaining parameters.

6.3.3   Fixing Borrowing Constraints

Before closing this section, we explore the effects of fixing the tightness of the borrowing constraint at some values that have been used in the literature on the estimated parameters. Column 7 of Table III displays the results when the borrowing constraint is chosen to be the natural borrowing limit, which is obtained by setting $\psi = 1$. Similarly, column 8 reports the results of the opposite exercise—of disallowing borrowing—obtained by setting $\underline{a}_t \equiv 0$. The value of risk aversion is set to 2 as in the benchmark case in Table I. Although there is some variation across the two columns and there are

---

[42]Notice that the role of δ as determining the wealth-to-income ratio is slightly different in our model compared to a standard calibration exercise. This is because here the amount of risk is not fixed (as would be the case in a calibration exercise where the income process is calibrated first and then δ is chosen). Instead the amount of risk and patience is jointly estimated.

Figure 11: Income and Consumption Inequality: Structural Model Calibrated Using Benchmark Estimates
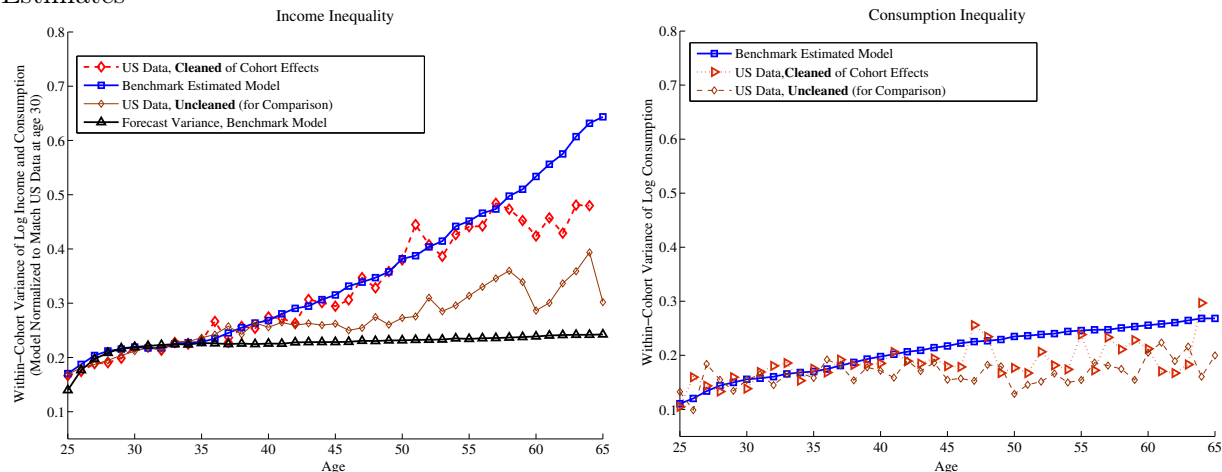
some differences from the benchmark case, by and large, these differences are quite minor. The main difference is in the fraction of households that are constrained, which is 38% when no borrowing is allowed, 14% in the benchmark estimation, and 7% with the natural borrowing limit. Thus, there is clearly sufficient information in the auxiliary model to allow us to pin down the value of the borrowing constraint, but its particular value does not seem to affect the other estimates substantially.

7   Implications for Calibrating Life Cycle Models

The model-data comparisons in the previous section are intended to evaluate the goodness of fit of the structural. We now ask how the implications of a life cycle model *calibrated* using the benchmark estimates would compare to the stylized facts on the life cycle behavior of consumption and income. For this purpose, the appropriate comparison is to empirical statistics that are computed by removing the nonstationary components—that is, cohort or time effects—from the data. Following the bulk of the literature (see, e.g., Deaton and Paxson (1994), Storesletten et al. (2004b), Guvenen (2009)), we now control for cohort effects in variances. The left panel of Figure 11 plots the profile of *cleaned* variances of income, which is now directly comparable to the existing studies mentioned above. The figure also reproduces (from Figure 8) the raw—uncleaned—profile for comparison, which shows that the profiles are very similar before age 45, but the cleaned profile is significantly *steeper* after that age, such that the variance is 15 log points higher between age 60 and 65. Similarly, the right panel plots the cleaned (and uncleaned) consumption variances in the data, which has a smaller effect—a rise of 3 log points after age 45.

Turning to the variance profiles implied by the calibrated model, somewhat surprisingly, they turn out to be more consistent with these cleaned variances, especially after the early 40s. This is especially pronounced for the income variance, which significantly overshoots the raw variances, but seems to track the cleaned variances remarkably well up to age 58. The improvement in the fit of the consumption variance profile is more subtle but still noticeable. It should be stressed, however, that

45

this is not a goodness-of-fit analysis, so these results should probably not be interpreted as a success of the estimation in fitting the data. The more accurate conclusion is that using the benchmark estimates for calibrating a life cycle model yields plausible implications for life cycle income and consumption inequality, consistent with the stylized facts from the US data.

Finally, as stated in the Introduction, a central goal of this paper is to quantify the magnitude of lifetime income risk *perceived* by households at different points in life, which we are now ready to do. The model we estimated contains two components that affect the *rise* in perceived income risk over time. The first one is the AR(1) income shocks, which is standard. The second one is the uncertainty about β^i. The left panel of Figure 11 plots (line with triangles) the forecast variance of predicted log income at different horizons, from the perspective of a 25-year-old: $\mathbb{E}_{25}\left[(y_{25+t} - \hat{y}_{25+t|25})^2\right]$ for $t = 1, 2, ..., 40$. As seen here, this profile overlaps very well with the variance of income graph in the first 10 years of the life cycle, implying that all the rise in inequality up to age 35 is considered as risk by households. This is largely because heterogeneity in β^i contributes little to income differences early on, which are instead driven by persistent income shocks. However, notice from the graph that the absolute amount of this risk is quite small. After age 35, however, inequality keeps rising—now almost entirely due to heterogeneous growth rates—whereas perceived risk tapers off, because the small initial uncertainty about β^i is largely resolved.

Overall, a 25-year-old household perceives that only about one-quarter (8 of the 31 log points) of the increase in income dispersion until age 65 is due to uncertainty, with the rest being perfectly predictable by households. In contrast, the standard RIP process estimated in the literature attributes *all* of the 31 log points rise in inequality to risk. The more general income process specification we study in this paper allows us to separate known heterogeneity from risk, which in turn reveals a much smaller amount of risk than what is currently used to calibrate models in the incomplete markets literature.

8 Conclusions

The joint dynamics of consumption and labor income contain rich information about the economic environment that individuals inhabit. In this paper, we have studied how such information can be extracted from choice data to shed light on different aspects of lifetime income risk. The Monte Carlo results suggest that the indirect inference method works very well, even in the presence of frequently binding borrowing constraints, missing observations, a concave pension function, and so on, that make the auxiliary model an imperfect approximation to the structural relationships that hold in the model. While in this paper we have focused entirely on the consumption-savings choice, the estimation method we use is general enough to accommodate a variety of other static or intertemporal decisions. Economic decisions that involve large fixed costs (and, hence, are made infrequently, such as fertility choice, house purchases, etc.) are likely to be especially forward looking and therefore useful for inferring the nature and amount of risk. We believe that the indirect inference methodology used in this paper can be fruitfully used in these alternative implementations.

Substantively, 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 the observable variables typically available to the econometrician, and finally; (iv) despite significant prior information, there is also some prior uncertainty that affects consumption behavior over the life cycle. 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. In a well-known calculation, Storesletten et al. (2004a) estimated that about half of lifetime income inequality is due to initial income differences (i.e., fixed effects), which maybe viewed as known by individuals. However, the fact that these authors assumed $\beta^i = 0$ means that that calculation attributes all the *rise* in inequality to shocks. In contrast, only a small part of the rise represents uncertainty in our framework.

We have abstracted from several important features that could be important. For example, labor supply, which is assumed to be inelastic here, could be an important tool for smoothing income shocks. Despite being methodologically feasible, such an extension raises some conceptual issues about how to model household labor supply. A sufficiently general specification that allows for an extensive margin (especially for female workers) would increase the computational demand significantly. We leave this important extension for future work. A similar issue is the treatment of partial insurance (beyond self-insurance) provided by informal channels and sources. The income data we use include several types of transfers that provide consumption smoothing relative to gross income (including gifts and help from family and friends, as well as unemployment insurance, welfare payments, etc.) There may still be further insurance available in real life and an important extension would be to model and estimate the extent of such risk, following the important work of Blundell et al. (2008). A final potentially restrictive assumption made here is that we do not distinguish between persistent and permanent shocks. This restriction was necessitated by the small sample size of the PSID and the difficulty of obtaining precise estimates once both permanent and AR(1) shocks are separately introduced (see Moffitt and Gottschalk (2008)).[43]

References

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

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

---

[43]Guvenen et al. (2010) tackle this issue using a new and confidential panel data set on labor income from the Internal Revenue Service that has more than 300,000 household-year observations, which potentially allows for the identification of a richer income process.

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

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

_ , **Luigi Pistaferri, and Ian Preston**, "Imputing consumption in the PSID using food demand estimates from the CEX," October 2006, (W04/27).

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

Bonhomme, S. and J-M. Robin, "Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990 a 2000," *Review of Economic Studies*, 2009, *76* (1), 63–92.

_ **and J.M. Robin**, "Generalized non-parametric deconvolution with an application to earnings dynamics," *Review of Economic Studies*, 2010, *77* (2), 491–533.

Carneiro, Pedro, Karsten Hansen, and James Heckman, "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College," *International Economic Review*, 2003, *24* (2), 361–421.

Cunha, Flavio, James Heckman, and Salvador Navarro, "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, April 2005, *57* (2), 191–261.

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

_ **and Christina Paxson**, "Intertemporal Choice and Inequality," *Journal of Political Economy*, June 1994, *102* (3), 437–67.

Gallant, A.R. and G. Tauchen, "Which moments to match?," *Econometric Theory*, 2006, *12* (04), 657–681.

Gourieroux, Christian, Alain Monfort, and Eric Renault, "Indirect Inference," *Journal of Applied Economics*, 1993, *8* (Supplement), S85–S118.

Gourinchas, Pierre-Olivier and Jonathan A. Parker, "Consumption over the Life Cycle," *Econometrica*, 2002, *70* (1), 47–89.

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

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

___ **and Burhanettin Kuruscu**, "A Quantitative Analysis of the Evolution of the U.S. Wage Distribution: 1970-2000," *NBER Macroeconomics Annual*, 2010.

___ **and** ___ , "Understanding the Evolution of the U.S. Wage Distribution: A Theoretical Analysis," *Journal of the European Economic Association*, forthcoming.

___ **, Bradley Heim, Vasia Panousi, Ivan Vidangos, Giovanni L. Violante, and Jason De Backer**, "Idiosyncratic Income Risk Estimated From IRS Administrative Wage Data," Technical Report, University of Minnesota 2010.

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

Hall, Robert E., "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *The Journal of Political Economy*, 1978, *86* (6), 971–987.

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

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

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante, "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 2010, *13* (1), 15–51.

Huggett, Mark, Gustavo Ventura, and Amir Yaron, "Sources of Lifetime Inequality," NBER Working Paper 13224 July 2007.

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

Krueger, Dirk and Fabrizio Perri, "How Do Households Respond to Income Shocks?," Technical Report, University of Minnesota 2009.

Lillard, Lee A. and Robert J. Willis, "Dynamic Aspetcs of Earnings Mobility," *Econometrica*, 1978, *46* (5), 985–1012.

MaCurdy, Thomas E., "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, January 1982, *18* (1), 83–114.

Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 2004, *72* (1), 1–32.

Mincer, Jacob, *Schooling, Experience and Earnings*, New York: Columbia University Press, 1974.

Moffitt, Robert and Peter Gottschalk, "Trends in the Transitory Variance of Male Earnings in the U.S., 1970-2004," Working Paper, Johns Hopkins University 2008.

Phelan, Chris, "Incentives, Insurance, and the Variability of Consumption and Leisure," *Journal of Economic Dynamics and Control*, 1994, *18* (3-4), 581–599.

Primiceri, Giorgio E. and Thijs van Rens, "Heterogeneous life-cycle profiles, income risk and consumption inequality," *Journal of Monetary Economics*, 2009, *56* (1), 20 – 39. Carnegie-Rochester Conference Series on Public Policy: The Causes and Consequences of Rising Economic Inequality April 25-26, 2008.

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

Smith, Anthony A, "Estimating Nonlinear Time–Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 1993, *8*, S63–S84.

Smith, Anthony A. Jr., "Thre Essays on the Solution and Estimation of Dynamic Macroeconomic Models." PhD dissertation, Duke University 1990.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, "Cyclical Dynamics in Idiosyncratic Labor Market Risk," *Journal of Political Economy*, June 2004, *112* (3), 695–717.

_ , **Christopher I. Telmer, and Amir Yaron**, "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, April 2004, *51* (3), 609–633.

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

NOT FOR PUBLICATION:
SUPPLEMENTAL APPENDIX

A Data Appendix

A.1 CE Data

A.1.1 1972–73 CE Data

We create a measure of nondurable consumption expenditures 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 nonzero food and income reports. In Table A.1 we report the number of households deleted from our sample during each sample selection requirement.

Table A.1: CE Sample Selection

Selection criterion:	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 1 to this variable in each subsequent year.

The income variable we use is total after-tax nonfinancial 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 to 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 to 1983 we take the family money variable and subtract the asset income of the head and the asset income of the wife. From 1975 to 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 to 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 to 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 to 1983 wife asset income is listed as one variable. From 1984 to 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 to 1992 to create the nonfinancial income variable we take family money and subtract head asset income, wife asset income, and asset income of other members of the household.

Adjusting for Taxes. From the nonfinancial income variable we need to subtract taxes paid on nonfinancial income. The PSID reports estimated total taxes for all households until 1991. 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. Finally, we subtract this estimated labor income tax from household income above to obtain the household after-tax labor income measure used in the estimation analysis.

A.2.2 Sample Selection

We start with a possible sample of 67,282 individuals interviewed between 1968 and 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 A.2.

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 $1,000. 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 nonfinancial 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.)

Table A.2: 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 & topcoding	71	23464
Male and head of household	19232	4232
Age restriction	429	3803
Five observations or more	1568	2235

A.3   Further Details of the Imputed Consumption Panel Construction

This section describes the details of the imputation procedures and reports some further validation tests on the quality of imputation.

As explained in the paper, we modify and extend BPP's procedure to make it suitable for our econometric analysis, which uses data from 1968 to 1992. This involves combining the 1980–92 CE survey with the 1972–73 waves, replacing the time dummies in BPP with different inflation variables, and using age dummies instead of a polynomial in age.

Specifically, we estimate a food demand system as a function of nondurable expenditures, demographics, and relative prices using an instrumental variables approach. To deal with endogeneity and measurement error, 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). The cohort-education-year specific averages of the log of husband's and wife's hourly wage rates are generated 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.

Table A.3 reports the results from the IV 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 nondurable consumption expenditures.

BPP used the evolution of the variance of consumption over time to check the quality of their imputation procedure. For completeness, here we discuss the results of our imputation for the same statistic. Figure A.1 plots the cross-sectional variance of log consumption over time. 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 kernel-smoothed version. 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 the 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.

B   A Monte Carlo Study

To investigate the ability of the proposed estimation method to uncover the true structural parameter vector with the specified auxiliary model, we begin by conducting a Monte Carlo study. The results are contained

Table A.3: 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 A.1: Cross-sectional Variance of Log Consumption in CEX and Imputed PSID Data: 1968–1992
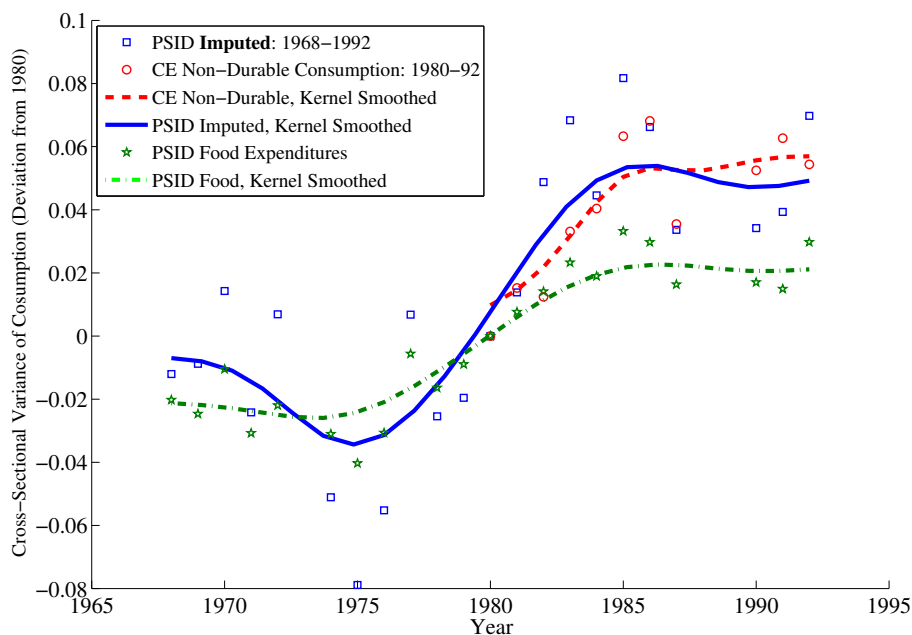
in Table A.4.[44] In column 1, the "true values" for the parameters are set to our benchmark estimates from PSID income and consumption data (column 1 of Table I). For each parameter, the initial values are drawn randomly from a uniform distribution centered around the true value but with a wide support.[45] The results discussed here are based on 100 replications, where each Monte Carlo run takes about 20 to 24 hours on a state-of-the-art workstation. Column 2 reports the results when both income and consumption data are used jointly for estimation. Clearly, the estimation method works well: bias is virtually absent for most parameters and is very small for the remaining few. Standard deviations are very small, indicating that all the parameters, with the exception of $corr_{\alpha\beta}$, can be identified in this framework. A useful question to ask is whether there are benefits to using consumption data in the estimation for (the six) parameters that can be identified with income data alone. To investigate this we use the same "true" values as in the previous exercise, but estimate the income process with income data alone and with equation (23) as the only auxiliary model regression (reported columns 4 and 5). Perhaps, unsurprisingly, the mean estimates have little if any bias, and although the precision of the estimates fall, this is minor for all parameters, except σ_β. However, as we shall see in the next section, when we estimate the income process from "real" (PSID) data, they turn out to be different from those in column of this table (i.e., A.4). Thus, another exercise we conduct is to take as the "true" values as those obtained from PSID with income data alone (column 2 of Table I). As seen in columns 7 and 8, the estimates are still largely unbiased, but the precision now has fallen significantly for some parameter values, most importantly for σ_β—going from 0.17 to 0.27—and for $corr_{\alpha\beta}$—going from 0.15 to 0.29. This reduced precision makes it harder to separate whether the rise in income inequality is coming from σ_β or from σ_α through the strong correlation.

Although it is difficult, if not impossible, to prove identification in this general setup, overall these 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 widely missing data, binding borrowing constraints, and

---

[44]We set $\overline{u}^{i,c} \equiv 0$ in the Monte Carlo analysis, because all households in the simulated data have the same demographics and zero initial wealth, making this fixed effect redundant.

[45]The exact ranges for the starting values are as follows: $\sigma_\alpha \in [0.2, 0.4]$, $\sigma_\beta \in [1.0, 3.0]$, $corr_{\alpha\beta} \in [-0.60, 0.35]$, $\rho \in [0.55, 0.95]$, $\sigma_\eta \in [0.15, 0.25]$, $\sigma_\varepsilon \in [0.001, 0.20]$, $\lambda \in [0.001, 0.75]$, $\sigma_y \in [0.11, 0.9]$, $\sigma_c \in [0.32, 0.39]$, $\sigma_{c_0} \in [0.38, 0.48]$, $\delta \in [0.92, 0.99]$, and $\psi \in [0.3, 0.99]$.

Table A.4: MONTE CARLO ANALYSIS

	Using **Y**+**C** Data			Using **Y** Data		Using **Y** Data		
	"True"	Estimates		Estimates		"True"	Estimates	
	Value 1	Mean	Sd. Err.	Mean	Std. Err.	Value 2	Mean	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income Processes Parameters								
σ_α	0.284	0.279	0.025	0.285	0.031	0.298	0.301	0.038
σ_β	1.852	1.815	0.176	1.834	0.220	1.343	1.377	0.271
$corr_{\alpha\beta}$	−0.162	−0.146	0.148	−0.161	0.173	0.558	0.531	0.289
ρ	0.754	0.758	0.025	0.754	0.027	0.783	0.780	0.022
σ_η	0.196	0.196	0.005	0.196	0.005	0.199	0.199	0.005
σ_ε	0.004	0.030	0.023	—	—	—		
Economic Model Parameters								
λ	0.345	0.318	0.084					
δ	0.950	0.950	0.002					
ψ	0.874	0.869	0.096					
Measurement Errors								
σ_y	0.147	0.142	0.007	0.147	0.005	0.147^a	0.146	0.005
σ_c	0.356	0.356	0.002	—		—		
σ_{c_0}	0.428	0.422	0.009	—		—		

[a]In estimations with income data alone, transitory shocks and measurement error cannot be identified separately. So we assume all i.i.d. shocks are measurement error with a standard deviation equal to $\sqrt{\sigma_y^2 + \sigma_\varepsilon^2}$ using the estimated value from the benchmark.

multiple sources of heterogeneity and randomness.

Finally, before settling down on the auxiliary model used in this paper (equations (22) and (23)), we explored a large number of alternatives. In Table A.5, we report a subset of our results from that work, which is representative of the issues we generally encountered. Columns 2 and 3 report the results when we use the same auxiliary model regressions as in the baseline case, but run them separately for *three* age groups (instead of two) yielding 75 parameters (instead of 50). In Columns 4 to 9, we go in the other direction and examine a sequence of auxiliary models that are successively more parsimonious. First (columns 4 and 5), we use the baseline auxiliary model but put no weight on the WY moment. Second (columns 6 and 7) we use the baseline auxiliary model but drop regressors that have t-statistics less than two. Finally, in columns 8 and 9, we use the same auxiliary model as in 6 and 7, but use only one set of equations for individuals of all age groups. The overall conclusion from these experiments is that the baseline auxiliary performs better in terms of both bias and precision of the estimates than the four alternatives that we explored. The main differences revolve around three parameters: λ, σ_β, and $corr_{\alpha\beta}$. For these parameters, we find that the alternative auxiliary models tend to generate estimates that exhibit both more bias and less precision. The differences are small in some cases, but quite large in the last experiment (with a single age group and dropping insignificant regressors). We have also experimented with other auxiliary models that seemed a prior plausible and differed more dramatically from those reported here and found them to perform very poorly. Therefore, we do not discuss them here.

Table A.5: Monte Carlo Analysis: Alternative Auxiliary Models

Auxiliary Model:		Baseline		No WY moment		Drop regressors with t-stats < 2.0		Drop regressors with t-stats < 2.0	
Number of Age Groups:[a]	"True"	Three		Two		Two		One	
	Value 1	Estimates		Estimates		Estimates		Estimates	
		Mean	Sd. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Income Processes Parameters									
σ_α	0.284	0.292	0.023	0.284	0.025	0.284	0.024	0.284	0.024
σ_β	1.852	2.000	0.163	1.814	0.193	1.821	0.207	1.758	0.397
$corr_{\alpha\beta}$	−0.162	−0.211	0.143	−0.162	0.189	−0.170	0.205	−0.09	0.228
ρ	0.754	0.765	0.027	0.756	0.025	0.760	0.029	0.757	0.038
σ_η	0.196	0.201	0.005	0.196	0.005	0.194	0.005	0.195	0.005
σ_ε	0.004	0.041	0.031	0.026	0.025	0.036	0.023	0.039	0.022
λ	0.345	0.320	0.110	0.291	0.110	0.310	0.094	0.272	0.114
δ	0.950	0.950	0.002	0.949	0.003	0.951	0.002	0.951	0.002
ψ	0.874	0.949	0.041	0.886	0.082	0.882	0.083	0.797	0.148
σ_y	0.147	0.146	0.010	0.142	0.008	0.142	0.007	0.141	0.007
σ_c	0.356	0.371	0.002	0.356	0.003	0.356	0.002	0.356	0.002
σ_{c_0}	0.428	0.439	0.010	0.421	0.010	0.420	0.010	0.418	0.010

[a]The baseline estimation uses two age groups in the auxiliary model.

C   Derivations and Proofs

C.1   Proofs of Propositions 1 and 2

We begin by establishing the following lemma that comes in handy in the proofs that follow.

Lemma A.1 *Assume $\sigma_\beta > 0$. Then $\partial\Pi_t/\partial\lambda > 0$.*

Proof.   First recall that $\Pi_t = \Phi_t(A_t/X_t) + \Psi_t(B_t/X_t)$. The dependence of Π_t on λ comes through A_t, B_t, and X_t, which all contain elements of $\mathbf{P}_{t+1|t}$, which in turn depend on λ. Thus, to establish $\partial\Pi_t/\partial\lambda > 0$ we first iterate on the recursions for updating the posterior covariance matrix (equation (3)). Specifically, start at $t = 0$: $\sigma^2_{\beta,1|0} = \lambda^2\sigma^2_\beta$, $\sigma^2_{z,1|0} = \sigma^2_\eta$ and $\sigma_{\beta z,1|0} = 0$. Then we can find $A_1 = \lambda^2\sigma^2_\beta$, $B_1 = \sigma^2_\eta$, and $X_1 = \lambda^2\sigma^2_\beta + \sigma^2_\eta$. Plugging these expressions into (7) and (8), we obtain $\sigma^2_{\beta,2|1} = \frac{\lambda^2\sigma^2_\beta\sigma^2_\eta}{\lambda^2\sigma^2_\beta+\sigma^2_\eta}$, $\sigma^2_{z,2|1} = \rho^2\frac{\lambda^2\sigma^2_\beta\sigma^2_\eta}{\lambda^2\sigma^2_\beta+\sigma^2_\eta}+\sigma^2_\eta$ and $\sigma^2_{\beta z,2|1} = \rho\frac{\lambda^2\sigma^2_\beta\sigma^2_\eta}{\lambda^2\sigma^2_\beta+\sigma^2_\eta}$. After a few iterations like this, the following recursive formulas emerge for the variances and covariances:

$$\sigma^2_{\beta,t+1|t} = K_t \tag{26}$$

$$\sigma^2_{z,t+1|t} = t^2\rho^2 K_t + \sigma^2_\eta \tag{27}$$

$$\sigma_{\beta z,t+1|t} = -t\rho K_t, \tag{28}$$

where $K_t \equiv \dfrac{\Theta\sigma^2_\eta}{\Theta(\overset{t}{\underset{s=0}{\Sigma}}(s-(s-1)\rho)^2-1)+\sigma^2_\eta}$ and $\Theta \equiv \dfrac{\lambda^2\sigma^2_\beta\sigma^2_\eta}{\lambda^2\sigma^2_\beta+\sigma^2_\eta}$.

Now, first, it is straightforward to show that $\Pi_t(\lambda = 0) > 0$. To see this, observe that if $\lambda = 0$ then $\Theta = 0$ and $K_t = 0$ for all t. So, $\sigma^2_{\beta,t|t-1} = \sigma_{\beta z,t|t-1} = 0$ and $\sigma^2_{z,t|t-1} = \sigma^2_\eta$. It follows that $A_t = 0$, $B_t = \sigma^2_\eta$ and $X_t = \sigma^2_\eta$. Plugging these values into the expression for Π_t shows that $\Pi_t = \Psi(t; T, r, \rho) \equiv \frac{1-\gamma}{1-\gamma\rho}\left[\frac{(1-(\gamma\rho)^{T-t+1})}{(1-\gamma^{T-t+1})}\right]$, which is always positive.

Second, to show that $\frac{\partial\Pi_t}{\partial\lambda} > 0$, we need to calculate the derivatives of A_t/X_t and B_t/X_t with respect to λ. First, we have

$$\frac{\partial\Theta}{\partial\lambda} = \frac{2\lambda^3\sigma^2_\beta(\sigma^2_z)^2}{(\lambda^2\sigma^2_\beta+\sigma^2_z)^2} > 0,$$

$$\frac{\partial K_t}{\partial\lambda} = \frac{\frac{\partial\Theta}{\partial\lambda}(\sigma^2_\eta)^2}{(\Theta(\overset{t}{\underset{s=0}{\Sigma}}(s-(s-1)\rho)^2-\rho^2)+\sigma^2_\eta)^2} > 0.$$

Using the chain rule and $\frac{\partial K_t}{\partial\lambda} > 0$ for all t, we find

$$\frac{\partial(A_t/X_t)}{\partial\lambda} = (t-\rho(t-1))K_{t-1}\frac{\partial K_{t-1}}{\partial\lambda}\sigma^2_\eta > 0$$

$$\frac{\partial(B_t/X_t)}{\partial\lambda} = -t(t-\rho(t-1))K_{t-1}\frac{\partial K_{t-1}}{\partial\lambda}\sigma^2_\eta < 0.$$

We can rewrite the derivative of Π_t with respect to λ as

$$\frac{\partial\Pi_t}{\partial\lambda} = [\Phi_t\frac{\partial(A_t/X_t)}{\partial\lambda} + \Psi_t\frac{\partial(B_t/X_t)}{\partial\lambda}] \tag{29}$$

$$= [\Phi_t - t\Psi_t](t-\rho(t-1))K_{t-1}\frac{\partial K_{t-1}}{\partial\lambda}\sigma^2_\eta. \tag{30}$$

Note that all terms outside of the square brackets are positive. Thus, $\frac{\partial\Pi_t}{\partial\lambda} > 0$ if and only if $[\Phi_t - t\Psi_t] > 0$.

To prove the latter, we proceed in two steps. First, the expression we are interested in is

$$\Phi_t - t\Psi_t = \left[\frac{\gamma}{(1-\gamma)} + \frac{t - (T+1)\gamma^{T-t+1}}{(1-\gamma^{T-t+1})} \right] - t \left[\frac{1-\gamma}{1-\gamma^{T-t+1}} \frac{1-(\gamma\rho)^{T-t+1}}{1-\gamma\rho} \right]. \tag{31}$$

It is straightforward to see that $\partial(\Phi_t - t\Psi_t)/\partial\rho < 0$, since ρ only appears in the second set of brackets (i.e, Ψ_t), which clearly becomes more negative as ρ rises.[46] Therefore, it is sufficient to prove that $\Phi_t - t\Psi_t > 0$ when $\rho = 1$ and the same will hold for all values of $\rho < 1$. This is how we shall proceed. Let the remaining planning horizon of an individual be denoted with $\tau \equiv T - t + 1$. When $\rho = 1$ we have $\Psi_t = 1$ and the expression simplifies to

$$\Phi_t - t\Psi_t = \left[\frac{\gamma}{(1-\gamma)} + \frac{t - (T+1)\gamma^{T-t+1}}{(1-\gamma^{T-t+1})} \right] - t = \frac{\gamma}{(1-\gamma)} - \frac{\tau\gamma^\tau}{(1-\gamma^\tau)}. \tag{32}$$

For $\tau = 1$, the expression equals zero. All we need to show is that the derivative of this expression with respect to τ is negative, which will then establish that $\Phi_t - t\Psi_t < 0$ for all $\tau > 1$. This is easy to do:

$$\frac{d}{d\tau} \left(\frac{\tau\gamma^\tau}{(1-\gamma^\tau)} \right) = \frac{\gamma^\tau(1 + \tau\log\gamma - \gamma^\tau)}{(1-\gamma^\tau)^2} < 0 \Leftrightarrow 1 + \tau\log\gamma - \gamma^\tau < 0 \Leftrightarrow \gamma^\tau < e^{\gamma^\tau - 1},$$

which is satisfied for all $\tau > 1$ as long as $\gamma < 1$ (i.e., $r > 0$). Since the second term in (32) is decreasing with the horizon, this establishes that $\Phi - t\Psi > 0$ for all $\tau > 1$ (alternatively $t < T$). $\blacksquare$

Proof of Proposition 1:

Part (i): Rewrite (14) as

$$\Delta C_t^i = \Pi_t \times (Y_t^i - \mathbb{E}_{t-1}(Y_t^i)) = \Pi_t \times (\Delta Y_t^i + Y_{t-1}^i - \mathbb{E}_{t-1}(Y_t^i))$$

$$= \Pi_t \times \Delta Y_t^i + \Pi_t \times \left(\beta^i(t-1) + z_{t-1}^i - (\widehat{\beta}_{t-1}^i t + \widehat{z}_{t-1}^i) \right)$$

$$= \Pi_t \times \Delta Y_t^i + \Pi_t \times \left(\left(\beta^i - \widehat{\beta}_{t-1}^i\right) t + (z_{t-1}^i - \widehat{z}_{t-1}^i) - \beta^i \right)$$

Taking the expectations of both sides with respect to the history up to time $t - 1$ (of prior beliefs and income realizations, $\{Y_1^i, Y_2^i, .., Y_{t-1}^i, \widehat{\beta}_{1|0}^i\}$) conditional on $\beta^i, \Delta Y_t^i$:

$$\mathbb{E}\left(\Delta C_t^i|\beta^i, \Delta Y_t^i\right) = \Pi_t \times \left(\Delta Y_t^i + \mathbb{E}\left(\beta^i - \widehat{\beta}_{t-1}^i|\beta^i, \Delta Y_t^i\right) \times t \right)$$
$$+ \Pi_t \times \mathbb{E}\left(z_{t-1}^i - \widehat{z}_{t-1}^i|\beta^i, \Delta Y_t^i\right) - \Pi_t \times \beta^i$$
$$\mathbb{E}\left(\Delta C_t^i|\beta^i, \Delta Y_t^i\right) = \Pi_t \times \Delta Y_t^i + \Pi_t \times (\beta^i - \beta^i) \times t$$
$$+ \Pi_t \times (0) - \Pi_t \times \beta^i.$$

On the last line, we made use of two facts: $\mathbb{E}\left(z_{t-1}^i - \widehat{z}_{t-1}^i|\beta^i, \Delta Y_t^i\right) = 0$ and $\mathbb{E}\left(\widehat{\beta}_{t-1}^i|\beta^i, \Delta Y_t^i\right) = \beta^i$, which yields

$$\mathbb{E}\left(\Delta C_t^i|\beta^i, \Delta Y_t^i\right) = \Pi_t \times \Delta Y_t^i - \Pi_t \times \beta^i. \tag{33}$$

Therefore, controlling for income growth, on average, consumption growth is decreasing in β^i. Furthermore, since $\partial\mathbb{E}(Y_{t-1}^i - Y_1^i|\beta^i)/\partial\beta^i > 0$, consumption growth is also decreasing in past income growth.

Part (ii): We need to establish three results. First the negative dependence proved above holds even when $\lambda = 0$, that is, when the individual has full information about his/her β^i. Second, the strength of consumption's response to past income growth becomes stronger (i.e, becomes more negative) as λ rises. From the expression for $\Delta\overline{C}_t^i$ given in (33), this is equivalent to showing $\partial\Pi_t/\partial\lambda > 0$, which was proved in Lemma 1. $\blacksquare$

Proof of Lemma 1. From the solution of the model, we know that consumption equals the annuity value of the physical wealth and expected lifetime discounted labor income:

---

[46]To see this, note that the ratio $\frac{1-(\gamma\rho)^{T-t+1}}{1-\gamma\rho}$ can be expanded as $1 + (\gamma\rho) + (\gamma\rho)^2 + ... + (\gamma\rho)^{T-t}$, which is clearly increasing in ρ.

$$C_t = \varphi_t[\frac{1}{\gamma}A_t + \sum_{s=0}^{T-t}\gamma^s\mathbb{E}_t(Y_{t+s})]. \tag{34}$$

Taking the expectation of the income process, $Y_t^i = \alpha^i + \beta^i t + z_t^i$, we find: $E_t(Y_{t+s}^i) = \alpha^i + \hat{\beta}_t^i(t+s) + \rho^s \hat{z}_t^i$. Plugging (9) in (34) yields:

$$C_t^i = \varphi_t[\frac{1}{\gamma}A_t^i + \sum_{s=0}^{T-t}\gamma^s(\alpha^i + \hat{\beta}_t^i(t+s) + \rho^s \hat{z}_t^i)] \tag{35}$$

$$= \varphi_t[\frac{1}{\gamma}A_t^i + \underbrace{(\alpha^i + \hat{\beta}_t^i t + \hat{z}_t^i)}_{Y_t^i} + \sum_{s=1}^{T-t}\gamma^s(\alpha^i + \hat{\beta}_t^i(t+s) + \rho^s \hat{z}_t^i)] \tag{36}$$

$$\Rightarrow C_t^i = \varphi_t(\underbrace{\omega_t^i}_{\text{Cash on hand}}) + \gamma\Phi(t+1;T,r)\hat{\beta}_t^i + \gamma\rho\Psi(t+1;T,r,\rho)\hat{z}_t^i, \tag{37}$$

which is equation (16) in Lemma 1.

C.2 Likelihood Approach vs Quadratic Objective: An Equivalence

Here we establish the asymptotic equivalence between the "likelihood approach" to indirect inference employed in our estimation and the quadratic objective—also called the Wald approach—that is often used in the literature. We prove the equivalence for a stylized case for clarity, although it will become clear that the proof 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. Now, consider the structural (i.e., "true") model:

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

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

$$\hat{h}_i(\beta) \equiv (\hat{\gamma}_{0,i}(\beta), \hat{\gamma}_{1,i}(\beta)) = \underset{\gamma_0, \gamma_1}{\text{argmin}} \sum_{t=1}^T (Y_t^i(\beta) - \gamma_0 - \gamma_1 Y_{t-1}^i(\beta))^2,$$

where i denotes the ith simulated data set, given β. Now define

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

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

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

The approach in this paper is (assuming M is large) to calculate

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

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

$$0 = \sum_t \left(Y_t - \gamma_0 \left(\beta \right) - \gamma_1 \left(\beta \right) Y_{t-1} \right) \gamma_0' \left(\beta \right)$$

$$+ \sum_t \left(Y_t - \gamma_0 \left(\beta \right) - \gamma_1 \left(\beta \right) Y_{t-1} \right) \gamma_1' \left(\beta \right) Y_{t-1}$$

$$= -\gamma_0' \left(\beta \right) \sum Y_t + \gamma_0 \left(\beta \right) \gamma_0' \left(\beta \right) T + \gamma_1 \left(\beta \right) \gamma_0' \left(\beta \right) \sum Y_{t-1}$$

$$- \gamma_1' \left(\beta \right) \sum Y_t Y_{t-1} + \gamma_0 \left(\beta \right) \gamma_1' \left(\beta \right) \sum Y_{t-1} + \gamma_1 \left(\beta \right) \gamma_1' \left(\beta \right) \sum Y_{t-1}^2, \tag{38}$$

where $\gamma_j'(\beta)$ is the derivative of γ_j, $j = 0, 1$. Now, as an alternative, consider minimizing the following quadratic form:

$$\left[\begin{array}{c} \gamma_0 \left(\beta \right) - \hat{\gamma}_0 \\ \gamma_1 \left(\beta \right) - \hat{\gamma}_1 \end{array} \right]' \left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right] \left[\begin{array}{c} \gamma_0 \left(\beta \right) - \hat{\gamma}_0 \\ \gamma_1 \left(\beta \right) - \hat{\gamma}_1 \end{array} \right],$$

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

$$a_{11} \left(\gamma_0 \left(\beta \right) - \hat{\gamma}_0 \right) \gamma_0' \left(\beta \right) + a_{12} \left(\gamma_0 \left(\beta \right) - \hat{\gamma}_0 \right) \gamma_1' \left(\beta \right) + a_{12} \left(\gamma_1 \left(\beta \right) - \hat{\gamma}_1 \right) \gamma_0' \left(\beta \right) + a_{22} \left(\gamma_1 \left(\beta \right) - \hat{\gamma}_1 \right) \gamma_1' \left(\beta \right) =$$

$$= \left(-a_{11} \hat{\gamma}_0 - a_{12} \hat{\gamma}_1 \right) \gamma_0' \left(\beta \right) - \left(a_{12} \hat{\gamma}_0 + a_{22} \hat{\gamma}_1 \right) \gamma_1' \left(\beta \right) + a_{11} \gamma_0 \left(\beta \right) \gamma_0' \left(\beta \right) + a_{12} \gamma_0 \left(\beta \right) \gamma_1' \left(\beta \right)$$

$$+ a_{12} \gamma_0' \left(\beta \right) \gamma_1 \left(\beta \right) + a_{22} \gamma_1 \left(\beta \right) \gamma_1' \left(\beta \right) = 0. \tag{39}$$

We want to make (39) look like condition (38). To do so, first set

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

Then the last four terms in (39) match four of the six terms in (38). But what about the remaining two terms in each equation? One can show that these terms match up asymptotically, as the observed sample size T grows large. To see this:

$$\operatorname{plim} \left[\gamma_0' \left(\hat{\beta}_T \right) \left(\frac{\sum Y_t}{T} - \frac{T \hat{\gamma}_0 \left(\hat{\beta}_T \right)}{T} - \left(\frac{\sum Y_{t-1}}{T} \right) \hat{\gamma}_1 \right) \right] = \gamma_0' \left(\beta_0 \right) \underbrace{\left(\mathbb{E} Y_t - \gamma_0 \left(\beta_0 \right) - \left(\mathbb{E} Y_{t-1} \right) \gamma_1 \left(\beta_0 \right) \right)}_{=0}$$

$$= \gamma_0' \left(\beta_0 \right) \times 0 = 0,$$

where β_0 is the "true" value of β (i.e., $\operatorname{plim} \hat{\beta}_T = \beta_0$). The terms in the set of parentheses on the right hand side is zero asymptotically as $T \to \infty$ because it is simply the F.O.C. that defines $\gamma_0 \left(\beta_0 \right) \left(= \operatorname{plim} \hat{\gamma}_0 \right)$. Similarly,

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

$$= \gamma_1' \left(\hat{\beta}_0 \right) \left(\mathbb{E} Y_t Y_{t-1} - \left(\mathbb{E} Y_{t-1} \right) \gamma_0 \left(\beta_0 \right) - \left(\mathbb{E} Y_{t-1}^2 \right) \gamma_1 \left(\beta_0 \right) \right) = 0,$$

again, because the second term is (asymptotically) the f.o.c. that defines $\gamma_1 \left(\beta_0 \right) = \operatorname{plim} \hat{\gamma}_1$. This shows that equations (38) and (39) are asymptotically equivalent, completing the proof.

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:

$$\left[\begin{array}{c} \gamma_0 \left(\beta \right) - \hat{\gamma}_0 \\ \gamma_1 \left(\beta \right) - \hat{\gamma}_1 \end{array} \right]' \left[\begin{array}{cc} 1 & \mathbb{E} Y_{t-1} \\ \mathbb{E} Y_{t-1} & \mathbb{E} Y_{t-1}^2 \end{array} \right] \left[\begin{array}{c} \gamma_0 \left(\beta \right) - \hat{\gamma}_0 \\ \gamma_1 \left(\beta \right) - \hat{\gamma}_1 \end{array} \right].$$

Note that the weighting matrix would be the optimal one if the auxiliary model were correctly specified, because it is proportional to the inverse of the asymptotic covariance matrix of $T^{1/2} \left[\begin{array}{cc} \gamma_0(\beta_0) - \hat{\gamma}_0 & \gamma_1(\beta_0) - \hat{\gamma}_1 \end{array} \right]'$, where β_0 is the true value of β.