Econometrica Supplementary Material

SUPPLEMENT TO "INFERRING LABOR INCOME RISK AND PARTIAL INSURANCE FROM ECONOMIC CHOICES": APPENDIXES

(*Econometrica*, Vol. 82, No. 6, November 2014, 2085–2129)

BY FATIH GUVENEN AND ANTHONY A. SMITH, JR.

APPENDIX A: DERIVATIONS AND PROOFS

A.1. *Kalman Filtering Equations*

FOR THE PROBLEM AT HAND, (3) and (4) can be manipulated to obtain some simple expressions. First, (3) simplifies to

$$\hat{\beta}^i_t - \hat{\beta}^i_{t-1} = (A_t/X_t)\hat{\xi}^i_t, \tag{28}$$

$$\hat{z}^i_t - \rho\hat{z}^i_{t-1} = (B_t/X_t)\hat{\xi}^i_t, \tag{29}$$

where $A_t \equiv t\sigma^2_{\beta,t|t-1} + \sigma_{\beta z,t|t-1}$, $B_t \equiv t\sigma_{\beta z,t|t-1} + \sigma^2_{z,t|t-1}$, and $X_t \equiv \mathrm{var}_{t-1}(y^i_t) = A_t t + B_t$. 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 $\hat{\xi}^i_t$ is split according to the perceived share of variance attributed to each component. Second, (4) reduces to

$$\sigma^2_{\beta,t+1|t} = \sigma^2_{\beta,t|t-1} - \frac{A^2_t}{X_t}, \tag{30}$$

$$\sigma^2_{z,t+1|t} = \rho^2\left[\sigma^2_{z,t|t-1} - \frac{B^2_t}{X_t}\right] + \sigma^2_\eta. \tag{31}$$

As shown by Guvenen (2007), an important feature of Bayesian learning in this framework is that beliefs about β^i change nonmonotonically 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. 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 ($\hat{\sigma}^2_{\beta|0}$).

A.2. *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 $\partial\Pi_t/\partial\lambda > 0$.*

 DOI: 10.3982/ECTA9446

PROOF: First recall that $\Pi_t = \Phi_t(A_t/X_t) + \Psi_t(B_t/X_t)$, where it can be shown (through straightforward but tedious algebra) that

$$\Phi(t; T, r) \equiv \left[\frac{\gamma}{(1-\gamma)} + \frac{t-(T+1)\gamma^{T-t+1}}{(1-\gamma^{T-t+1})}\right]$$

and

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

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 (4)). 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 (30) and (31), 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:

$$\sigma^2_{\beta,t+1|t} = K_t, \tag{32}$$

$$\sigma^2_{z,t+1|t} = t^2\rho^2K_t + \sigma^2_\eta, \tag{33}$$

$$\sigma_{\beta z,t+1|t} = -t\rho K_t, \tag{34}$$

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

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 in these values shows that $\Pi_t = \Psi(t; T, r, \rho)$, 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

$$\frac{\partial \Theta}{\partial \lambda} = \frac{2\lambda^3 \sigma_\beta^2 (\sigma_z^2)^2}{(\lambda^2 \sigma_\beta^2 + \sigma_z^2)^2} > 0,$$

$$\frac{\partial K_t}{\partial \lambda} = \frac{\frac{\partial \Theta}{\partial \lambda} (\sigma_\eta^2)^2}{\left(\Theta \left(\sum_{s=0}^{t} (s - (s-1)\rho)^2 - \rho^2 \right) + \sigma_\eta^2 \right)^2} > 0.$$

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

$$\frac{\partial (A_t/X_t)}{\partial \lambda} = \big(t - \rho(t-1)\big) K_{t-1} \frac{\partial K_{t-1}}{\partial \lambda} \sigma_\eta^2 > 0,$$
$$\frac{\partial (B_t/X_t)}{\partial \lambda} = -t\big(t - \rho(t-1)\big) K_{t-1} \frac{\partial K_{t-1}}{\partial \lambda} \sigma_\eta^2 < 0.$$

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

$$\frac{\partial \Pi_t}{\partial \lambda} = \left[\Phi_t \frac{\partial (A_t/X_t)}{\partial \lambda} + \Psi_t \frac{\partial (B_t/X_t)}{\partial \lambda} \right] \tag{35}$$

$$= [\Phi_t - t\Psi_t]\big(t - \rho(t-1)\big) K_{t-1} \frac{\partial K_{t-1}}{\partial \lambda} \sigma_\eta^2. \tag{36}$$

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

$$\Phi_t - t\Psi_t = \left[\frac{\gamma}{(1-\gamma)} + \frac{t - (T+1)\gamma^{T-t+1}}{(1-\gamma^{T-t+1})} \right] - t\left[\frac{1-\gamma}{1-\gamma^{T-t+1}} \frac{1-(\gamma\rho)^{T-t+1}}{1-\gamma\rho} \right]. \tag{37}$$

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.[33] 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

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

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

$$\Phi_t - t\Psi_t = \left[\frac{\gamma}{(1-\gamma)} + \frac{t-(T+1)\gamma^{T-t+1}}{(1-\gamma^{T-t+1})}\right] - t \tag{38}$$

$$= \frac{\gamma}{(1-\gamma)} - \frac{\tau\gamma^\tau}{(1-\gamma^\tau)}.$$

For $\tau = 1$, the expression equals zero. All we need to show is that the derivative[34] of the second term 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,

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

$$\Leftrightarrow \quad 1 + \tau\log\gamma - \gamma^\tau < 0 \quad \Leftrightarrow \quad \gamma^\tau < e^{\gamma^\tau - 1},$$

which is satisfied for all $\tau > 1$ as long as $\gamma < 1$ (i.e., $r > 0$). Since the second term in (38) is decreasing with the horizon, this establishes that $\Phi - t\Psi > 0$ for all $\tau > 1$ (alternatively $t < T$). *Q.E.D.*

PROOF OF PROPOSITION 1: (i) Rewrite (13) as

$$\begin{aligned}
\Delta C_t^i &= \Pi_t \times \left(Y_t^i - \mathbb{E}_{t-1}(Y_t^i)\right) = \Pi_t \times \left(\Delta Y_t^i + Y_{t-1}^i - \mathbb{E}_{t-1}(Y_t^i)\right) \\
&= \Pi_t \times \Delta Y_t^i + \Pi_t \times \left(\beta^i(t-1) + z_{t-1}^i - \left(\hat{\beta}_{t-1}^i t + \hat{z}_{t-1}^i\right)\right) \\
&= \Pi_t \times \Delta Y_t^i + \Pi_t \times \left(\left(\beta^i - \hat{\beta}_{t-1}^i\right)t + \left(z_{t-1}^i - \hat{z}_{t-1}^i\right) - \beta^i\right).
\end{aligned}$$

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

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

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

$$\mathbb{E}\left(\Delta C_t^i | \beta^i, \Delta Y_t^i\right) = \Pi_t \times \Delta Y_t^i - \Pi_t \times \beta^i. \tag{39}$$

[34]Here we are treating τ as a continuous variable, when in fact time is discrete. It is easy to see that this is an innocuous assumption in this context.

Therefore, controlling for income growth, on average, consumption growth is decreasing in β^i. Furthermore, since $\partial\mathbb{E}(Y^i_{t-1} - Y^i_1|\beta^i)/\partial\beta^i > 0$, consumption growth is also decreasing in past income growth.

(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 $\Delta\overline{C}^i_t$ given in (39), this is equivalent to showing $\partial\Pi_t/\partial\lambda > 0$, which is now proved in Lemma 1. *Q.E.D.*

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:

$$C_t = \varphi_t\left[\frac{1}{\gamma}A_t + \sum_{s=0}^{T-t}\gamma^s\mathbb{E}_t(Y_{t+s})\right]. \tag{40}$$

Taking the expectation of the income process, $Y^i_t = \alpha^i + \beta^i t + z^i_t$, we find $E_t(Y^i_{t+s}) = \alpha^i + \hat{\beta}^i_t(t+s) + \rho^s\hat{z}^i_t$. Plugging (6) into (40) yields

$$C^i_t = \varphi_t\left[\frac{1}{\gamma}A^i_t + \sum_{s=0}^{T-t}\gamma^s\big(\alpha^i + \hat{\beta}^i_t(t+s) + \rho^s\hat{z}^i_t\big)\right] \tag{41}$$

$$= \varphi_t\left[\frac{1}{\gamma}A^i_t + \underbrace{\big(\alpha^i + \hat{\beta}^i_t t + \hat{z}^i_t\big)}_{Y^i_t} + \sum_{s=1}^{T-t}\gamma^s\big(\alpha^i + \hat{\beta}^i_t(t+s) + \rho^s\hat{z}^i_t\big)\right] \tag{42}$$

$$\Rightarrow \quad C^i_t = \varphi_t\big(\underset{\text{Cash on hand}}{\omega^i_t}\big) + \gamma\Phi(t+1; T, r)\hat{\beta}^i_t + \gamma\rho\Psi(t+1; T, r, \rho)\hat{z}^i_t, \tag{43}$$

which is equation (15) in Lemma 1. *Q.E.D.*

A.3. *Partial Insurance*

Following the same steps as in the proof of Lemma 1 above and replacing Y_t with Y^{disp}_t yields the following expression for consumption growth in the presence of partial insurance:

$$\Delta C_t = \hat{\xi}_t\left\{\frac{A_t}{X_t}\big[\Phi(t; T, r) - t\varphi_t\theta\big] + \frac{B_t}{X_t}\big[\Psi(t; T, r) - \varphi_t\theta\big]\right\}.$$

This expression can be further simplified by rearranging terms and recognizing that $A_t t + B_t \equiv X_t$. We get

$$\Delta C_t = (\Pi_t - \varphi_t \theta) \times \hat{\xi}_t.$$

A.4. *Likelihood Approach versus Quadratic Objective: An Equivalence*

Here we establish the asymptotic equivalence between the "likelihood ratio (LR) 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 $\epsilon_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$, σ^2 is known, and Y_0 is given. Consider the auxiliary model $Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \eta_t$, $\eta_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. The auxiliary-model likelihood is $-\sum_{t=1}^{T}(Y_t - \gamma_0 - \gamma_1 Y_{t-1})^2$. Define

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

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

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

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

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

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

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

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

$$
\begin{aligned}
(44)\qquad 0 &= \sum_t \big(Y_t - \gamma_0(\beta) - \gamma_1(\beta)Y_{t-1}\big)\gamma_0'(\beta) \\
&\quad + \sum_t \big(Y_t - \gamma_0(\beta) - \gamma_1(\beta)Y_{t-1}\big)\gamma_1'(\beta)Y_{t-1} \\
&= -\gamma_0'(\beta)\sum Y_t + \gamma_0(\beta)\gamma_0'(\beta)T + \gamma_1(\beta)\gamma_0'(\beta)\sum Y_{t-1} \\
&\quad - \gamma_1'(\beta)\sum Y_t Y_{t-1} + \gamma_0(\beta)\gamma_1'(\beta)\sum Y_{t-1} + \gamma_1(\beta)\gamma_1'(\beta)\sum Y_{t-1}^2,
\end{aligned}
$$

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

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

where $\hat{\gamma}_T \equiv \arg\min_{\gamma_0,\gamma_1} \sum_{t=1}^T (Y_t - \gamma_0 - \gamma_1 Y_{t-1})^2$. The first-order condition (F.O.C.) with respect to β is

$$
\begin{aligned}
(45)\qquad & a_{11}\big(\gamma_0(\beta) - \hat{\gamma}_0\big)\gamma_0'(\beta) + a_{12}\big(\gamma_0(\beta) - \hat{\gamma}_0\big)\gamma_1'(\beta) \\
&\quad + a_{12}\big(\gamma_1(\beta) - \hat{\gamma}_1\big)\gamma_0'(\beta) + a_{22}\big(\gamma_1(\beta) - \hat{\gamma}_1\big)\gamma_1'(\beta) \\
&= (-a_{11}\hat{\gamma}_0 - a_{12}\hat{\gamma}_1)\gamma_0'(\beta) - (a_{12}\hat{\gamma}_0 + a_{22}\hat{\gamma}_1)\gamma_1'(\beta) + a_{11}\gamma_0(\beta)\gamma_0'(\beta) \\
&\quad + a_{12}\gamma_0(\beta)\gamma_1'(\beta) + a_{12}\gamma_0'(\beta)\gamma_1(\beta) + a_{22}\gamma_1(\beta)\gamma_1'(\beta) \\
&= 0.
\end{aligned}
$$

We want to make (45) look like condition (44). To do so, first set

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

Then the last four terms in (45) match four of the six terms in (44). 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,

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

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

$$\begin{aligned}
&\text{plim}\Big[\gamma_1'(\hat{\beta}_T)\Big(T^{-1}\sum Y_t Y_{t-1} - \Big(T^{-1}\sum Y_{t-1}\Big)\hat{\gamma}_0 \\
&\qquad - \Big(T^{-1}\sum Y_{t-1}^2\Big)\hat{\gamma}_1\Big)\Big] \\
&\quad = \gamma_1'(\hat{\beta}_0)\big(\mathbb{E}Y_t Y_{t-1} - (\mathbb{E}Y_{t-1})\gamma_0(\beta_0) - \big(\mathbb{E}Y_{t-1}^2\big)\gamma_1(\beta_0)\big) = 0,
\end{aligned}$$

again, because the second term is (asymptotically) the F.O.C. that defines $\gamma_1(\beta_0) = \text{plim}\,\hat{\gamma}_1$. This shows that equations (44) and (45) are asymptotically equivalent, completing the proof.

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

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

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

$$T^{1/2}\big[\gamma_0(\beta_0) - \hat{\gamma}_0 \quad \gamma_1(\beta_0) - \hat{\gamma}_1\big]',$$

where β_0 is the true value of β.

APPENDIX B: DETAILS OF THE ECONOMETRIC PROCEDURE

B.1. *Numerical Solution*

This section describes how we compute a numerical approximation to a household's optimization problem. We iterate backward from the final period of retirement (after which it is assumed that all members of the household die) to compute the household's value function at every age. During the retirement period this iteration can be accomplished analytically because the value function has a known functional form when the period utility function exhibits constant relative risk aversion. During the working life, given a value, α^i, for the level of the household's income profile, there are three state variables: a household's wealth and its beliefs about both the slope of its income profile and the value of its persistent shock, both of which are summarized by conditional means given the history of the household's income shocks.

We choose a (coarse) grid for each of these variables (in particular, 40 points for wealth, with a higher concentration of points at low levels of wealth, 8 points for the mean belief about the slope, and 7 points for the mean belief about the persistent shock) and then interpolate between grid points to evaluate the value function at points off the grid. We combine two interpolation schemes. First, given a level for wealth, we use bilinear interpolation for the two mean beliefs; second, we use a cubic spline in wealth. Given a value function at age $t+1$ (i.e., the value function's value at each of the $40 \times 8 \times 7 = 2{,}240$ grid points), we compute the optimal savings decision at time t at every grid point. We accomplish this first by checking whether the borrowing constraint binds (by checking whether the slope of the right-hand side of the Bellman equation is negative at the constraint) and then, if not, by using a standard one-dimensional root-finding algorithm to set the first-order condition for savings to zero. Using the optimal decisions at time t, we calculate the value function at time t at each of the grid points and then proceed backward to time $t-1$. To compute the expected value (given current beliefs) of the value function on the right-hand side of the Bellman equation, we use Gauss–Hermite quadrature with four points.

We compute value functions at each age for six different values of α^i (recall that the level of a household's income profile is fixed across time and known with certainty). To compute optimal savings decisions for each of the six αs given a household's three state variables, we use an interpolation scheme like the one above (i.e., bilinear interpolation in beliefs followed by cubic spline interpolation in wealth). We then use cubic spline interpolation across the values of α^i to calculate optimal savings for a household's specific value of α^i.

Finally, the grids for α^i and the mean beliefs depend on the structural parameters. First, given a value for the standard deviation, σ_α, of α^i in the population, we construct a set of values for α^i by choosing a uniform grid of probabilities on $[0.005, 0.995]$ and then calculating arguments for a normal cumulative distribution function (c.d.f.) with mean zero and variance σ_α^2 that deliver these probabilities. Second, we simulate the cross section of beliefs in the population and then choose lower and upper bounds at each age that correspond to lower and upper tails of 0.5% in the simulated population. We then pick a uniform grid for each of the beliefs with these lower and upper bounds as end points of the grid.

We check the accuracy of the approximated value functions by calculating Euler equation errors at different points in the state space and verifying that they are close to zero when the household is unconstrained and positive when the household is constrained. Finally, we find that our structural parameter estimates are not very sensitive to increases in the numbers of points in the various grids.

B.2. *Estimation via Global Methods*

The indirect inference objective (equation (47) below) is a function of 14 variables, and is highly jagged and nonlinear. Therefore, we employed the stochastic global optimization routine described in Guvenen (2013) for maximizing this objective. In a nutshell, the algorithm begins by pre-testing a large number (N) of uniformly spaced candidate points in the parameter space and uses a multistart algorithm from the most promising (i.e., with best objective value) n^* of these pre-tested points. For the baseline estimation we used $N = 5{,}000$ and $n^* = 1{,}000$. Local optimization from these candidate points is performed via Nelder–Mead's downhill simplex algorithm and new candidate points are chosen in a way that concentrates around the most promising region of the parameter space after a certain (large) number of local optimization has been performed. Once the algorithm converges to a narrow range, we also do multiple additional restarts using Davidon–Fletcher–Powell's derivative-based optimization routine to further polish up the optimum. The global algorithm is parallel so that different local optimizations are run on separate central processing units (CPUs) to speed up computational time. For more details, see Guvenen (2013).

B.3. *Implementation: A Gaussian Objective Function*

This section describes the details of how we implement the indirect inference estimator. Loosely speaking, the indirect inference estimator is obtained by choosing the values of the structural parameters so that the estimated model and the U.S. data look as similar as possible when viewed through the lens of the auxiliary model. More concretely, define

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

to be the residuals of estimated equations (23) and (24), which is understood to equal zero when data for household i in year t are missing. The superscript "Data" specifies the data source used in the regression, which is either the PSID or the structural model (indicated by SIM (simulated)). The objective function we use is

$$\mathcal{L}(\mathbf{a}, \mathbf{b}, \Sigma, \text{Data}) = |\Sigma|^{-J/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{2,235}\sum_{t=1968}^{1993} \boldsymbol{\epsilon}_t^{i,\text{Data}} \Sigma^{-1} \left(\boldsymbol{\epsilon}_t^{i,\text{Data}}\right)'\right), \tag{46}$$

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 (23) and (24). 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 $\mathcal{L}$ as a "Gaussian objective function."

To implement the estimator, we first maximize the objective function in (46) using real data (i.e., from the PSID) to obtain a set of reduced-form parameters, denoted by $(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\Sigma})$. Next, we follow a similar procedure using simulated data. The vector of structural parameters that we estimate is

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

For a given $\boldsymbol{\Omega}$, 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 (23) and (24), which yields $\tilde{\mathbf{a}}_1$, $\tilde{\mathbf{b}}_1$, and $\tilde{\Sigma}_1$. 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 $\tilde{\Sigma}_2$. Repeat this N^{SIM} times and construct the averages

$$\tilde{\mathbf{a}} = \frac{1}{N^{\mathrm{SIM}}} \sum_{n=1}^{N^{\mathrm{SIM}}} \tilde{\mathbf{a}}_n,$$

and analogously for $\tilde{\mathbf{b}}$ and $\tilde{\Sigma}$. Then we use these averaged parameter values—estimated from *simulated* data—to evaluate the objective function (46) using the observed (PSID) data:

$$\mathcal{L}\big(\tilde{\mathbf{a}}(\boldsymbol{\Omega}, \mathrm{SIM}), \tilde{\mathbf{b}}(\boldsymbol{\Omega}, \mathrm{SIM}), \tilde{\Sigma}(\boldsymbol{\Omega}, \mathrm{SIM}), \mathrm{PSID}\big).$$

If the simulated data look 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 $(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\Sigma})$ than at $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}, \tilde{\Sigma})$ because the latter does not maximize (46) with PSID data (but instead with simulated data). Finally, the indirect inference estimator is defined as

$$\begin{aligned}(47)\qquad \hat{\boldsymbol{\Omega}} = \arg\min_{\boldsymbol{\Omega}} \big[& \mathcal{L}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\Sigma}, \mathrm{PSID}) \\ & - \mathcal{L}\big(\tilde{\mathbf{a}}(\boldsymbol{\Omega}, \mathrm{SIM}), \tilde{\mathbf{b}}(\boldsymbol{\Omega}, \mathrm{SIM}), \tilde{\Sigma}(\boldsymbol{\Omega}, \mathrm{SIM}), \mathrm{PSID}\big) \\ & + 10 \times (WY_{\mathrm{PSID}} - WY_{\mathrm{SIM}})^2\big],\end{aligned}$$

where WY is the wealth-to-income ratio defined in the text.

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 A.4, 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.

Computation of Model Specification Test Statistic. 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 ($\mathcal{L}$) given in (46) 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 $\mathcal{L}$ 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.

Specifics of the Filling-in Procedure. 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 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.

B.4. *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 in Table S.I.[35]

[35]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.

TABLE S.I
MONTE CARLO ANALYSIS

	Using Y+C Data			Using Y Data		Using Y Data		
	True	Estimates		Estimates		True	Estimates	
	Value 1	Mean	Std. Err.	Mean	Std. Err.	Value 2	Mean	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income Processes Parameters								
σ_α	0.288	0.293	0.017	0.285	0.031	0.298	0.301	0.038
σ_β	1.764	1.735	0.137	1.834	0.220	1.343	1.377	0.271
$\text{corr}_{\alpha\beta}$	−0.127	−0.106	0.102	−0.161	0.173	0.558	0.531	0.289
ρ	0.756	0.755	0.023	0.754	0.027	0.783	0.780	0.022
σ_η	0.227	0.227	0.007	0.196	0.005	0.200	0.199	0.005
σ_ε	0.100	0.105	0.016	–	–	–		
Economic Model Parameters								
λ	0.438	0.410	0.045					
$\overline{\delta}$	0.953	0.952	0.001					
ψ	0.582	0.610	0.040					
θ	0.451	0.447	0.028					
Measurement Errors								
σ_y	0.165	0.163	0.006	0.147	0.005	0.147[a]	0.146	0.005
σ_c	0.355	0.356	0.007	–		–		
σ_{c_0}	0.430	0.428	0.011	–		–		

[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.

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.[36] The results discussed here are based on 140 replications, where each Monte Carlo run takes about 20–24 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 $\text{corr}_{\alpha\beta}$, 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

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

the same true values as in the previous exercise, but estimate the income process with income data alone and with equation (24) as the only auxiliary-model regression (reported in columns 4 and 5). Perhaps, unsurprisingly, the mean estimates have little if any bias, and although the precision of the estimates falls, 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., Table S.I). Thus, another exercise we conduct is to take as the "true" values as those obtained from the 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 $\text{corr}_{\alpha\beta}$—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 multiple sources of heterogeneity and randomness.

Finally, before settling down on the auxiliary model used in this paper (equations (23) and (24)), we explored a large number of alternatives. In Table S.II, we report a subset of our results from that work, which is representative of the issues we generally encountered.[37] 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 4–9, 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 2. 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 model 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 $\text{corr}_{\alpha\beta}$. 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

[37]The Monte Carlo analysis in Table S.II was conducted using the baseline model in the working paper version (Guvenen and Smith (2010)), which differs in minor ways from the version in this text.

TABLE S.II

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]		Three		Two		Two		One	
	"True"	Estimates		Estimates		Estimates		Estimates	
	Value 1	Mean	Std. 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
$\mathrm{corr}_{\alpha\beta}$	−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.

age group and dropping insignificant regressors). We have also experimented with other auxiliary models that seemed a priori plausible and differed more dramatically from those reported here, and we found them to perform very poorly. Therefore, we do not discuss them here.

Two main conclusions emerge from those analyses. (i) 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 $\text{corr}_{\alpha\beta}$, can be pinned down fairly precisely. (ii) 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 $\text{corr}_{\alpha\beta}$ 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.

APPENDIX C: DATA APPENDIX

C.1. *Consumer Expenditure Survey Data*

1972–1973 Waves

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 1972–1973 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 S.III, we report the number of households deleted from our sample during each sample selection requirement.

1980–1992 Waves

We merge the 1972–1973 CE data with the 1980–1992 data used in Blundell, Pistaferri, and Preston (2006) (hereafter 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 1982–1984 dollars. Income is deflated using the Consumer Price Index (CPI). Total food expenditures are deflated using the average food price series provided by the

TABLE S.III

CE SAMPLE SELECTION

Selection Criterion	Dropped	Remain
Initial sample	–	19,975
Male head	4,470	15,505
Age restriction	5,200	10,305
Nonzero income and food	709	9,596
Married	874	8,722
Nonmissing education	213	8,509

Bureau of Labor Statistics. The inflation rates for food, fuel, alcohol, and transportation were determined by the corresponding price series provided by the Bureau of Labor Statistics (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.

C.2. *PSID Data*

C.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 asked only 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, the wife's asset income is listed as one variable. From 1984 to 1991, we generate the wife's asset income as the sum of the wife's share of asset income and the wife's other asset income. For 1992, the wife's asset income is the sum of the 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, the wife asset income, and asset income of other members of the household.

C.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 S.IV.

1. The individual must be from the original main PSID sample (not from the Survey of Economic Opportunities or Latino subsamples).

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

TABLE S.IV

PSID SAMPLE SELECTION

Criteria	Dropped	Remain
Initial sample	–	67,282
Main sample	39,906	27,376
Continuously married	2,805	24,571
No major composition change	4	24,567
Missing data	1,032	23,535
Outlier or top-coded	71	23,464
Male and head of household	19,232	4,232
Age restriction	429	3,803
Five observations or more	1,568	2,235

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-top-coded observations for the labor income of the head and wife and non-top-coded observations for total nonfinancial income.

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

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 National Bureau of Economic Research (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.

Measure of Net Worth. Our wealth measure includes cash and demand deposits; time and saving deposits, certificates of deposit (CDs), and money market accounts; stocks, bonds, and mutual fund holdings (including independent retirement accounts (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.

C.3. *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 the CE under additional assumptions.

In a recent paper, Blundell, Pistaferri, and Preston (2006) (hereafter, BPP) developed 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 of 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 1980–1992 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 by BPP to cover the period 1968–1992. Here we provide a brief overview of our method and a discussion of the quality of the imputation.

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 are 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.[38] The data in this earlier period also appear to be of superior quality in certain respects compared with those from subsequent waves.[39] 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 C.3.

We begin with two dimensions of consumption data that are crucial for our estimation exercise. First, the left panel of Figure S.1 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).[40] 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.[41] 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 ages 25 and

[38]The sample size is around 9,500 units in 1972–1973 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.

[39]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 do the CE waves after 1980. For example, in 1973, the fraction of 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, the fraction of consumer services in the CE accounts for 93% of the same category in NIPA in 1973 but drops to only 66% in 1989.

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

[41]If we do not use the 1972–1973 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 S.1.

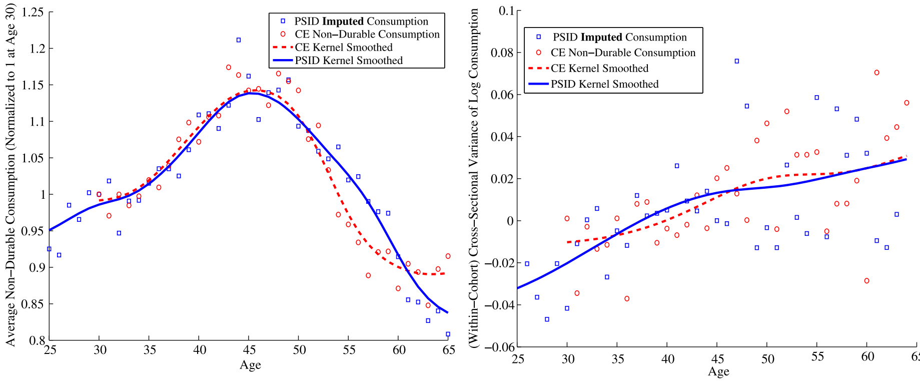

FIGURE S.1.—Mean and variance profile of log consumption over the life cycle.

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, Telmer, and Yaron (2004)), but is consistent with more recent papers that have used samples extending to the late 1990s (cf. Heathcote, Perri, and Violante (2010)). This finding is important in understanding some of the estimates (especially, λ).

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).[42] Figure S.2 plots the actual consumption of the control group against the imputed consumption 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.[43] 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 overesti-

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

[43]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.

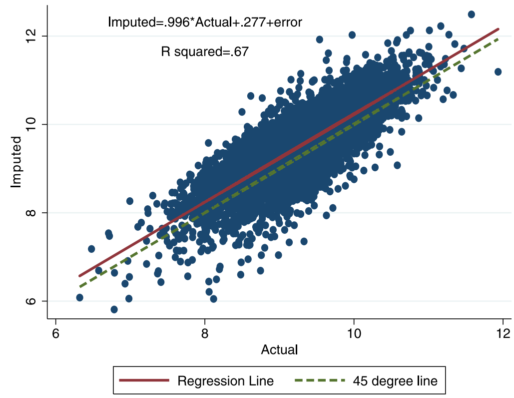

FIGURE S.2.—Out-of-sample predictive power of the imputation method in the CE. This plot is obtained by estimating the instrumental variable (IV) food demand system on a randomly chosen half of the CE sample and then imputing the consumption for the other half (control group). The figure plots the actual consumption of the control group versus their imputed consumption. The average regression slope is 0.996, the average constant is 0.24, and the average R^2 is 0.67 over 200 repetitions.

mation 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 demographic groups. Overall, we conclude that the imputation procedure works fairly well and does not result in any systematic over- or underprediction of actual consumption.

Further Details

This section describes the details of the imputation procedures and reports some further validation tests on the quality of imputation. 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 their 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–year–education specific averages of the log of the 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 S.V 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 S.3 plots the cross-sectional variance of log consumption over time. In the figure, the circles mark the CE data, whereas the squares show the imputed consumption in the PSID. Similarly, the dashed line and the solid 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 dashed-dotted 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 with 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.

APPENDIX D: ROBUSTNESS ANALYSIS

We now present results from robustness exercises. These experiments have been conducted with a version of the model in which the probability of death is set to zero until age $T = 80$, and households have access to self-insurance only (the baseline case in the working paper version, Guvenen and Smith (2010)). These two changes do not make an appreciable difference in the results so we present the robustness analysis using this slightly simpler version of our model.

D.1. *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

TABLE S.V

INSTRUMENTAL VARIABLES ESTIMATION OF DEMAND FOR FOOD IN THE CE[a]

Variable	Estimate	Variable	Estimate
$\ln(c)$	0.798***	$\ln(c) \times I\{11\% \leq \Delta \log p_{\text{fuel}}\}$	0.00386*
	(26.80)		(1.83)
$\ln(c) \times \text{age} \times I\{\text{age} < 37\}$	0.00036***	$\ln(c) \times (\text{year} - 1980)$	−0.00057
	(3.38)		(−0.68)
$\ln(c) \times \text{age} \times I\{37 \leq \text{age} < 47\}$	0.00048***	One child	0.149
	(5.45)		(1.16)
$\ln(c) \times \text{age} \times I\{47 \leq \text{age} < 56\}$	0.00042***	Two children	0.564***
	(5.75)		(3.98)
$\ln(c) \times \text{age} \times I\{56 \leq \text{age}\}$	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)
$\ln(c) \times I\{5\% \leq \Delta \log p_{\text{food}} < 8\%\}$	0.00096	Family size	0.0509***
	(1.01)		(16.20)
$\ln(c) \times I\{8\% \leq \Delta \log p_{\text{food}} < 11\%\}$	0.00858***	$\ln p_{\text{food}}$	0.581**
	(4.25)		(2.28)
$\ln(c) \times I\{11\% \leq \Delta \log p_{\text{food}}\}$	−0.00091	$\ln p_{\text{fuel}}$	−0.117
	(−0.39)		(−0.97)
$\ln(c) \times I\{5\% \leq \Delta \log p_{\text{fuel}} < 8\%\}$	0.00074	White	0.0824***
	(0.66)		(11.38)
$\ln(c) \times I\{8\% \leq \Delta \log p_{\text{fuel}} < 11\%\}$	0.00091	Constant	−1.822***
	(0.53)		(−2.65)
		Observations	21,864

[a] We pool the data from the 1972–1973 waves of the CE with the 1980–1992 waves. We instrument log food expenditures (and their interactions) with the cohort–year–education 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.

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 1 of Table S.VI, 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—

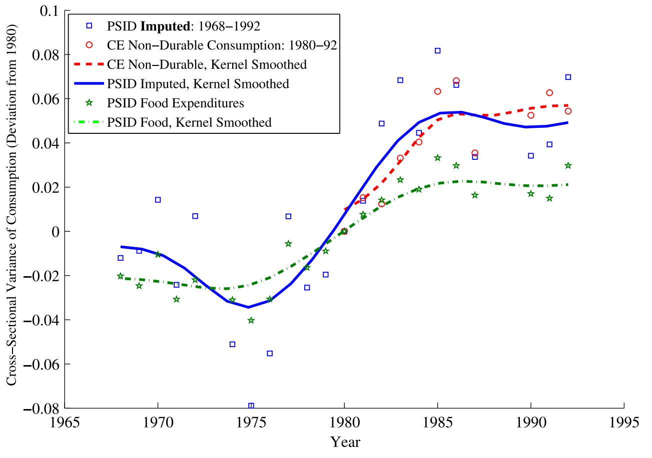

FIGURE S.3.—Cross-sectional variance of log consumption in CE and imputed PSID data: 1968–1992.

both presumably to make the model better match the wealth-to-income ratio moment. Overall, the most important parameters about the income process as well as λ seem very robust to reasonable changes in the interest rate.

D.2. *Alternative Filling-in Method*

We now examine the robustness of the estimation results to the method chosen for filling in missing data (column 2 of Table S.VI). 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 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.

D.3. *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 (column 3). 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: $\underline{a}_{25}$ rises to 41% of average income.

TABLE S.VI

SENSITIVITY ANALYSIS: ALTERNATIVE ASSUMPTIONS

	Low Interest Rate $\gamma = 0.97$ (1)	Alternative Filling-in Method (2)	Doubling Minimum Income $\underline{Y} = 0.10$ (3)	Use Data Up to Age 65 (4)	No Prior Uncertainty $\lambda = 0$ (5)	Maximum Prior Uncertainty $\lambda = \lambda^{\max}$ (6)
Income Processes Parameters (can be identified with income data alone)						
σ_α	0.284	0.220	0.293	0.228	0.268	0.248
σ_β	1.856	1.916	1.886	1.088	1.756	1.04
$\text{corr}_{\alpha\beta}$	−0.164	0.003	−0.166	−0.161	−0.086	0.751
ρ	0.755	0.760	0.759	0.801	0.777	0.806
σ_η	0.196	0.200	0.208	0.200	0.195	0.196
σ_ε	0.005	0.006	0.007	0.003	0.006	0.010
Economic Model Parameters (need consumption data)						
λ	0.380	0.327	0.374	0.520	0.0 (fixed)	0.656[a]
δ	0.964	0.950	0.951	0.943	0.954	0.951
ψ	0.790	0.921	0.757	0.992	0.761	0.895
Measurement Error and Transitory Shocks (need consumption data)						
σ_y	0.147	0.145	0.156	0.152	0.148	0.151
σ_c	0.356	0.356	0.356	0.356	0.355	0.356
σ_{c_0}	0.429	0.414	0.427	0.432	0.424	0.433
Max % constrained...	16.1%	10.2%	12.5%	9.1%	14.1%	13.2%
...at age	30	31	30	35	30	33
$\underline{a}_{25}$/mean income	0.35	0.44	0.41	0.93	0.21	0.37
$\underline{a}_{55}$/mean income	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.

D.4. *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.[44] Still, it is useful to examine how the results would change if the entire sample (up to age 65) were used in the estimation. As seen in column 4 of Table S.VI, 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 $1.08 \times 0.51 = 0.55$. Now both the variance of log income and the 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 $\lambda \times \sigma_\beta$, 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.

D.5. *Fixing Borrowing Constraints*

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 S.VII 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 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.

[44] 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.

TABLE S.VII

SENSITIVITY OF STRUCTURAL ESTIMATES TO RESTRICTIONS ON ECONOMIC MODEL PARAMETERS[a]

	Role of Preference Parameters						Borrowing Limit	
	Low RRA	*High RRA*	*δ Fixed*	*δ Fixed*	δ Estim.	*δ Fixed*	$\psi = 1$	$\underline{a}_t \equiv 0$
Preset Parameters↓	(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
$\text{corr}_{\alpha\beta}$	−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*	$\underline{a}_t \equiv 0$
Max % constrained	13.9%	10.5%	52.2%	21.7%	14.3%	12.2%	7.2%	38.1%
Max constrained age	31	33	25	35	35	33	35	25
Wealth-to-income ratio	1.08	1.08	0.94	−0.03	0.15	1.03	1.08	1.07
$\underline{a}_{25}$/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]The values of measurement errors and transitory shocks are not reported because they vary by little across specifications. The range for σ_y is from 0.129 to 0.149, for σ_c it is 0.353 to 0.356, and for σ_{c_0}, it is from 0.42 to 0.429, with the exception of column 3 for which it is 0.514. *Fixed a priori at the value indicated.

D.6. *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 5.1). 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 relative risk aversion (RRA) at, respectively, 1 (column 1 of Table S.VII) and 3 (column 2). A quick glance across these two columns 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.

D.6.1. *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 (23) and (24), 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 estimates 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.

D.6.2. *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 S.VII). 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 $\text{corr}_{\alpha\beta}$ 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 to be 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 versus 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.[45] 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 S.VII 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.)

[45]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.

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.

REFERENCES

BLUNDELL, R., L. PISTAFERRI, AND I. PRESTON (2006): "Imputing Consumption in the PSID Using Food Demand Estimates From the CEX," Working Paper W04/27. [16,20]

DEATON, A., AND C. PAXSON (1994): "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102 (3), 437–467. [22]

GOURINCHAS, P.-O., AND J. A. PARKER (2002): "Consumption Over the Life Cycle," *Econometrica*, 70 (1), 47–89. [30]

GUVENEN, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review*, 97 (3), 687–712. [1]

——— (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12 (1), 58–79. [21]

GUVENEN, F. (2013): "Quantitative Economics With Heterogeneity," Book Manuscript, University of Minnesota. [10]

GUVENEN, F., AND A. A. SMITH (2010): "Inferring Labor Income Risk From Economic Choices: An Indirect Inference Approach," Working Paper 16327, National Bureau of Economic Research. [14,24]

HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006," *Review of Economic Dynamics*, 13 (1), 15–51. [22]

JOHNSTON, D., AND S. SHIPP (1997): "Trends in Inequality Using Consumption-Expenditures: The US From 1960 to 1993," *Review of Income and Wealth*, 43, 133–152. [21]

SLESNICK, D. T. (1992): "Aggregate Consumption and Saving in the Postwar United States," *Review of Economics and Statistics*, 74 (4), 585–597. [21]

STORESLETTEN, K., C. I. TELMER, AND A. YARON (2004): "Consumption and Risk Sharing Over the Life Cycle," *Journal of Monetary Economics*, 51 (3), 609–633. [22]

Dept. of Economics, University of Minnesota, 4-101 Hanson Hall, 1925 South 4th Street, Mineapolis, MN 55455, U.S.A. and NBER; guvenen@umn.edu

and

Dept. of Economics, Yale University, 28 Hillhouse Avenue, Room 306, New Haven, CT 06520, U.S.A. and NBER; tony.smith@yale.edu.

Manuscript received July, 2010; final revision received April, 2014.