CONSUMPTION AND RISK
SHARING OVER THE LIFE CYCLE

Kjetil Storesletten
Chris I. Telmer
Amir Yaron

Consumption and Risk Sharing Over the Life Cycle
Kjetil Storesletten, Chris I. Telmer, and Amir Yaron
NBER Working Paper No. 7995
November 2000
JEL No. E21, D31

ABSTRACT

A striking feature of U.S. data on income and consumption is that inequality increases with age. Using both panel data and an equilibrium life cycle model, we argue that this is informative for understanding the importance and the characteristics of idiosyncratic labor market risk. We find that uncertainty distributed throughout the working years accounts for 40 percent of life time uncertainty, with the remainder being realized prior to entering the labor market. We estimate that the shocks received over the life cycle contain a highly persistent component, with an autocorrelation coefficient between 0.98 and unity. The joint behavior of earnings and consumption inequality, interpreted using our model, adds to the body of evidence suggesting that labor market risks are imperfectly pooled and that a precautionary motive is an important aspect of U.S. savings behavior. The restrictions imposed by general equilibrium theory play an important role in arriving at each of these conclusions.

Kjetil Storesletten
Institute for International Economics Studies
University of Stockholm
S-106 91 Stockholm, SWEDEN
and CEPR
storeslk@iies.su.se

Chris I. Telmer
Graduate School of IndustrialAdministration
Frew and Tech Streets
Carnegie-Mellon University
Pittsburgh, PA 15213
chris.telmer@cmu.edu

Amir Yaron
The Wharton School
University of Pennsylvania
2256 Steinberg-Dietrich Hall
Philadelphia, PA 19104-6367
and NBER
yarona@savage.wharton.upenn.edu

1 Introduction

Understanding the determinants of economic inequality is important for many questions in economics. It bears directly on issues as wide ranging as education policy, economic growth and the equity premium puzzle. Most existing work has focused on inequality in income, wealth and a variety of individual-specific characteristics such as educational attainment and labor market status. Relatively little attention has been paid to inequality in what these items ultimately lead to: consumption. This is unfortunate. The reason, presumably, that income and wealth inequality are of such interest is that they have an important impact on consumption inequality and, as a result, on inequality in economic welfare.

The basic question this paper asks is how important idiosyncratic income shocks are for understanding consumption inequality. To answer this, we follow an approach similar to Deaton and Paxson (1994). We first present evidence on consumption and earnings from U.S. panel data and then examine the implications for theory. The model we use is a general equilibrium life-cycle model with imperfect risk sharing. Our main finding is that idiosyncratic shocks are important: without them our model cannot account for the joint behavior of U.S. data on consumption and earnings inequality. This then leads to a number of implications. First, we find that an alternative model — a life-cycle version of the permanent income hypothesis — cannot account for the data. Data on inequality, therefore, add to the existing evidence in support of a precautionary savings motive. Second, a key ingredient is persistence. To account for consumption inequality, we find that idiosyncratic shocks must contain a persistent component, with an autocorrelation coefficient exceeding 0.75. To account for *both* consumption and earnings inequality, the autocorrelation must exceed 0.98. Finally, and most importantly, we use our model to attribute cross-sectional variation in lifetime consumption, earnings and welfare to either (a) characteristics determined before age 23, or (b) shocks realized between age 23 and retirement. It is difficult to understate the importance of this decomposition. Should the determinants of inequality be fixed early in life, for instance, then there is no risk-sharing role for either financial markets or government policies aimed at the working population such as unemployment insurance. We argue that this is not the case, concluding that working-age shocks account for at least 40 percent of total cross-sectional variation. This stands in contrast to recent work by Keane and Wolpin (1997), who find a value closer to 10 percent.

Most of our conclusions are driven by three simple features of U.S. data: (a) age-dependent inequality in labor earnings and consumption increases substantially between ages 23 and 60, (b) the increase in consumption is less than the increase in earnings, and (c) the increase in both is approximately linear. For labor earnings, we document these facts using household-level data from the Panel Study on Income Dynamics (PSID). For consumption, we cite evidence from Deaton and Paxson (1994), who use Consumer Expenditure Survey (CEX) data. They also document these

patterns in CEX labor earnings, so the facts on earnings are representative of two independent data sources.

The way in which we interpret these stylized facts lies at the heart of our results. We interpret the increase in inequality with age as arising from idiosyncratic labor market shocks which are imperfectly pooled across agents due to incomplete asset markets. There are of course alternative interpretations, but we argue that a fairly broad array of evidence supports ours. For example, we show that the pattern in PSID earnings inequality remains if one conditions on educational cohorts. This casts some doubt on an interpretation based upon predetermined heterogeneity in skills. In a related paper (Storesletten, Telmer, and Yaron (2000)) we make a similar point, using data on inequality in hours-worked to argue against a model where consumption inequality arises under complete risk sharing, driven by skill heterogeneity and non-separability between leisure and consumption. Finally, we argue that while there are many potential explanations of the earnings data which do not involve idiosyncratic risk, they are unlikely to be consistent with with increasing inequality in consumption. The simple reason is that consumption smoothing behavior tends to distribute deterministic sources of inequality evenly over the life cycle, resulting in a flat age profile.

Given this interpretation of what drives inequality, our model places restrictions on risk sharing behavior, consumption decisions and the specifics of the process governing idiosyncratic risk. Specifically, equilibrium conditions pin down the capital to output ratio and, since trading in capital is the means with which agents self-insure, the amount of risk sharing which is feasible. This in turn determines the pattern of consumption inequality, both the overall level and the differences across ages. By comparing this to the data, a number of insights emerge. For instance, our model does a good job, quantitatively, at accounting for observed consumption inequality. This lends support to papers such as Altonji, Hayashi, and Kotlikoff (1991), Attanasio and Davis (1996), Attanasio and Weber (1992), Cochrane (1991) and Mace (1991), which argue that risk sharing, especially for high frequency shocks, is considerable but far from complete. We also find support for the theory of precautionary savings by showing that, absent the precautionary motive, consumption inequality in our model is far in excess of that in the data. Finally, by experimenting with the idiosyncratic risk process, we are able to rule out certain parts of the parameter space. For instance, we are able to place a lower bound on the persistence and the volatility of idiosyncratic shocks, with values below these bounds being incapable of accounting for the increase in consumption inequality. Theory, then, in conjunction with evidence on consumption, places restrictions on the stochastic process for earnings.

Theory also influences how we approach our econometric analysis. Our approach is somewhat non-standard in that we interpret data as arising from finite stochastic processes and make strong assumptions about initial conditions. The lynchpin is the ability to condition on age. We formulate moments based on finite sequences

2

of random variables and, as a result, obtain estimators which are not subject to the well-known problems associated with non-stationary time series processes. Based only on age-dependent cross-sectional variances, we use PSID data to obtain an estimate of the autocorrelation coefficient of idiosyncratic shocks of essentially unity. The driving factor is the linear shape of the empirical age profile, less than unit-root shocks being associated with a concave shape. When we add more conventional moments to our system — autocovariances — we obtain only slightly lower estimates, in the neighborhood of 0.98. Cross-sectional variances, therefore, seem quite informative.

Further details regarding related work are as follows. Blundell and Preston (1998), Cutler and Katz (1992), Krueger and Perri (2000) and Slesnick (1993) also examine how income and consumption inequality interact, but they focus on changes over time as opposed to age. Smith and Wang (2000) show that increasing consumption inequality with age is an important aspect of how dynamic contracting works towards mitigating adverse effects of private information. Deaton, Gourinchas, and Paxson (2000) explore how risk sharing via a social security system is related to consumption inequality, something we also examine in Section 5 but in far less detail. Finally, a rich literature — including Aiyagari (1994), Chatterjee (1994), Huggett (1996), Quadrini (2000), Krusell and Smith (1998) and Castañeda, Díaz-Giménez, and Ríos-Rull (1998) — focuses on wealth inequality using a class of models similar to ours. Huggett's (1996) paper is of particular interest in that he shows how a model similar to ours can account for how wealth inequality varies over the life cycle.

The remainder of the paper is organized as follows. Section 2 presents stylized facts on labor earnings and consumption, develops a novel econometric approach and estimates a model for idiosyncratic earnings shocks. Section 3 outlines our life-cycle model, Section 4 reports quantitative results, Section 5 conducts a sensitivity analysis and Section 6 concludes.

2 Evidence

The focal point of our paper is the mapping between agent-specific labor income and consumption. Accordingly, we begin by documenting a selected set of characteristics of these variables pertaining to U.S. households. For labor income (and other types of non-wage income) we obtain data from the Panel Study on Income Dynamics (PSID) for the years 1969-1992. For consumption we exploit information from Deaton and Paxson (1994), who obtain data from the U.S. Consumer Expenditure Survey (CEX) for the years 1980-1990. The reasons for not using a single data source for both income and consumption are related to data quality. The PSID is, arguably, the highest quality source for income-related data, but is very narrow in terms of consumption, being limited to the the consumption of food. The CEX, in contrast, provides data on a broad set of consumption categories but is inferior to the PSID in terms of income. Our approach — merging the two datasets — comes at a cost in terms of consistency,

but, we believe, ultimately provides a clearer picture than would either dataset in isolation. In Appendix A we elaborate further, and also discuss the extent to which inconsistencies between the two data sources can be reconciled.

A brief overview of the salient features of our data is as follows. The consumption data from Deaton and Paxson (1994) corresponds to nonmedical and nondurable expenditures on goods and services by urban households over the years 1980-1990. We refer the reader to Deaton and Paxson (1994) for further details. Labor market and other income — what we will refer to as 'earnings' — is from the PSID and corresponds to total household wage income, before taxes, plus 'transfers' such as unemployment insurance, workers compensation, transfers from nonhousehold family members, and so on (an exact definition is provided in Appendix A). Transfers are included so that our measure of earnings risk is *net* of the risks which are pooled through these mechanisms, such mechanisms being absent from our theory. Our selection criteria *does not* focus exclusively on male-headed households, something which distinguishes it from many previous PSID-related studies, including our own (Storesletten, Telmer, and Yaron (1999a)). In addition, we explicitly allow for within-sample changes in family structure such as marriage, divorce and death. This broader-based sampling criteria — motivated in part to be consistent with CEX consumption data — is more comprehensive in terms of measuring idiosyncratic shocks (*e.g.*, divorce), while still allowing for the identification of time series parameters. The cost, of course, is increased sensitivity to measurement error, something we argue is mitigated by our focus on cross-sectional properties of the data.

In the upper panel of Figure 1 we report sample moments for what will be the focus of our theoretical explorations: the age-dependent cross-sectional variance of the logarithm of earnings and consumption. An important aspect of each of these loci is that they explicitly control for 'cohort effects' in the cross-sectional variance. Specifically, if we define an 'age-time cohort' to be all households with a head of a given age, born in the same year, then our measures of cross-sectional dispersion are *net* of dispersion which is unique to a given cohort. For instance, if dispersion among young households in the 1982 cohort was uncharacteristically high, then we identify the incremental amount of dispersion as a 'cohort effect' and remove it using methods outlined in Appendix A. These cohort effects turn out to be quantitatively important, something which Deaton and Paxson (1994) also document for CEX data. By not removing them, for instance, our estimate of the cross-sectional variance for the young (old) increases (decreases) by roughly 50 percent (20 percent), thereby making for a substantially flatter age profile. Furthermore, in Storesletten, Telmer, and Yaron (1999a), where we postulate an explicit model of cohort effects in volatility (driven by idiosyncratic shocks becoming more volatile during aggregate downturns), we find strong evidence of that such effects are an important feature of PSID data. Disentangling cohort and age effects, therefore, is an important ingredient of this paper, which abstracts from both aggregate shocks and cohort effects.

4

The important features of Figure 1, for our purposes, are as follows. Both earnings inequality and consumption inequality increase over the working part of life cycle, however only in the case of earnings does it decline at retirement. Inequality among the young is roughly the same for earnings and consumption, but earnings inequality increases at a substantially larger rate as agents age. For example, over the working years the cross-sectional standard deviation of earnings increases by roughly 80 percent, whereas that of consumption increases by only 40 percent. Finally, the rate of increase for both earnings and consumption is, loosely speaking, linear up until the retirement years. Each of these features, which are qualitatively similar to what Deaton and Paxson (1994) find using only CEX data (see their Figure 6), will be informative for our theoretical explorations. The shapes of the age-inequality profiles, for instance, provide stark restrictions on the parametric model of earnings uncertainty we formulate below. The location of consumption profile relative to the income profile is informative for issues related to risk sharing and how risk is distributed over the life cycle.

A theme of our paper is that we interpret the evidence in Figure 1 — a positive relationship between inequality and age — as indicating that households receive persistent idiosyncratic earnings shocks throughout their working lives. Moreover, because consumption inequality increases alongside earnings inequality, we interpret the data as being indicative of incomplete risk sharing. There are, however, alternative interpretations which have very different economic implications. For example, as Deaton and Paxson (1994) and others have noted, heterogeneity in skills along with non-separability between leisure and consumption can, in a complete markets world, generate consumption inequality which increases with age. In a related paper (Storesletten, Telmer, and Yaron (2000)), we use data on hours worked, consumption and earnings to argue that this is an unlikely explanation. A second alternative, relating more directly to the pattern in earnings inequality, also involves heterogeneity in skills. Suppose that wages grow faster for high skilled workers relative to low skilled workers. Suppose also that, contrary to our interpretation, most of the idiosyncratic risk which individuals face either relates to the acquisition of skills or to the return to acquiring those skills. In this case, — if idiosyncratic risk *within* skill cohorts is relatively unimportant — one would expect to see increasing inequality across the entire population, but not across households with similar skill levels. In the lower panel of Figure 1 we show that, for our PSID earnings data, this is not the case. We use educational attainment as a proxy for skills. We see that earnings inequality increases at the same rate for households with heads who have not received a high school diploma, who have received a high school diploma and who have received a college degree. Loosely speaking, the lower panel of Figure 1 shows that the overall measure of earnings inequality in the upper panel is not misleading in terms of how earnings inequality changes within education cohorts. We find this encouraging in terms of our interpretation of the data. We also use this as justification for pooling together educational cohorts for the remainder of the paper, something which is convenient

given that our model abstracts from educational choice and (explicit) heterogeneity in skills.

2.1 A Parametric Model for Earnings

There are three features of Figure 1 which are particularly important for our analysis: the level of inequality among the young, the increase in inequality with age and the rate of change of this increase with age. The level will identify the relative importance of 'fixed effects:' characteristics determined at birth and carried throughout life. The increase — a factor of three between ages 23 and 60 — will identify the conditional variance of the idiosyncratic shocks which are received over the working part of the life cycle. Finally, the (roughly) linear rate at which inequality increases with age will identify how persistent these shocks are. We now turn to a parametric model of earnings uncertainty in order to tie these three features together in a coherent manner. What follows is a relatively brief, informal treatment, with additional details relegated to Appendix A.

We associate an individual household with the age of the household head, denoted h, and the *cohort* (*i.e.*, the birth year) to which the household belongs, denoted c. Denoting the natural logarithm of earnings for the ith household of age h, belonging to cohort c, as y_{ih}^c, we follow Deaton and Paxson (1994) in decomposing the cross-sectional variance of earnings into cohort and age effects, a_c and b_h, respectively:

$$Var(y_{ih}^c) = a_c + b_h \quad . \tag{1}$$

The age effects b_h are, in fact, what are plotted in Figure 1 (after being additively scaled so as to match the overall level of dispersion in the raw data). They will also be the focal point of our theory. The cohort effects a_c play no role in our theory (it abstracts from aggregate variability) but, as we've argued above, represent an important source of variability in the data. Our statistical approach, therefore, retains Deaton and Paxson's (1994) relatively high dimensional representation of the cohort effects (*i.e.*, one parameter a_c for each cohort c), but reduces the dimensionality of the age effects to four parameters. More specifically, we think of variation in y_{ih}^c as arising from two independent sources, one common to all agents of a given cohort and the other idiosyncratic with respect to each individual agent:

$$y_{ih}^c = x_c + u_{ih} \quad ,$$

where x_c is a cohort specific shock and u_{ih} is specific to the ith individual of age h (we omit the cohort superscript from u_{ih} because its distribution does not vary across cohorts). The process for u_{ih} is

$$u_{ih} = \alpha_i + z_{ih} + \varepsilon_{ih} \tag{2}$$
$$z_{ih} = \rho z_{i,h-1} + \eta_{ih} \quad ,$$

6

where $\alpha_i \sim \mathrm{N}(0, \sigma_\alpha^2)$, $\varepsilon_{ih} \sim \mathrm{N}(0, \sigma_\varepsilon^2)$, $\eta_{ih} \sim \mathrm{N}(0, \sigma_\eta^2)$, $z_{i0} = 0$ and, therefore, $E(u_{ih}) = 0$ in the cross section. The random variable α_i — commonly called a 'fixed effect' — is realized at birth and then retained throughout life. The variables z_{ih} and ε_{ih} are realized at each period over the life cycle and are what we refer to as persistent and transitory 'life-cycle shocks,' respectively. Empirically, the latter is important for capturing actual transitory shocks as well as *i.i.d.* measurement error, a particular concern for us given our relatively broad PSID selection criteria. Theoretically, because most transitory risk get pooled in our model, ε_{ih} only matters insofar as it affects our estimate of σ_η. In this sense, the quantitative features of our theory are unaffected by our inability to distinguish between which component of ε_{ih} is measurement error and which component is a transitory shock.

Equation (2) gives rise to the following interpretation of the age effects, b_h, in equation (1).

$$b_h = Var(u_{ih}) = \sigma_\alpha^2 + \sigma_\eta^2 \sum_{j=0}^{h-1} \rho^{2j} + \sigma_\varepsilon^2 \ . \tag{3}$$

Moments related to equation (3) form the foundation of our estimators and represent a hallmark of our statistical approach. The idea is to interpret household-level data as arising from a class of finite processes and make strong assumptions about initial conditions. The resulting interaction between age and persistence can have important quantitative implications. For instance, the summation term in equation (3) is the finite analog of the familiar expression for the unconditional variance of a first-order autoregression: $\sigma_\eta^2/(1 - \rho^2)$. For small values of h coupled with large values of ρ, these two quantities are quite different and, as we'll see shortly, can lead to drastically different estimates. This approach has its drawbacks (mainly associated with initial conditions) but also its benefits. The interface with our OLG theory, for instance, is seamless. Statistically, the well-known problems associated with non-stationary time series analysis are absent since our framework is well defined for values of ρ greater than or equal to unity.

2.2 Implications of Cross-Sectional Variances

Before proceeding to formal estimation, it is informative to highlight the implications of the age-related variation in the cross-sectional variance documented in Figure 1. Specifically, we can exploit the manner in which the population variances in equation (3) change with age, h, in order to identify three of the four parameters of interest. There are many ways to do this, but the most natural is as follows. The amount of initial dispersion in the data, a variance of 0.29, identifies the sum $\sigma_\alpha^2 + \sigma_\varepsilon^2$ (these two parameters are not separately identified using only variances). The rate at which dispersion increases on average — *i.e.*, the first derivative — identifies the conditional

7

variance of the persistent shocks, σ_η^2. Finally, the curvature in the earnings locus — *i.e.*, the second derivative — identifies the autocorrelation, ρ.

This algorithm is impressionistic (things are more simultaneous than it suggests), but it captures the essence of what is easily implemented on the computer. Specifically, given a value for $\sigma_\alpha^2 + \sigma_\varepsilon^2$, values for σ_η^2 and ρ can be found to match the values b_h in equation (3) with the corresponding PSID sample moments from Figure 1. The result is depicted in Figure 2, where we plot both the PSID earnings profile and the population moments from equation (3). The associated parameter values are $\sigma_\alpha^2 + \sigma_\varepsilon^2 = 0.2735$, $\sigma_\eta^2 = 0.0166$ and $\rho = 0.9989$. We find these values to be robust to alternative algorithms, including visually matching the age profiles, using an exactly identified objective function, and using an overidentified objective function (*i.e.*, using more than three points from the PSID age profile to select 3 parameters from equation (3)).

At this informal level, then, the data suggest that idiosyncratic earnings shocks contain a component which is highly persistent, perhaps permanent. The conditional magnitude of this component is dwarfed by that of the fixed effects but, as we show below in section 2.4, its magnitude is substantial when viewed in terms of its impact on lifetime earnings as a whole. What drives these findings, in particular the near unit root, is straightforward. They are an inescapable implication of viewing the linearly increasing earnings profile in Figure 2 through the window of the age-dependent autoregressive process represented in equations (2) and (3). Given this particular process, a linear increase in dispersion can only be consistent with $\rho = 1$.

2.3   GMM Estimation

Estimates based on Figure 2 alone are useful in that the mapping between moments and parameters is very transparent. The limitations, however, are that we are unable to separately identify σ_α and σ_ε and, just as importantly, that we may be ignoring other important aspects of the data. Most obvious are autocovariances; we have estimated a dynamic parameter, ρ, without using any (explicit) dynamic information on individual households. We now turn to a more formal GMM-based approach in order to rectify this and ask to what extent exclusively focusing on Figure 2 distorts our view of the data.

In order to maintain an understanding of which moments are most informative for which parameters, we incorporate additional information in a cumulative fashion, beginning with an exactly identified system and ending with an overidentified one. An exactly identified system requires us to add one additional moment to the three cross-sectional variances we've used above. The moment we add is the cross-sectional variance of the *summation* of three time series observations on individual-specific earnings. Specifically, define s_{ih}^c as the summation of an h year old individual's

earnings over three adjacent years. Given equations (1) and (2), the variance of this sum is:

$$
\begin{aligned}
Var(s_{ih}^c) &= Var(y_{ih}^c + y_{i,h+1}^c + y_{i,h+2}^c) \quad\quad\quad (4)\\
&= 9\sigma_\alpha^2 + 3\sigma_\varepsilon^2 + \left((1+\rho)^2 + 1\right)\sigma_\eta^2 + \left(1 + \rho + \rho^2\right)^2 Var(z_{ih}) + 9a_c
\end{aligned}
$$

where $Var(z_{ih}) = \sigma_\eta^2 \sum_{j=0}^{h-1} \rho^{2j}$. The usefulness of the summation is apparent if one considers the variance of each component of the sum:

$$
Var(y_{ih}^c) = \sigma_\alpha^2 + \sigma_\varepsilon^2 + Var(z_{ih}) + a_c
$$

In equation (4), the coefficient of 9 multiplying σ_α^2, versus that of 3 multiplying σ_ε^2, is what identifies σ_ε and σ_α. It reflects the fact that three independent realizations of ε_{ih} occur over adjacent years, whereas α_i is only realized once. While the same information is contained in the autocovariances, $Cov(y_{ih}^c, y_{i,h+1}^c)$ and $Cov(y_{ih}^c, y_{i,h+2}^c)$, the summation is more useful in that we are able to incorporate information from multiple autocovariances while still maintaining an exactly identified system.

One last transformation is required. For each age, h, the last term on the right side of equation (4) implies that the moment is cohort-specific. Our approach — which is to formulate an exactly identified system of four age-specific moments — requires that we eliminate cohort specificity. For the summation moments, (4), we do so by scaling our data on s as follows.

$$
s_{ih} = \frac{s_{ih}^c}{Var(s_{ih}^c)^{1/2}} \left(9\sigma_\alpha^2 + 3\sigma_\varepsilon^2 + \left((1+\rho)^2 + 1\right)\sigma_\eta^2 + \left(1 + \rho + \rho^2\right)^2 Var(z_{ih})\right)^{1/2},
$$

where we take as given the coefficients a_c (obtained through a first-pass OLS regression, as in Deaton and Paxson (1994)). The variance of s_{ih}, which is now absent cohort effects, is simply the term in parentheses.

Table 1 reports our estimates. The moments underlying the exactly identified system are the three cross-sectional variances denoted with asterisks in Figure 2 and the variance of s_{ih}, where $h = 42$. In Appendix A we provide more details as well a discussion indicating that our estimates are not sensitive to formulating moments based on different ages, h. What we see is that the incorporation of autocovariances reduces our estimate of ρ, but only slightly. We now obtain $\rho = 0.985$ as opposed to $\rho = 0.998$ previously. We get a slightly higher estimate of σ_η^2 — 0.022 versus 0.017 — and find that the variance of the transitory shocks is roughly 23% as large as that of the fixed effect shock ($\sigma_\varepsilon^2 = 0.057$ and $\sigma_\alpha^2 = 0.242$, whereas we previously found that $\sigma_\varepsilon^2 + \sigma_\alpha^2 = 0.271$). Overall, it appears that our informal estimates based on Figure 2 are robust to this particular means of incorporating more dynamic information contained in the data.

In the third panel of Table 1 we carry the process one step further, specifying an overidentified system which consists of the same three cross-sectional variances (the

asterisks in Figure 2) but summations corresponding to *all* ages h (thus giving us 41 overidentifying restrictions). In other words, whereas the estimates in panel B rely on us choosing a particular age h in order to identify each of the four parameters, we now include moments associated with all ages and let the GMM estimator weight each moment according to its informativeness (see Appendix A). By doing so not only do we provide a check on age sensitivity, we also open up the possibility that less weight is given to what drives our near-unit-root estimate of ρ — the linear, upward sloping shape of the age profile of cross-sectional variance — and more weight is given to autocovariances, the focus of much previous work which as often found substantially smaller values for ρ. As is seen in Table 1, panel C, our estimates are barely changed. In particular, the estimate of ρ is 0.982 and the magnitude of the innovation variances is essentially the same.

Our final set of estimates, reported in Table 1, Panel D, relax the restriction that the cohort effects, a_c, are equal to those from preliminary OLS estimation. Instead, we estimate the four parameters of our time series model simultaneously with the cohort parameters, a_c, thereby achieving our initial goal of a decomposition similar to that of Deaton and Paxson (1994) but with a lower dimensional representation of the age effects. Once again we see that our previous estimates appear to be quite robust, the variances being essentially unchanged and the estimate of ρ falling to only 0.977.

The overall message at this point is that the age pattern in the cross-sectional variances (Figure 2) is quite informative. The incorporation of autocovariances implies smaller estimates of persistence — something we find reassuring given previous work — but the magnitude of the decrease is very small. Appendix A provides additional evidence, based on more conventional moment specifications, showing that as long as cross-sectional variances of the form in equation (3) are incorporated, their implication of very high persistence seems to be a robust feature of the data.

2.4 How is Uncertainty Distributed Over the Life Cycle?

Before moving on to our theory, it's informative to ask what our estimates suggest about the extent to which earnings uncertainty is resolved early in life versus being distributed throughout the life cycle. Keane and Wolpin state that

> *According to our estimates, unobserved endowment heterogeneity, as measured at age 16, accounts for 90 percent of the variance in lifetime utility. Alternatively, time-varying exogenous shocks to skills account for only 10 percent of the variation.* Keane and Wolpin (1997, pg. 515).

Our estimates of the relative magnitudes of the fixed effects, α_i, and the life-cycle shocks, η_{ih} and ε_{ih}, are informative along these lines, with fixed effects representing

10

'unobserved endowment heterogeneity' and life-cycle shocks represent 'time varying exogenous shocks.'

Specifically, representative estimates (from Table 1) of the conditional variance of the persistent and transitory shocks are $\sigma_\eta^2 = 0.023$ and $\sigma_\varepsilon^2 = 0.060$, respectively. The variance of the fixed effect shock is substantially larger, at $\sigma_\alpha^2 = 0.245$. Comparing these values, however, is misleading in that such a comparison ignores the effect of persistence on the unconditional variance as well the fact that the persistent shocks are not just realized once — as is the case for the fixed effects — but at each point along the life cycle. A more relevant comparison is the following variance ratio:

$$\frac{Var\left(\sum_{h=23}^{65} R^{-h} e^{u_{ih}} \mid \sigma_\eta = \sigma_\varepsilon = 0\right)}{Var\left(\sum_{h=23}^{65} R^{-h} e^{u_{ih}}\right)} \ , \tag{5}$$

where the discount rate, R, is set to 4.5 percent, an equilibrium outcome of our model. Equation (5) is simply the variance of the discounted stream of earnings, absent life-cycle shocks, relative to the variance inclusive of life-cycle shocks. Using the above values, with the conservative choice of $\sigma_\varepsilon = 0$ (*i.e.*, the transitory shocks are interpreted as measurement error), we compute this ratio to be 58 percent. If we incorporate a realistic life cycle pattern in the mean of earnings — described explicitly in the next section — we find the ratio to be slightly lower, at 54 percent. In either case, our results stand somewhat in contrast to those of Keane and Wolpin (1997). While we find, as they do, that a majority of a household's earnings risk is realized early in life, we differ substantially in terms of the magnitude. Their estimates suggest that just 10 percent of the uncertainty is attributable to life-cycle shocks whereas our estimate is roughly 46 percent.

An obvious limitation of these results is that they are based on the simple statistical model in equation (2) as opposed to the rich model of occupational choice which Keane and Wolpin (1997) advocate. They are also based exclusively on earnings data and ignore the implications of how U.S. consumption inequality changes with age (see Figure 1). We therefore revisit this issue after describing our theory, which imposes strong restrictions on the joint distribution of earnings and consumption. What we'll see is that while the quantitative message does not change dramatically, the evidence supportive of this message is strengthened.

3   Theory

We now formulate a stationary life cycle model in order to ask how the earnings inequality described above might manifest itself in terms of consumption inequality. The model will deliver restrictions on U.S. consumption inequality which are informative for the model itself as well as the class of earnings processes which, when viewed

11

through the window of the model, are consistent with data on consumption. Our theory has its roots in the Bewley (1986) class of models, building on previous studies by Huggett (1996), İmrohoroğlu, İmrohoroğlu, and Joines (1995) and Ríos-Rull (1994), as well as subsequent work in Storesletten, Telmer, and Yaron (1999a).

The economy is populated by H overlapping generations of agents, each generation consisting of a large number of atomistic agents. Agents are indexed by age, h, with $h \in \mathcal{H} = \{1, 2, \ldots, H\}$. Lifetimes are uncertain. We use ϕ_h to denote the unconditional probability of surviving up to age h, with $\phi_1 = 1$, and use $\xi_h = \phi_h/\phi_{h-1}$, $h = 2, 3, \ldots, H$, to denote the probability of surviving up to age h, conditional on being alive at age $h - 1$. The fraction of the total population attributable to each age cohort is fixed over time at φ_h and the population grows at rate ϑ.

Each individual agent is characterized by their age, a preference ordering over consumption distributions, an endowment process and an asset market position. Preferences for an unborn agent are identical across agents and are represented by,

$$ E \sum_{h=1}^{H} \beta^h \phi_h U(c_h) \ , \tag{6} $$

where U is the standard twice differentiable, strictly concave utility function and the expectation is taken with respect to the economy's stationary probability distribution.

Agents begin working at age 22 and, conditional on surviving, retire at age 65. After retirement they must finance consumption entirely from an existing stock of assets. Prior to retirement an agent of age h receives an annual endowment, n_h, of an age-specific amount of labor hours (or, equivalently, productive efficiency units) which they supply inelastically to an aggregate production technology. Individual labor income is then determined as the product of hours worked and the market clearing wage rate.

Aside from age, heterogeneity is driven by idiosyncratic labor market uncertainty. We adopt the following process for the logarithm of hours worked,

$$ \log n_h = \kappa_h + \alpha + z_h + \varepsilon_h \ , \tag{7} $$

where κ_h are a set of parameters used to characterize the cross-sectional distribution of mean income across age cohorts, $\alpha \sim \mathrm{N}(0, \sigma_\alpha^2)$ is a 'fixed effect,' determined at birth, $\varepsilon_h \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ is a transitory shock received each period, and z_h is a persistent shock, also received each period, which follows a first-order autoregression:

$$ z_h = \rho z_{h-1} + \eta_h \ , \ \ \eta_h \sim \mathrm{N}(0, \sigma_\eta^2) \ , \ \ z_0 = 0 \ . $$

This process is a direct analog of what we estimated in the previous section, the only difference being that, here, we specify a process for hours worked and not labor

earnings. The difference, however, will be an additive constant (the wage rate in our stationary equilibrium will be constant), thereby making the distinction innocuous.

Output in this world is produced by an aggregate technology to which individuals rent their labor services and capital. Denoting per-capita output as Y, the production technology is of the form

$$Y = Zf(K, N) \ ,$$

where K and N denote per-capita capital and labor supply, respectively, and Z is a constant which governs secular growth. We appeal to a cross-sectional law of large numbers to argue that K and N are constants and normalize N to unity. The competitively determined wage and capital rental rates are denoted W and R, respectively. Denoting aggregate consumption as C, and given a rate of depreciation on aggregate capital of δ, the law of motion for aggregate capital is

$$K' = Y - C + (1 - \delta)K \ .$$

There is one asset – capital – which pays a return R in addition to a survivor's premium which we use to represent an actuarily fair annuity. Specifically, the total return on capital invested at age $h - 1$, conditional on surviving to age h, is R/ξ_h (recall that ξ_h is the conditional survival probability). Agents who do not survive to age h receive zero return and their assets are, effectively, redistributed throughout the surviving population.

Dynamic Programming Problem and Equilibrium

The cross-sectional distribution for this economy can now be represented as a function μ, defined over an appropriate family of subsets of $S = (\mathcal{H} \times \tilde{\mathcal{Z}} \times \mathcal{A})$, where $\tilde{\mathcal{Z}}$ is the product space containing all possible idiosyncratic shocks and $\mathcal{A}$ is the set of possible capital holdings. In words, μ is simply a distribution of agents across ages, idiosyncratic shocks, and asset holdings. Because our economy does not feature aggregate shocks, we can appeal to cross-sectional law of large numbers to ensure that, in any stationary equilibrium, the distribution μ is unique (Aiyagari (1994)).

Recalling that R and W denote the (market clearing) return on capital and the wage rate, respectively, the decisions of an agent of age h are governed by the following constraints.

$$
\begin{aligned}
c_h + a'_{h+1} &\leq a_h R/\xi_h + n_h W & (8)\\
a'_{h+1} &\geq \underline{a}(\alpha, z) \ , \quad a_{H+1} \geq 0 \ ,
\end{aligned}
$$

where a_h denotes beginning of period asset (or capital) holdings, a'_{h+1} denotes end of period asset holdings, and $\underline{a}(\alpha, z)$ denotes a (possibly) state-dependent borrowing

constraint. Our timing convention is that savings decisions are made at the end of the current period, and returns are paid the following period at the realized capital rental rate (adjusted by the annuity factor, ξ_h).

Denoting the value function of an h year old agent as V_h, their choice problem can be represented as,

$$V_h(\alpha, z_h, \varepsilon_h, a_h) = \max_{a'_{h+1}} \left\{ U(c_h) + \beta \frac{\phi_{h+1}}{\phi_h} E \left[V'_{h+1}(\alpha, z'_{h+1}, \varepsilon'_{h+1}, a'_{h+1}) \right] \right\} \quad (9)$$

subject to equations (8).

An equilibrium is defined as prices, R and W and a set of cohort-specific functions, $\{V_h, a'_{h+1}\}_{h=1}^{H}$ such that (a) the firm's profit maximization problem is satisfied ($R = Zf_1(K, N) - \delta + 1$ and $W = Zf_2(K, N)$), (b) individual optimization problems are satisfied ($\{V_h, a'_{h+1}\}_{h=1}^{H}$ satisfy equations (9)), and (c) markets clear and aggregate quantities result from individual decisions ($K = \int_S a_h \, d\mu$ and $N = \int_S n_h \, d\mu$).

We solve the individuals' optimization problems based on piecewise linear approximation of the decision rules, and follow Huggett (1993) and Aiyagari (1994) in solving for the equilibrium interest rate.

4   Quantitative Results

We interpret one period in our model as corresponding to one year of calendar time. The aggregate production technology is Cobb-Douglas:

$$Y = ZK^\theta N^{1-\theta} \ ,$$

Following Cooley and Prescott (1995), we set θ equal to 0.4 (which corresponds to capital's share of national income being 40%) and allow for a 7.8% annual depreciation rate on the aggregate capital stock. The secular growth rate in GNP per capita, by which we normalize all individual quantities in our model, is chosen to be 1.5% per year.

Each agent's preferences are identical (up to age-dependent mortality risk) and are described by equation (6). We parameterize the period utility function with the standard isoelastic specification,

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

We set γ to 2 and the utility discount factor, β, is set so as to match the average capital to output ratio in the U.S of 2.9. This gives rise to a real return on capital, R, of 5.99 percent. Demographic variables are chosen to correspond to simple properties

14

of the U.S. population. Agents are 'born' at age 22, retire at age 65 and are dead by age 100. 'Retirement' is defined as having one's labor income drop to zero. Mortality rates are chosen to match those of the U.S. females in 1991 and population growth is set to 1.0%.

The process for idiosyncratic labor income, equation (7), is implemented as a discrete approximation to the autoregressive time series model in equation (2).[1] Parameter values are chosen to correspond to our point estimates in Table 1 Panel A, except that we choose $\rho = 1$ instead of $\rho = 0.9989$. The choice of a unit root specification is motivated by several factors. First, it allows us to exploit a normalization which greatly reduces our computational burden.[2] Second, unit root economies facilitate a useful comparison with other theories of consumption choice, most notably the life cycle version of the permanent income hypothesis. Third, much of the focus of our quantitative questions is on the age profile of the cross-sectional variance of earnings and consumption. Economies with $\rho < 1$ will, in general, overstate the degree of curvature in the earnings locus, thereby making for a more difficult interpretation of what drives differences in theoretical and empirical consumption profiles. A unit root, on the other hand, matches the theoretical and empirical earnings profiles, allowing us to look elsewhere for what drives differences in consumption. Finally, in Section 5 we demonstrate that the implications of using $\rho = 1$ instead of $\rho = .98$ (*i.e.*, values from Table 1, Panels B-D) are of minor importance for consumption behavior. Our results, therefore, are equally valid for near-unit-root shocks — values of ρ between 0.98 and 1.00 — as they are for unit-root shocks.

One final parameter is the the lower bound on asset holdings or, equivalently, the borrowing constraint. For the case of unit root shocks (where we solve a normalized version of our economy) the constraint is $\underline{a}(\alpha, z) = \underline{a}e^{\alpha+z}$, with the constant $\underline{a}$ set so that the average agent cannot borrow in excess of the (annual) level of GNP per capita.[3] For economies with $\rho < 1$, $\underline{a}(\alpha, z)$ is a constant, set equal to the negative of annual GNP per capita.

---

[1] In the case of $\rho = 0.98$, we approximate the autoregressive process with a 21-state Markov chain, following the approach in Huggett (1996).

[2] The unit root in endowments implies that the ratio of future endowments over the exponent of the sum of previous shocks, e^z, is identically distributed, irrespective of e^z. Moreover, since preferences are homothetic and current endowments and borrowing constraints are constant fractions of e^z, one can show that the ratio of optimal consumption to e^z is independent of z. Thus, one does not have to keep track of z as a state variable, and one needs only take into account the distribution of the *innovation* to the shock, which we approximate with an *i.i.d.*, two-state Markov chain.

[3] In our benchmark economy, with unit-root shocks, the borrowing constraint does not bind for any agent in any state of nature, irrespective of $\underline{a}$. Only for economies with $\rho < 1$ is the value of $\underline{a}$ relevant.

4.1 Risk Sharing

A useful context in which to interpret our results is the process for marginal utility implied by the first-order conditions of an unconstrained agent,

$$U'(c^h) = \frac{1}{R\beta} U'(c^{h-1}) + v^h \ , \tag{10}$$

where $v^h \equiv U'(c^h) - E[U'(c^h)]$ is the innovation in marginal utility experienced by an h year old agent. This innovation is, by definition, conditionally uncorrelated with anything in the agent's information set when they were $h-1$ years old. Its variance, however, can display a variety of different patterns. Most models with a life cycle component, for instance, feature the variance of v^h being age dependent.

Equation (10) is relevant for interpreting our results because the variance of v^h is at the heart of what we think of as risk sharing and the implications for consumption inequality. If markets were complete, for instance, the variance of v^h would be zero — recall that our economy abstracts from aggregate uncertainty — and, given homothetic preferences and separability between consumption and leisure, consumption inequality would be constant over the life cycle. Autarky, on the other hand, would result in the variance of v^h being closely associated with that of the innovations in labor earnings, and the pattern of consumption inequality would be similar to the pattern of earnings inequality we see in Figure 1. Partial risk sharing, which characterizes our model, implies that the variance of v^h lies somewhere in between, the exact location being governed by the risk sharing technologies with which we endow our economy.

Equation (10) also captures the sense in which the equilibrium aspect of our model plays a fundamental role. Equilibrium conditions impose restrictions on the variance of v^h — and therefore restrictions on the pattern of consumption inequality — which go beyond those obtainable by studying the first-order condition (10) in isolation. In simple terms, an individual who dissaves in the face of an adverse shock (or saves less than they would otherwise) is only able to do so insofar as some other agent saves in response to a good shock. Moreover, the set of feasible exchanges between these agents is constrained by the financial market structure. These constraints have a direct bearing on risk sharing allocations and, therefore, on the variance of v^h. In more specific terms, our model's specification for preferences and technologies (including the stochastic processes for labor earnings) imply a particular level of aggregate capital which must serve as the vehicle with which agents collectively buffer their idiosyncratic shocks. In relation to equation (10), this is the same as saying that the rate of return on savings, R, is endogenous and cannot be chosen independently of time preference, β, and the function $U(\cdot)$. As we'll see, the interaction of these variables and their impact on savings behavior is quantitatively important for the amount of risk sharing which financial markets can facilitate and, therefore, for the variance of v^h.

4.2 Implications of the Life Cycle/Permanent Income Hypothesis

In order to best understand our results, note that agents in our model have two motives for saving: a life cycle motive and a precautionary motive. It is informative to begin by abstracting from the latter. The dotted line in Figure 3 reports the implications of the life cycle version of the permanent income hypothesis (PIH), a framework characterized by certainty equivalence and, therefore, the absence of a precautionary motive. More precisely, the dotted line represents the implications for consumption inequality, should each agent face an environment identical to that in our benchmark economy, but have quadratic preferences and a rate of time preference equal to the benchmark rate of return on capital. What we see is that the PIH overstates the degree of consumption inequality, both relative to the data and relative to our benchmark economy. For our particular parameterization, the PIH predicts that consumption inequality will *exceed* earnings inequality over most of the working years. A striking and robust feature of the data is that the opposite is true.

The economics underlying the PIH graph are straightforward and can be derived analytically.[4] Because of a life-cycle pattern in average income (which we calibrate to U.S. data), PIH consumers hold negative financial wealth, on average, until age 40. That is, they borrow against future wage receipts. A crucial variable for determining at what rate consumption inequality disperses is, for a given age, the average fraction of total wealth which is represented by financial wealth, where total wealth is defined as financial wealth plus human wealth (*i.e.*, claims to future wage payments). If financial wealth is a large fraction of total wealth, then (persistent) idiosyncratic shocks will have a small impact on total income, on consumption decisions, and therefore on the cross-sectional variance of consumption. The opposite holds if financial wealth is small relative to total wealth. The pattern we see in the PIH line of Figure 3 reflects these effects. During the years in which the average household borrows, consumption inequality increases at a more rapid rate than earnings inequality. Once the average household begins to save (in order to finance retirement consumption), consumption inequality still increases, but at slower rate relative to earnings inequality. Quantitatively speaking, the degree of convexity and concavity which this gives rise to is difficult to distinguish from the data. What is true, however, is that the overall level of consumption inequality, relative to earnings inequality, is substantially larger than what we see in the U.S. economy.

Our model, in contrast, generates a far more realistic picture. The dashed line in Figure 3 shows that consumption inequality is everywhere less than earnings inequality, as is the case in the data. The overall increase in inequality — which is roughly 100 percent in the data — is 125 percent in our model versus over 300 percent for the PIH. We attribute the differences to two features of our framework: a precautionary

---

[4]We have done so based on a generalization of the calculations in Deaton and Paxson (1994), page 457.

savings motive and a general equilibrium restriction. The latter is that time preference, β, is chosen not to equal the rate of return on savings, as in the PIH, but to pin down a realistic level of aggregate savings. Aggregate savings is the vehicle with which society collectively pools risk, so this restriction is crucial for the location of the consumption inequality profile relative to the earnings profile. The former — the precautionary savings motive — is a characteristic of our model's convex marginal utility (Kimball (1990)) and has played an important role in accounting for data on *the level* of life cycle savings (*c.f.*, Carroll (1997), Carroll and Samwick (1998), Deaton (1992), and Zeldes (1989)). Our results provide additional support for a precautionary motive by focusing on its importance in accounting for consumption *dispersion*. We find the stock of savings generated by a precautionary motive to be an important factor in generating the lower level of consumption inequality relative to earnings inequality — that is, the risk sharing behavior — which we see in the data. The key, as above, is the ratio of financial wealth to total wealth. With a precautionary motive, agents begin to accumulate wealth very early in life, not even choosing to borrow against an increasing average profile of lifetime earnings. As financial wealth grows relative to human wealth the impact of idiosyncratic shocks on consumption decisions, and on consumption inequality, is mitigated.

4.3   Understanding the Benchmark Economy

Figure 3 shows that, while our economy does a better job than the PIH, it still generates consumption inequality which is excessive relative to the data. We now present results from computational experiments which illustrate why this is, how our model might be modified, and what we learn about risk sharing in the real world.

The fact that risk sharing seems understated in our economy leads us to question two sets of assumptions: those related to the statistical processes governing earnings uncertainty and those related to the technologies with which agents share risk. We defer the latter until Section 5 and begin with the former. There are three important aspects of earnings uncertainty which are important for understanding our results.

A. *A realistically shaped profile of earnings inequality is neither necessary nor sufficient to generate a realistically shaped profile of consumption inequality.*

The importance of this is that consumption inequality does not simply reflect whatever properties we assign to earnings inequality. What matters goes deeper than that, reflecting how consumption and risk sharing respond to the statistical properties of earnings uncertainty. To illustrate this we consider two polar cases. First we show that we can formulate an earnings process with a cross sectional variance identical to that in Figure 3, but which generates a strongly counterfactual pattern in consumption inequality. Next, we do the opposite and show that there exist counterfactual earnings processes which generate reasonable looking consumption profiles.

Each of these cases are illustrated in Figure 4. The dashed lines demonstrate insufficiency. These lines represent earnings and consumption inequality in an economy with *i.i.d.* earnings shocks (*i.e.*, $\rho = 0$) which are heteroskedastic in terms of age. In other words, increasing earnings inequality is generated by increasing conditional variance, as opposed to constant conditional variance in conjunction with persistence. We see that the implications for consumption are dramatic; inequality increases by less than 15% whereas it more than doubles in both the data and our benchmark economy (Figure 3). Persistence is therefore crucial.

The dotted lines in Figure 4 demonstrate that the converse holds. That is, they represent an economy in which consumption inequality is similar to what we see in the data but earnings inequality is strongly counterfactual. Again, autocorrelation plays an important role. The earnings process in this case has homoskedastic shocks, but $\rho = 0.86$. What generates the flat dispersion profile is that we set the variance of the initial conditions (the variables z_{i0}) equal to the variance of the the stationary distribution of the z process. Increasing earnings inequality, then, is not necessary for increasing consumption inequality.

From an econometric perspective, the upshot of this point is that the process for earnings is not identified by the age profile of consumption inequality alone. It is identified, however, by the combination of the age profiles of earnings and consumption (*i.e.*, there are lots of earnings processes which give rise to a realistic consumption profile, but only one which also fits the facts on earnings). This is the sense in which our model is an effective way to discriminate between alternative processes for earnings.

B. *Highly persistent idiosyncratic shocks are necessary to account for increasing consumption inequality*

The necessity of persistence has already been suggested by the preceding discussion. A stronger statement is as follows. Processes for earnings inequality with $\rho < 0.75$ cannot account for the magnitude of the increase in consumption inequality which we see in the data. This is true, irrespective of the implications of such processes for earnings inequality. It is also true irrespective of whether or not we allow ourselves the flexibility of letting the conditional variance function be age dependent.

These statements are based on computational experiments in which we choose a particular value for ρ and then select the remaining parameter values to maximize the increase in consumption inequality, subject to the constraint that the average level of earnings inequality matches that of the U.S. data. In each case, a value of $\rho < 0.75$ turns out to be insufficient for generating a realistic increase in consumption inequality. Several examples are illustrated in Figure 5. Each theoretical line represents consumption inequality for an economy in which the conditional variance of the persistent shocks is chosen to match the average level of earnings inequality.

We see lines associated with $\rho < 0.75$ are deficient relative to the data. Thus, even if we ignore the counterfactual implications for earnings inequality of ρ being substantially less than unity (earnings inequality is far too concave in these cases), we need a relatively large value of ρ in order to account for consumption.[5]

C. *A substantial component of overall earnings uncertainty must arise in the form of persistent shocks, received over the life cycle.*

Equivalently, the conditional variance of the persistent shocks, σ_η, must be of sufficient magnitude if our theory is to be consistent with the increase in consumption inequality. While this is to a certain extent obvious, given the previous point about the necessity of persistent shocks, it is worth emphasizing in a different context. As we discussed in Section 2.4, Keane and Wolpin's (1997) work suggests that 90 percent of the average household's lifetime earnings uncertainty is resolved at quite an early age. Our results suggest that this may be inconsistent with the joint behavior of earnings and consumption inequality which we see in U.S. data. More specifically, we find that, in the context of our benchmark economy with unit-root shocks, the smallest value of the conditional variance, σ_η^2, which can generate a realistic increase in consumption inequality is 0.0154, which is only 7 percent smaller than our benchmark value of 0.0166.[6]

It is informative to view this result — the minimum volatility in the persistent shocks which can account for the increase in U.S. consumption inequality — in the context of the variance decomposition of Section 2.4. Using the above parameter values, $\rho = 1$, $\sigma_\eta^2 = 0.0154$, $\sigma_\alpha^2 = 0.2970$ and $\sigma_\varepsilon^2 = 0.0041$, we compute the variance ratio in equation (5) to be 55 percent (the value from our unit-root, benchmark economy is 51 percent). The upshot, therefore, is that stochastic life-cycle shocks must account for *at least* 45 percent of the lifetime variation in labor earnings in order to fit the facts on increasing consumption inequality.

4.4   Discussion: A Puzzle and a Solution

We take several lessons from our analysis to this point. First, there seems to be a tension between accounting for the earnings data and accounting for the consumption data, where the battle is fought primarily in terms of the autocorrelation coefficient. If

---

[5]Note that relatively small values for ρ also generate the counterfactual implication that earnings inequality decreases slightly between ages 35 and 45. This is a result of an increasingly important role played by borrowing constraints, something we discuss further in Section 5.

[6]We arrive at this conclusion by lowering the magnitude of σ_η^2 while simultaneously increasing the size of the fixed-effect shocks, σ_α^2, so as to maintain the same average level of earnings dispersion. The resulting value for σ_α^2 is 0.2970, a 9 percent increase relative to the benchmark. The value of $\sigma_\eta^2 = 0.0154$ is that which delivers an average increase in consumption dispersion which matches the U.S. data.

we restrict ourselves to shocks which are homoskedastic in terms of age, the earnings data are strongly suggestive of a unit root. The consumption data, on the other hand, suggest more risk sharing than our model can generate with unit-root shocks. Consumption, therefore, seems to indicate less persistent shocks. The important caveat is the linear shape of the consumption profile. Even with unit-root shocks, our model generates a consumption locus which appears to be excessively concave. This concavity is driven primarily by the same type of 'portfolio effect' as was discussed above: agents accumulate wealth as they age and, as a result, the consumption effect of an idiosyncratic shocks diminishes with age.

A second lesson is that theory is helpful in distinguishing between alternative statistical models for earnings uncertainty. In other words, because there are many ways to generate a given pattern of earnings inequality, our model plays an important role in identifying which are consistent with the evidence on consumption inequality. A stark example is inherent in point A of the previous section. The dashed earnings profile in Figure 4 is identical to that from our benchmark economy, but, instead of being generated by persistent shocks, is generated by $i.i.d.$ shocks with a conditional variance which increases with age. This process matches the evidence on earnings but, through the window of our model, has counterfactual implications for consumption (the lower, dashed line in Figure 4). Our model, therefore, imposes informative restrictions on earnings uncertainty, suggestive of the importance of persistence.

Finally, the tension related to persistence is suggestive of one of two things. Either we loosen-up our benchmark statistical model for earnings uncertainty, or we explore alternative risk sharing technologies in our theory. We turn to the latter in the next section. An example of the former is that we can allow the conditional variance of the persistent shocks to change with age, something we've already employed above in point A. We pursue this further in Figure 6, where we show the implications of alternative values of ρ, where the age-dependent heteroskedasticity in the persistent shocks is specified so as to match the evidence on increasing earnings uncertainty. We see that a value of $\rho = 0.85$ succeeds in accounting for both consumption and earnings inequality. Apparently, the flexibility afforded by age-dependent heteroskedasticity resolves the tension between consumption and earnings by allowing for lower auto-correlation — and therefore more risk sharing — while maintaining an earnings profile which increases linearly.

How plausible is this? The key feature of the economy with heteroskedasticity (the graph with $\rho = 0.85$ in Figure 6) is that idiosyncratic shocks become increasingly volatile with age. We are unaware of the empirically plausibility of older workers receiving idiosyncratic shocks which are more volatile than those faced by younger workers.[7] What is clear, however, is that this is a strong, testable restriction implied by our model. Moreover, it is a restriction which applies to models of consumption choice more generally ($e.g.$, the PIH), many of which also generate concave dispersion

---

[7]See, however, the model and discussion of health shocks in Hubbard, Skinner, and Zeldes (1994)).

profiles due to the life-cycle pattern in wealth accumulation which we've highlighted above. In a nutshell, linearity in both the earnings and the consumption profile, with the latter lying beneath the former, may be difficult to account for without this particular pattern of heteroskedasticity over the life cycle.

4.5 The Life Cycle Distribution of Uncertainty Revisited

In Section 2.4 we decomposed the uncertainty earnings into a component which is realized early in life and a component which is realized along the life cycle. We found that 46 percent of the variability in lifetime earnings is attributable to life-cycle shocks. Our theory now allows us to revisit this question in a richer economic context, conducting the decomposition with regard to consumption and welfare as opposed to the exogenously-given earnings process.

We begin with a calculation which is directly comparable to that in Keane and Wolpin (1997): a decomposition of the variability in lifetime utility. Denote realized utility, along some random path, as $w(\alpha, \eta, \varepsilon)$:

$$w(\alpha, \eta, \varepsilon) = \sum_{h=1}^{H} \beta^h \phi_h U(c_h) \ .$$

If w is evaluated at the optimum, then the conditional mean, $E(w \,|\, \alpha)$, is the value function V conditional on some realized fixed effect, α. We denote $V_\alpha = E(w \,|\, \alpha)$. The total variance in realized utility can therefore be decomposed into variance attributable to fixed effects (i.e., variance in these conditional value functions) and the average variance, conditional on fixed effects:

$$\begin{aligned} Var(w) &= Var(E[w \,|\, \alpha]) + E(\,Var[w \,|\, \alpha]) \\ &= Var(V_\alpha) + E(\,Var[w \,|\, \alpha]) \ . \end{aligned}$$

Our interpretation of Keane and Wolpin (1997) is that $Var(V_\alpha)/Var(w)$ is roughly 90 percent. For our economy, after converting w into monetary equivalents (to be consistent with Keane and Wolpin (1997)) we find this ratio to be 0.65. Fixed effects, therefore, account for more of the variation in lifetime utility than they do the variation in the present value of lifetime earnings. Given that utility is defined over consumption, and that agents are able to use financial markets to self insure against life-cycle shocks but not fixed-effect shocks, this is not surprising.

Accounting for a majority of the variation in lifetime utility, however, does not necessarily imply precedence in terms of welfare. A more direct decomposition of how fixed and life cycle effects impact welfare is as follows. Consider an unborn agent. Denote ψ_α as the percentage reduction in per-period consumption which, when combined with receiving a fixed effect shock $\alpha = 0$ with certainty, would make them indifferent to living in our benchmark economy which does include fixed-effect

uncertainty. In Appendix B we show that $\psi_\alpha = 1 - \exp(-\gamma\sigma_\alpha^2/2)$. Denote ψ_u as the analogous consumption reduction, only where we eliminate both the life-cycle shocks and the fixed-effect shocks. Appendix B also shows that ψ_u can be computed up to a scale factor which must be computed by simulation. The ratio ψ_α/ψ_u reflects the relative valuation, in percentage units of consumption, of the fixed-effect shocks. Our benchmark economy yields values of $\psi_\alpha = 0.24$, $\psi_u = 0.40$ for a ratio of

$$\frac{\psi_\alpha}{\psi_u} = 0.60 \ \ .$$

Alternatively, we might express the value of eliminating the fixed effects relative to ψ_l, the value of getting rid of just the life cycle shocks. We compute ψ_l to be 0.22, implying that the welfare cost of fixed effects is just 9 percent larger than that of life-cycle effects. In either case the implication is that fixed effects are relatively less important from a welfare perspective than they are in accounting for variance in lifetime utility. The reason is related to how the life-cycle shocks affect welfare. Agents dislike life-cycle shocks for two reasons: the variability which they represent and the fact that they motivate the accumulation of a precautionary buffer stock of savings. The latter has a utility cost in and of itself (it generates a less flat life pattern in average consumption), thereby implying that the welfare implications of life-cycle shocks are larger than the fraction of the variability in realized utility which they account for.

In summary, we find that uncertainty which is realized early in life accounts for somewhere between 54 and 65 percent of the lifetime uncertainty which agents face. The lower number is based on a decomposition of the variance in lifetime earnings (Section 2.4). The higher is based on the variance in lifetime utility. To us, however, the most natural measure of the importance of a particular source of uncertainty is how much an agent will pay to eliminate it. By this metric, fixed effects represent roughly 60 percent of the welfare cost of total idiosyncratic risk.

These results stand in contrast to those of Keane and Wolpin (1997). We attribute the differences to two factors. They assume, for computational reasons, that idiosyncratic shocks are *i.i.d.* whereas high persistence plays a central role in both our model and our empirical analysis. They limit their attention to ages 16 through 26, whereas we consider the entire life cycle. The latter is important for our results, which depend on the overall increase in inequality between age 23 and retirement.

5 Sensitivity Analysis

The implications of our model are robust to a number of alternative statistical processes for earnings uncertainty and a number of alternative risk sharing arrangements. The specifics are as follows.

(a) *Stationary shocks.* Most of our results have been based on an economy with unit-root shocks. In Section 4 we justified this on several grounds, but the fact remains that the bulk of the evidence in Section 2 and Table 1 points to estimates of ρ less than unity but greater than 0.96 (incorporating the standard errors). In Figure 7 we report results analogous to our main results (Figure 3), only where we use parameter values which are representative of those from Table 1. We set $\rho = 0.984$ and, in order to maintain a clean comparison, keep σ_α^2 fixed at its benchmark value. We then choose σ_η^2 to match the average level of earnings dispersion in the data. We see that the theoretical earnings locus is now more concave than in the data, owing to a value of ρ less than unity, and that consumption dispersion is slightly *higher* than in our benchmark economy, Figure 3.[8] Overall, we find that a calibration based on Panels B-D of Table 1 generates earnings inequality which is somewhat inconsistent with the data, but consumption inequality which is quite similar to that in our benchmark economy. The former is not surprising, given that these estimates incorporate information far beyond that contained in the cross-sectional variances for earnings. The latter is the sense in which our results are robust to the near-unit-root behavior which we conclude is a good characterization of our PSID data.

(b) *Social security.* In Section 4.3 we attributed excessive consumption inequality to either the statistical process for earnings uncertainty — which we subsequently examined — or the variety of risk sharing technologies from which our economy abstracts. In regard to the latter, we have analyzed versions of our model which incorporate a pay-as-you go social security system similar to that in Storesletten, Telmer, and Yaron (1999b). As one would expect, the inclusion of a socially provided risk sharing technology reduces consumption inequality relative to our benchmark economy, thus providing a better account of U.S. data. The magnitude, however, is not large. In contrast, social security generates relatively more concavity in the consumption profile — it reduces inequality among the elderly — and thus offers a relatively inferior account of the linear shape of the U.S. profile. Deaton, Gourinchas, and Paxson (2000) provide an more in-depth analysis of how social security and consumption inequality interact. They find that social security reduces inequality, as do we, and go on to examine the impact of reforms such as individual accounts.

(c) *Borrowing constraints.* In economies with unit-root shocks the borrowing constraint $\underline{a}$ is never binding, irrespective of the level at which we set $\underline{a}$. Our benchmark results are therefore independent of issues related to borrowing constraints. For economies with decreasingly less persistence, borrowing constraints

---

[8]The fact that consumption dispersion is higher with $\rho = 0.984$ than with $\rho = 1.0$ is, interestingly, due to the average increase in consumption dispersion over the lifetime being hump-shaped in ρ, with the peak occurring somewhere between 0.90 and unity. This property is also a feature of the version of the PIH discussed in Section 4.2, something we have verified analytically.

bind more frequently and generate a hump in the consumption dispersion profile. This is evident in Figure 5. For values of ρ in the range of our estimates, however, we find that our results are quite robust to alternative specifications of the borrowing constraint.

(d) *Risk aversion.* By increasing the risk aversion coefficient γ to as much as 7, and simultaneously reducing the value of the utility discount factor β to match the U.S. capital to output ratio, we find that the implications for consumption dispersion are inconsequential. Our results are therefore not sensitive to changes in risk aversion, *per se*. Of course, were we to hold time preference fixed while increasing γ, the level of precautionary savings would increase, the economy's capital to output ratio would become unrealistically high, the amount of risk sharing would increase, and the level of consumption inequality would decrease.

(e) *Initial wealth.* We use evidence on the distribution of wealth among young U.S. workers documented by Díaz-Giménez, Quadrini, and Ríos-Rull (1997). Economies with an average level of initial wealth (at age 23) of zero, but with dispersion equal to what is observed in U.S. data, generate consumption inequality which is qualitatively similar to our benchmark economy. Consumption dispersion is slightly higher for agents aged 23-29, but slightly lower for the remaining age cohorts. The magnitude of these differences is not large and the average level of dispersion is quite similar to our benchmark model.

(f) *Initial distribution of idiosyncratic shocks.* Our econometric methodology and theoretical calculations all assume that there is no initial distribution of persistent idiosyncratic shocks: $z_{i0} = 0$. In terms of earnings, it is conceivable that the variance of a mean zero distribution for these shocks can identified. Given our PSID sample, however, we find doing so problematic (*i.e.*, both the variance of the fixed-effects and the variance of initial distribution of the z's do not appear to be empirically identifiable, something which is primarily due to the linear shape of the earnings profile in Figure 1). In terms of theory, we find the effects on consumption inequality of alternative assumptions about the distribution of z_{i0} to be inconsequential.

(g) *Annuities.* Our model incorporates perfect annuity markets via the term in the budget constraint (8) which reflects the survivor's premium. We find that the elimination of annuities, again holding fixed the level of wealth, makes the consumption profile slightly less concave than our benchmark economy. The size of this effect, however, is quite small and bridges very little of the gap between the concavity in our theory and the linearity in data.

25

6 Conclusions

This paper studies risk-sharing behavior. An inescapable aspect of risk which arises in the labor market is age. Young people face more risk than old people, especially if shocks are persistent. If risk sharing is imperfect, therefore, variation across age should be informative, both for the degree of risk sharing and the technology which gives rise to it. We find that a relatively simple buffer-stock savings model provides a good account of how the age-dependent cross-sectional variance of U.S. income and consumption are related.

Our results depend less on risk sharing technologies — *i.e.*, number of assets, annuities markets, social security — than they do on the characteristics of idiosyncratic risk. Most important is the degree of persistence, something which plays a prominent role in many other studies in macroeconomics and finance. We find that very high persistence, in the neighborhood of a unit root, is necessary to account for how inequality increases with age. Theoretically, this means that models with low persistence generate far more risk sharing than is indicated by data on income and consumption inequality. Empirically, it is a result of emphasizing cross-sectional variances as opposed to the more conventional autocovariances. The former are less susceptible to measurement error than the latter, something which lends credence to our estimates. In addition, when both cross-sectional variances and autocovariances are allowed to weigh-in simultaneously (in the usual GMM sense), the implication of very high persistence remains.

General equilibrium theory plays an important role in the following sense. What ultimately determines the degree of risk sharing in a buffer-stock model is the amount of capital available. If preferences can be chosen arbitrarily, any level of capital can be chosen and any level of risk sharing can be generated. The restriction delivered by equilibrium is that the level of capital should match that of the U.S. economy: 2.9 times the level of output. Given this, the level of risk sharing (*i.e.*, the level of consumption inequality relative to income inequality) is endogenous and provides a useful assessment of the model. We find our results to be robust to different preference parameters in this sense; alternative combinations of risk aversion and time preference which lead to a capital output ratio of 2.9 generate similar risk sharing implications.

From a normative perspective, the most important implication of our paper involves the issue highlighted by Keane and Wolpin (1997). They ask a very important question: what fraction of lifetime uncertainty is resolved before individuals enter the labor market? They reach an equally important, and provocative, conclusion: 90 percent. This, obviously, has profound implications. It suggests a minimal insurance role for financial markets, calls into question most notions of precautionary savings, and suggests that policy aimed at combating inequality should almost exclusively focus (at least indirectly) on school children. We reach a less extreme conclusion, suggesting that at least 40 percent of lifetime uncertainty is resolved over the working

years.[9]

Relative to Keane and Wolpin (1997), our analysis has its limitations. They use a dynamic model of occupational choice which can say a great deal about the ultimate determinants of inequality. Our model is driven by a reduced-form statistical process for earnings, so inequality is not driven by decisions, *per se*. This limitation, however, can also be a strength. Our conclusion rests on one simple observation: without substantial life-cycle shocks, it is difficult to account for the extent to which inequality in earnings *and* consumption increases between age 23 and retirement. What makes this argument particularly convincing is consumption. Should earnings inequality be driven by forces other than shocks — educational and occupational decisions, for instance — consumption inequality will tend to be flat across age due to the standard life-cycle smoothing motive (and ignoring liquidity constraints). Labor market risk, therefore, is necessary to account for increasing consumption inequality with age.

---

[9]Section 4.5 discusses the specifics of what distinguishes our approach and what gives rise to this difference.

References

Aiyagari, S. R., (1994), Uninsured idiosyncratic risk and aggregate saving, *Quarterly Journal of Economics* 109, 659–684.

Altonji, J. G., F. Hayashi, and L. J. Kotlikoff, (1991), Risk sharing, altruism, and the factor structure of consumption, NBER working paper number 3834.

Attanasio, O. and S. J. Davis, (1996), Relative wage movements and the distribution of consumption, *Journal of Political Economy* 104, 1227–1262.

Attanasio, O. and G. Weber, (1992), Consumption growth and excess sensitivity to income: evidence from U.S. micro data, Unpublished manuscript, Stanford University.

Bewley, T. F., (1986), Stationary monetary equilibrium with a continuum of independently fluctuating consumers, in *Contributions to Mathematical Economics in Honor of Gerard Debreu*, edited by Werner Hildenbrand and Andreu Mas-Colell. Amsterdam: North Holland.

Blundell, R. and I. Preston, (1998), Consumption inequality and income uncertainty, *Quarterly Journal of Economics* 113, 603–640.

Carroll, C., (1997), Buffer stock savings and the life cycle and permanent income hypothesis, *Quarterly Journal of Economics* 112, 1–55.

Carroll, C. and A. Samwick, (1998), How important is precautionary savings, *Review of Economics and Statistics* 80, 410–419.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull, (1998), Earnings and wealth inequality and income taxation: Quantifying the trade-offs of switching the U.S. to a proportional income tax system, Mimeo, University of Pennsylvania.

Chatterjee, S., (1994), Transitional dynamics and the distribution of wealth in a neoclassical growth model, *Journal of Public Economics* 54, 97–119.

Cochrane, J. H., (1991), A simple test of consumption insurance, *Journal of Political Economy* 99, 957–76.

Cooley, T. F. and E. C. Prescott, (1995), Economic growth and business cycles, In *Frontiers of Business Cycle Research*, edited by Thomas F. Cooley. Princeton University Press: New Jersey.

Cutler, D. M. and L. F. Katz, (1992), Rising inequality? Changes in the distribution of income and consumption in the 1980's, *American Economic Review* 82, 546–551.

Deaton, A., (1992), *Understanding Consumption*, Oxford University Press, Oxford.

Deaton, A., P.-O. Gourinchas, and C. Paxson, (2000), *Social security and inequality over the life cycle*, in *The Distributional Effects of Social Security Reform*, edited by Martin Feldstein and Jeffrey Liebman. Chicago: University of Chicago Press.

Deaton, A. and C. Paxson, (1994), Intertepmoral choice and inequality, *Journal of Political Economy* 102, 437–467.

Díaz-Giménez, J., V. Quadrini, and J. V. Ríos-Rull, (1997), Dimensions of inequality: facts on the U.S. distribution of earnings, income and wealth, *Federal Reserve Bank of Minneapolis Quarterly Review* 21, 3–21.

Hubbard, R. G., J. Skinner, and S. P. Zeldes, (1994), The importance of precautionary motives in explaining individual and aggregate savings, *Carnegie Rochester Conferance Series on Public Policy* 40, 59–126.

Huggett, M., (1993), The risk free rate in heterogeneous-agents, incomplete insurance economies, *Journal of Economic Dynamics and Control* 17, 953–969.

Huggett, M., (1996), Wealth distribution in life-cycle economies, *Journal of Monetary Economics* 38, 469–494.

İmrohoroğlu, A., S. İmrohoroğlu, and D. H. Joines, (1995), A life cycle analysis of social security, *Economic Theory* 6, 83–114.

Keane, M. P. and K. I. Wolpin, (1997), The career decisions of young men, *Journal of Political Economy* 105, 473–522.

Kimball, M. S., (1990), Precautionary saving in the small and in the large, *Econometrica* 58, 53–73.

Krueger, D. and F. Perri, (2000), Does income inequality lead to consumption inequality? Empirical findings and a theoretical explanation, Manuscript, Stanford University.

Krusell, P. and A. A. Smith, (1998), Income and wealth heterogeneity in the macroeconomy, *Journal of Political Economy* 106, 867–896.

Mace, B. J., (1991), Full insurance in the presence of aggregate uncertainty, *Journal of Political Economy* 99, 928–56.

Quadrini, V., (2000), Entrepreneurship, saving, and social mobility, *Review of Economic Dynamics* 3, 1–40.

Ríos-Rull, J. V., (1994), On the quantitative importance of market completeness, *Journal of Monetary Economics* 34, 463–496.

Slesnick, D. T., (1993), Gaining ground: Poverty in the postwar united states, *Journal of Political Economy* 101, 1–38.

Smith, A. A. J. and C. Wang, (2000), Dynamic credit relationships in general equilibrium, Unpublished manuscript, Carnegie Mellon University.

Storesletten, K., C. I. Telmer, and A. Yaron, (1999a), Asset pricing with idiosyncratic risk and overlapping generations, GSIA working paper no. 1998-E226, Carnegie Mellon University.

Storesletten, K., C. I. Telmer, and A. Yaron, (1999b), The risk sharing implications of alternative social security arrangements, *Carnegie Rochester Conferance Series on Public Policy* 50, 213–259.

Storesletten, K., C. I. Telmer, and A. Yaron, (2000), Accounting for idiosyncratic risks over the life cycle: Theory and evidence, Unpublished manuscript, Carnegie Mellon University.

Zeldes, S. P., (1989), Optimal consumption with stochastic income: Deviations from certainty equivalence, *Quarterly Journal of Economics* 104, 275–298.

Appendix A
Panel Data and Estimation

Our data source is the family files and the individual files of the Panel Study on Income Dynamics (PSID), covering the years 1969-1992. Since each PSID cross section covers income earned the previous year, we refer to the time dimension as being 1968-1991. We base our analysis on a sequence of 22 overlapping panels, each of which has a time dimension of 3 years. For instance, the first panel, which we refer to as having a 'base year' of 1968, consists of earnings data from the years 1968, 1969, and 1970. The panel with a base year of 1969 contains data from 1969, 1970 and 1971. These overlapping panels allow for the identification of our model's time series parameters while at the same time maintaining a broad cross section (due to the introduction of new households) and a stable age distribution. In addition, our statistical methods explicitly incorporate the overlapping nature of the panels into our estimate of the covariance matrix.

We define a household's total earnings as wage earnings plus transfers. Wage earnings are defined as the sum of the wage earnings of the household head plus those of their spouse. 'Transfers' include a long list of variables defined by the PSID (the 1968 variable name, for instance, is V1220), but the lion's share is attributable to unemployment insurance, workers compensation, and transfers from non-household family members. Total earnings are converted to real earnings per household member by using the CPI deflator and by dividing by the number of household members.

Given a specific base year, a household is selected into the associated panel if the following conditions are met for the base year and each of the two subsequent years:

- Total earnings are positive in each year.

- Total earning growth rates are no larger than 20 and no less than 1/20 in any consecutive years.

In addition, we follow standard practice in excluding households which were originally included in the Survey of Economic Opportunity. This selection criteria is quite broad relative to previous work, including our own (Storesletten, Telmer, and Yaron (1999a)). Additional restrictions are typically imposed that the household head is male and that there is no change in family structure over time. We choose not to impose these restrictions because (a) we don't need to: our focus on cross-sectional moments greatly mitigates the problems associated with estimating a dynamic model with data in which household structure changes, and (b) we don't want to: by doing so we may eliminate an important component of the idiosyncratic variation we are trying to measure. In addition, this broad selection criteria is chosen to be consistent with the consumption data we use, taken from Deaton and Paxson (1994).

Much of our analysis is based on the age profile of the cross-sectional variance of the logarithm of earnings: see Figure 1 for instance. These cross-sectional variances have been computed so as to remove "cohort effects:" variation which is specific to a cohort of individuals born during the same year. For our graphical analysis, we use the same method as Deaton and Paxson (1994). We remove cohort and age means and then regress the residuals on age and cohort dummy variables. The graphs report the coefficients on the age dummies, rescaled so as to match the average level of dispersion in the raw data. For our GMM analysis, we take a different approach, using the moments underlying the OLS estimates to formulate a lower-dimensional model of the age effects.

A.1 Reconciliation with Consumer Expenditure Survey (CEX) Data

Our paper is based on earnings data from the PSID and consumption data from the CEX, the idea being that these are the best data sources. The cost, of course, is consistency. While it is not possible to compare consumption across both sources — the PSID contains only data on food consumption — we can check how labor earnings measure up. We find that average income for various age groups in the CEX is very similar to our PSID sample. For example, the mean income of households in the 1997 CEX survey is \$40,247, \$48,788 and \$55,260 for ages 30,40,50, respectively. The corresponding means in the PSID, given our selection criteria, are essentially identical. More importantly, we find that, once one uses our relatively broad PSID selection criteria, cross sectional variances are quite similar across sources. This can be informally verified by comparing Figure 1 to the associated graphs in Deaton and Paxson (1994).

A.2 Parametric Model and GMM Estimation

We associate an individual household with the age of the household head, denoted h, and the *cohort* (*i.e.*, the birth year) to which the household belongs, denoted c. There are H ages and C cohorts. The logarithm of demeaned earnings for the ith household of age h belonging to cohort c is modeled as

$$y_{ih}^c \quad = \quad u_{ih} + x_{ic}$$

where x_{ic} and u_{ih} represent mean zero 'cohort' and 'age' shocks respectively.

Deaton and Paxson (1994) use a dummy variable regression to recover the variances of u_{ih} and x_{ic}, which we denote b_h and a_c. They then scale the age effects so that, on average, the coefficients match the unconditional variance of some reference age group (we use age 42). We denote the scaled age effects as b_h^*. The solution to

the following system of moment equations yields the OLS estimates from this dummy variable regression:

$$
\begin{aligned}
E[(y_{ih}^c)^2|h] - \bar{a} - b_h &= 0, & \text{H such moments} \\
E[(y_{ih}^c)^2|c] - \bar{b} - a_c &= 0 & \text{C such moments} \\
b_h + (m - b_{42}) - b_h^* &= 0 & \text{(A1)} \\
E[(y_{ih}^c)^2 \,|\, h = 42] - m &= 0
\end{aligned}
$$

The first two equations are the OLS moments while the second two represent the scaling.

As discussed in the text, our approach consists of putting structure on u_{ih} and, therefore, on b_h^*. Specifically, we let

$$
\begin{aligned}
u_{ih} &= \alpha_i + z_{ih} + \varepsilon_{ih} \\
z_{ih} &= \rho z_{i,h-1} + \eta_{ih} \ ,
\end{aligned}
$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\varepsilon_{ih} \sim N(0, \sigma_\varepsilon^2)$, $\eta_{ih} \sim N(0, \sigma_\eta^2)$ and $z_{i0} = 0$. This implies that the age effects have the following form

$$
b_h^* = Var(u_{ih}) = \sigma_\alpha^2 + \sigma_\eta^2 \sum_{j=0}^{h-1} \rho^{2j} + \sigma_\varepsilon^2 \ . \tag{A2}
$$

Our estimator is based on an amalgamation of equations (A1) and (A2). This system, however, does not separately identify σ_α^2 and σ_ε^2. We therefore formulate additional moment equations based upon summations of y_{ih}^c:

$$
s_{ih}^c = y_{ih}^c + y_{i,h+1}^c + y_{i,h+2}^c \ .
$$

Variances of s_{ih}^c identify σ_α^2 and σ_ε^2 because they involve autocovariances. We prefer these variances to individual autocovariances because (a) they incorporate information from multiple autocovariances in an economical way (this is helpful for the exactly-identified system described below), and (b) they are less susceptible to $i.i.d.$ measurement error.

The variances of s_{ih}^c are

$$
Var(s_{ih}^c) = 9\sigma_\alpha^2 + 3\sigma_\varepsilon^2 + \left((1+\rho)^2 + 1\right) \sigma_\eta^2 + \left(1 + \rho + \rho^2\right)^2 Var(z_{ih}) + 9a_c \ , \tag{A3}
$$

where $Var(z_{ih}) = \sigma_\eta^2 \sum_{j=0}^{h-1} \rho^{2j}$. These moments are cohort specific. We eliminate cohort specificity by scaling the s_{ih}^c data as follows:

$$
s_{ih} = \frac{s_{ih}^c}{Var(s_{ih}^c)^{1/2}} \left(9\sigma_\alpha^2 + 3\sigma_\varepsilon^2 + \left((1+\rho)^2 + 1\right)\sigma_\eta^2 + \left(1 + \rho + \rho^2\right)^2 Var(z_{ih})\right)^{1/2} \ ,
$$

33

where the term in the denominator, $Var(s_{ih}^c)^{1/2}$, is the expression on the right hand side of equation (A3). The variance of s_{ih} is

$$Var(s_{ih}) = 9\sigma_\alpha^2 + 3\sigma_\varepsilon^2 + \left((1+\rho)^2 + 1\right)\sigma_\eta^2 + \left(1 + \rho + \rho^2\right)^2 Var(z_{ih}) \qquad (A4)$$

which is now absent cohort effects. Equations (A4) therefore represent $H-2$ moments which we use in varying degrees in the discussion below. Initially, we treat the a_c's as given when constructing the s_{ih}'s (incorporating, however, the sampling variability in our standard errors), but subsequently we estimate them jointly with the other parameters.

The moment equations (A1), (A2) and (A4) form the foundation of our estimators. An obvious place to begin might be to substitute equations (A2) into equations (A1) and then add K equations of the form (A4). This would give an overidentified system with $H + C + K$ equations and $C + 4$ parameters. As is discussed in the text, however, we find it informative to begin with a much smaller, exactly-identified, system and then work our way up to the larger system.

Panel B of Table 1

These estimates represent an exactly identified system. Taking as given the estimates of the a_c's from (A1), we choose three ages from equations (A1) — ages 23, 42 and 60 — and one age from equations (A4): age 42. The estimates of ρ, σ_η, σ_ϵ, and σ_α are simply a solution to the non-linear set of equations which result from equating the three b_h^* parameters to the corresponding expressions from equations ((A2) and then incorporating the one equation from (A4). The standard errors are corrected for the estimation error embedded in the estimates of b_h^* and a_c and are computed using the delta method.

We repeated this estimation procedure for moments based on different age-sets. That is, we maintained $h = 42$ for equation (A4) while using $h = 30, 45, 60$ and $h = 35, 45, 55$ in equation (A2). The results were essentially unaltered. We also tried $h = 30, 40$ and 50 in equation (A4). Again, the results were very similar to those displayed in Table 1.

Panel C of Table 1

Panel C moves to an overidentified GMM system by incorporating all ages from equations (A4). That is, we incorporate information from many more age-dependent autocovariances. We continue to use only ages $h = 23, 42, 60$ from equations (A2). As discussed in the text, the additional autocovariance moments reduce the estimated ρ, but only marginally. This result is robust to changing the age set used in equation (A2).

34

Panel D of Table 1

Panels A through C treat the a_c parameters as fixed. In panel D, we jointly estimate the a_c's in both (A1) and (A3-A4). The procedure is in essence similar to the one used for Panels B and C except that the a_c's are iterated upon jointly with $\rho, \sigma_\eta, \sigma_\alpha, \sigma_\epsilon$ using equations (A1)-(A4). The starting values are the a_c values of the OLS estimation in Panel B. The cohort effects a_c do not change dramatically but do result in a slightly lower estimate of ρ.

Additional Evidence

1. Estimates using moments based directly on individual autocovariances instead of the summation moments (A4) — perhaps a more conventional approach — yield slightly lower estimates of ρ, between 0.95 and 0.97.

2. Year, age and cohort dummies cannot be simultaneously estimated as they are linear combinations of one another. To check whether our results are sensitive to the use of cohort dummies, we estimated the age effects in conjunction with year dummies instead of cohort dummies. The pattern in the cross-sectional variances depicted in Figure 1 is quite robust along these lines (Deaton and Paxson (1994) reach a similar conclusion).

3. We estimated the earnings process separately for different educational groups: no high school diploma, high school graduates, and college graduates. The estimates for all these education groups are quite similar to those presented in Table 1. For example, for high school graduates we estimated ρ to be 0.974 and 0.968 when cohort effects were set at their OLS estimates and were jointly estimated, respectively. Our estimates of persistence using pooled data (across educational groups) are therefore robust in this sense. Similar findings apply to the other parameters.

Appendix B
Welfare Calculations

Details regarding the welfare gain of removing fixed effects are as follows. We want to answer the question "how much consumption is the agent willing to give up to get rid of the fixed effects?" We claim that this amount is $\psi_\alpha = 1 - \exp\left(-\gamma \sigma_\alpha^2\right)$.

To see this, start with the following expression which implicitly defines ψ_α:

$$\int \int \sum_{h=1}^{H} \beta^h \phi_h \; U\left(c_h\left(\alpha, z_h, \varepsilon_h, a_h\right)\right) d\mu = \int_{I\{\alpha=1\}} \sum_{h=1}^{H} \beta^h \phi_h \; U\left((1-\psi_\alpha)\, c_h\left(1, z_h, \varepsilon_h, a_h\right)\right) d\mu \tag{B5}$$

Since preferences are homothetic and fixed effects simply scale earnings and borrowing constraints proportionally in each state, each individual's consumption must be proportional to their fixed effect. Thus, $c_h\left(\alpha, z, \varepsilon, a\right) = \alpha c_h\left(1, z, \varepsilon, \frac{a}{\alpha}\right)$, so (B5) can be rewritten as

$$\int \alpha^{1-\gamma} \int_{I\{\hat{\alpha}=\alpha\}} \sum_{h=1}^{H} \beta^h \phi_h \; U(c_h(\hat{\alpha}, z_h, \varepsilon_h, \frac{a_h}{\hat{\alpha}})) d\mu = (1-\psi_\alpha)^{1-\gamma} \int_{I\{\alpha=1\}} \sum_{h=1}^{H} \beta^h \phi_h \; U(c_h(1, z_h, \varepsilon_h, a_h)) d\mu$$

$$\int \alpha^{1-\gamma} V_1(1, z, \varepsilon, 0) d\alpha = (1-\psi_\alpha)^{1-\gamma} V_1(1, z, \varepsilon, 0)$$

$$(1-\psi_\alpha)^{1-\gamma} = \int e^{(1-\gamma)\log\alpha} d\alpha = \exp\left(-(1-\gamma)\sigma_\alpha^2 \gamma\right)$$

$$\psi_\alpha = 1 - \exp\left(-\gamma \sigma_\alpha^2\right)$$

QED

Now, let us compute the welfare gain of removing life-cycle shocks, i.e. to obtain an answer the question "how much consumption is the agent willing to give up to get rid of the life-cycle shocks?" This amounts to solving for ψ_l implicitly given by

$$\int \int \int V_1\left(\alpha, z, \varepsilon, 0\right) \mu\left(\alpha, z, \varepsilon\right) d\alpha dz d\varepsilon = (1-\psi_l)^{1-\gamma} \int \hat{V}_1\left(\alpha, 0\right) d\alpha$$

where $\hat{V}_1\left(\alpha, 0\right)$ is the value function of a newborn agent in who has fixed-effect α and zero wealth in an economy with $\sigma_\eta = \sigma_\varepsilon = 0$. The term $(1-\psi_l)$ is raised to the power $1-\gamma$ and premultiplying the value function since the current utility is CRRA (see above). This means that the welfare gain ψ_l can be expressed by the following expression (which must be computed numerically by simulation)

$$\psi_l = 1 - \left(\frac{EV_1\left(\alpha, z, \varepsilon, 0\right)}{E\hat{V}_1\left(\alpha, 0\right)}\right)^{\frac{1}{1-\gamma}}.$$

Table 1
Idiosyncratic Endowment Process: Parameter Estimates

	ρ	σ_α^2	σ_η^2	σ_ε^2
A. Informal Estimates, Based on Figure 2[*]				
Estimate	0.9989	0.2735	0.0166	n.a.
Standard Error				
B. Exactly Identified System				
Estimate	0.984	0.242	0.022	0.057
Standard Error	0.009	0.016	0.007	0.028
C. Overidentified System, Predetermined Cohort Effects				
Estimate	0.982	0.247	0.024	0.061
Standard Error	0.011	0.013	0.006	0.022
D. Overidentified System, Simultaneously Estimated Cohort Effects				
Estimate	0.977	0.244	0.024	0.063
Standard Error	0.018	0.021	0.010	0.029

[*]The parameters σ_α and σ_ε are not separately identified in this case.

Entries describe GMM estimates for the idiosyncratic endowment process described in the text:

$$u_{ih} = \alpha_i + z_{ih} + \varepsilon_{ih}$$
$$z_{ih} = \rho z_{i,h-1} + \eta_{ih} \;,$$

where $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\varepsilon_{ih} \sim N(0, \sigma_\varepsilon^2)$, $\eta_{ih} \sim N(0, \sigma_\eta^2)$ and $z_{ic}^0 = 0$. Estimates in Panel A are based on the level, slope and curvature of the age profile of cross-sectional variances shown in Figure 1. In Panel B, one additional moment is added vis-a-vis those implicit in Panel A: the variance of the sum, $u_{ih} + u_{i,h+1} + u_{i,h+2}$. The estimates are based on one specific age, $h = 42$, but alternatives are discussed in Appendix A. Panel C represents the same approach as Panel B, however one moment is added for each age cohort. In Panel D we estimate 'cohort effects' simultaneously with the four parameters of interest, whereas in Panels A–C cohort effects are pre-estimated and removed from our data through cohort-specific scaling.

37

Figure 1
Cross-sectional Variance by Age
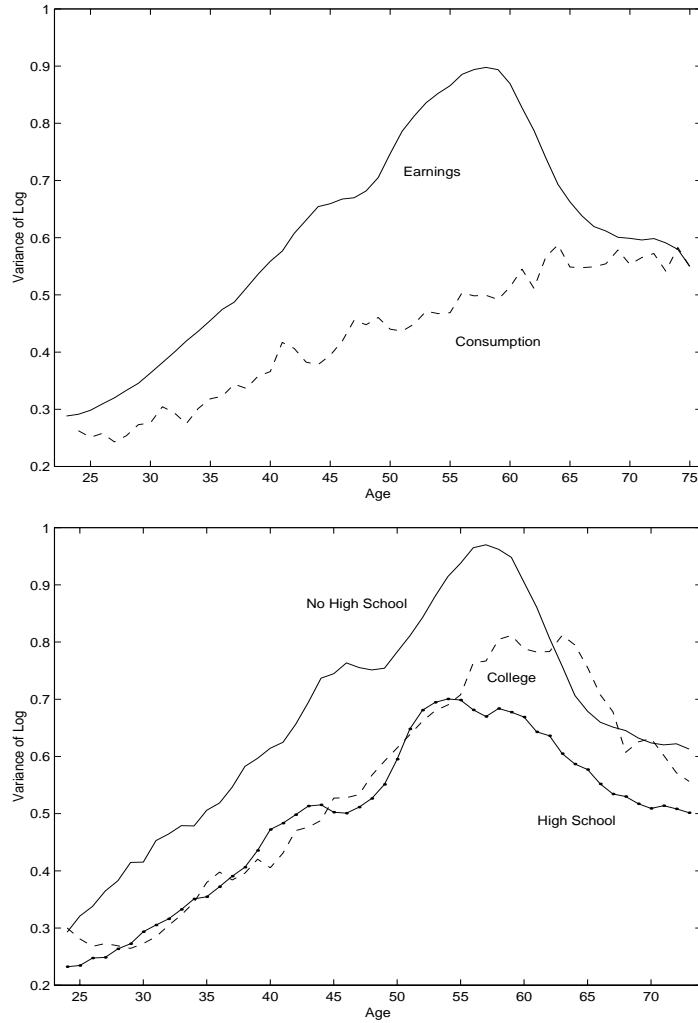
Each line represents the age-dependent, cross-sectional variance of the logarithm of the quantity in question. The basic data unit is the household. Consumption data are from the Consumer Expenditure Survey (CEX) and are taken directly from Deaton and Paxson (1994). Earnings data are taken from the Panel Study on Income Dynamics (PSID). Exact details are in Appendix A. The lower panel reports earnings dispersion broken down into three educational cohorts whereas the upper panel aggregates educational cohorts. All graphs represent variances which are net of 'cohort effects:' dispersion which is unique to a group of households with heads born in the same year. This is accomplished, as in Deaton and Paxson (1994), via a cohort and age dummy variable regression. The graphs are the coefficients on the age dummies, scaled so as to mimic the overall level of dispersion in the data.

Figure 2
## Calibration of Earnings Process
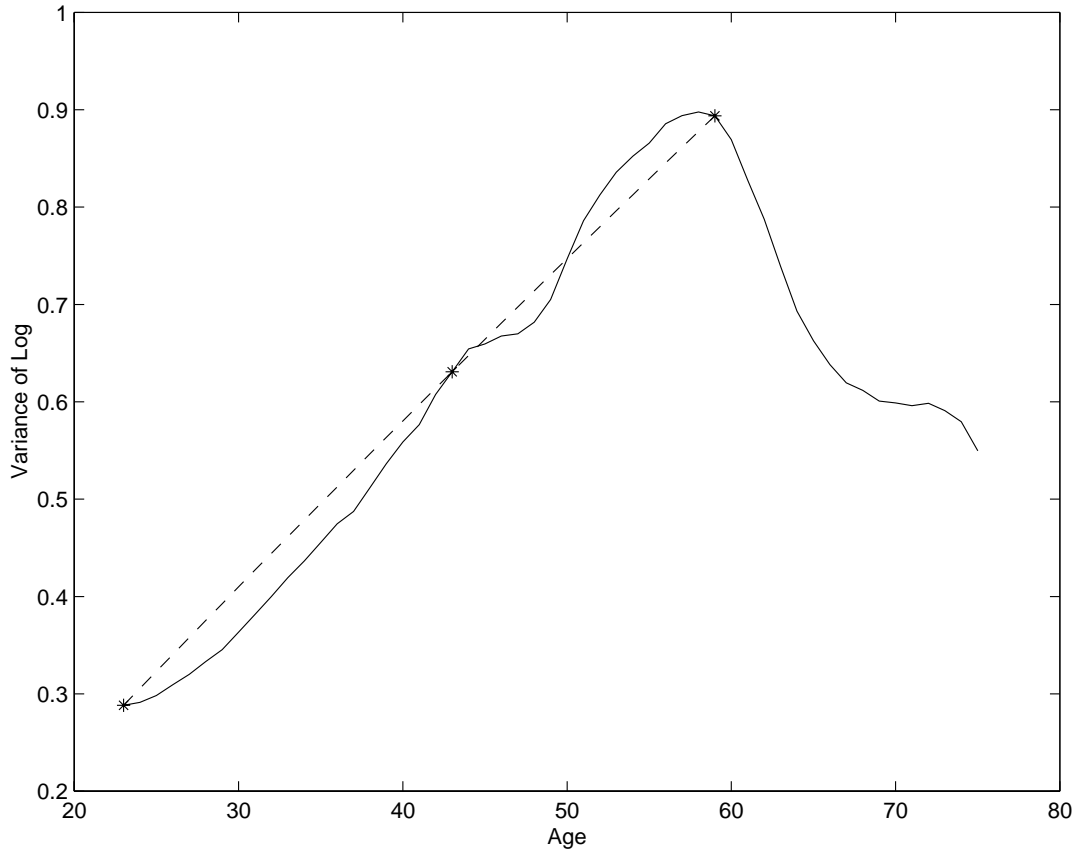
The solid line is the cross-sectional variance of earnings, based upon PSID data, and is taken directly from Figure 1. The dashed line represents the population cross-sectional variances associated with the process formulated in Section 2.1, with parameter values chosen to best match the level (*i.e.*, the intercept), the slope and the curvature of the empirical age profile. The resulting parameter values (discussed in Section 2.1) are $\sigma_\alpha^2 + \sigma_\varepsilon^2 = 0.2735$, $\sigma_\eta^2 = 0.0166$ and $\rho = 0.9989$. The asterisks represent the three sample variances which this informal calibration explicitly matches.

Figure 3
## Cross-sectional Variance: Theory and Data
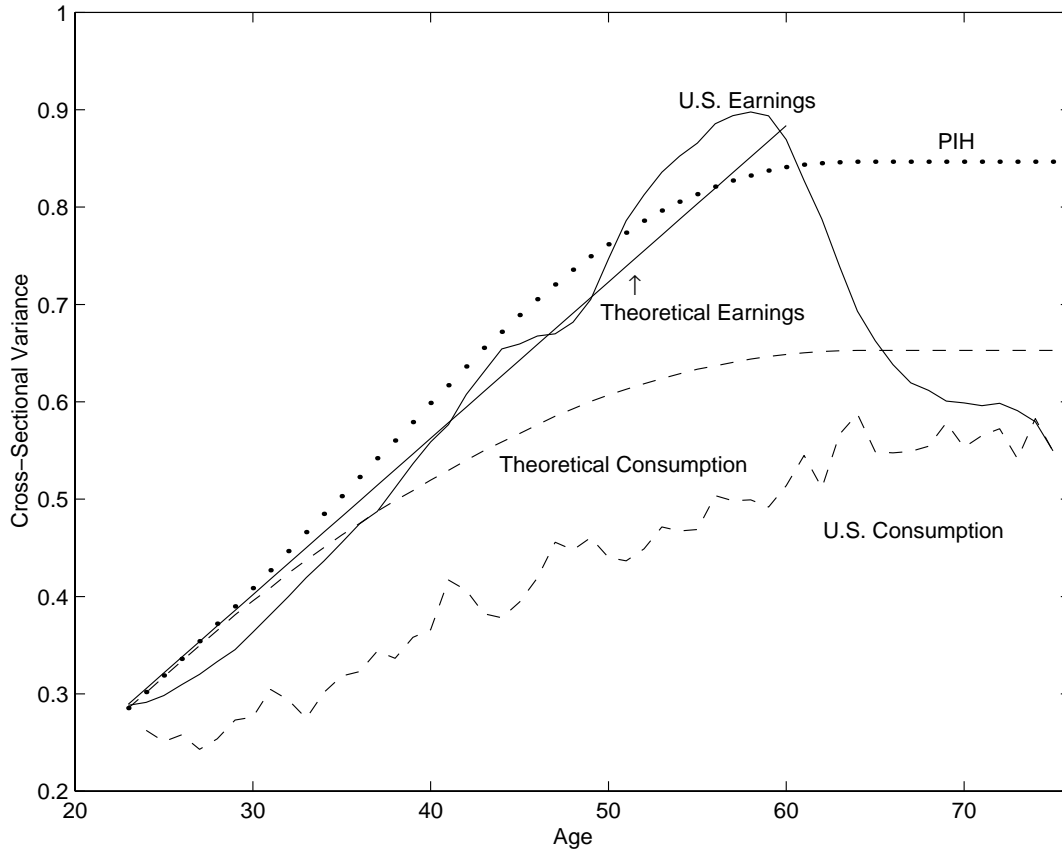
This graph compares the population moments from our benchmark model with those of the data. Solid lines represent theoretical and empirical earnings. Dashed lines represent consumption. The dotted line represents the cross-sectional variance of consumption generated by the life-cycle version of the permanent income hypothesis, discussed in Section 4.2.

Figure 4
Effect of Shape of Earnings Profile on Shape of Consumption Profile
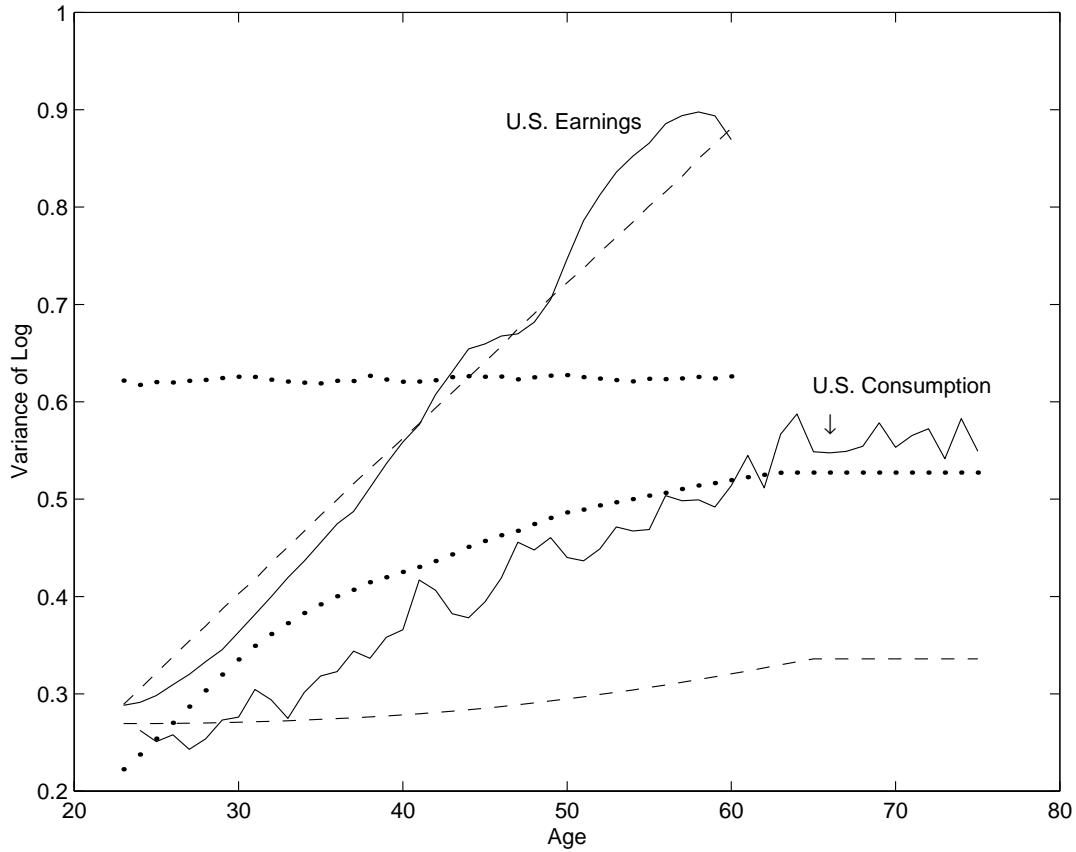

This graph supports the first statement in Section 4.3; a realistically-shaped profile of earnings inequality is neither necessary nor sufficient to generate a realistically shaped profile of consumption inequality. Solid lines represent sample moments. The dotted lines show that an unrealistic pattern of earnings dispersion (the uppermost dotted line) can be associated with a realistic pattern of consumption dispersion (the lower dotted line). The dashed lines show the opposite, that a realistic pattern of earnings dispersion (the uppermost dashed line) can generate a counterfactual pattern of consumption dispersion (the lower dashed line). The critical ingredient is persistence. The dotted lines are associated with a high-persistence earnings process, but with an initial distribution equal to the stationary distribution. The dashed lines represent an *i.i.d.* earnings processes, but with volatility which increases with age.

Figure 5
## Effect of Persistence on Consumption Inequality
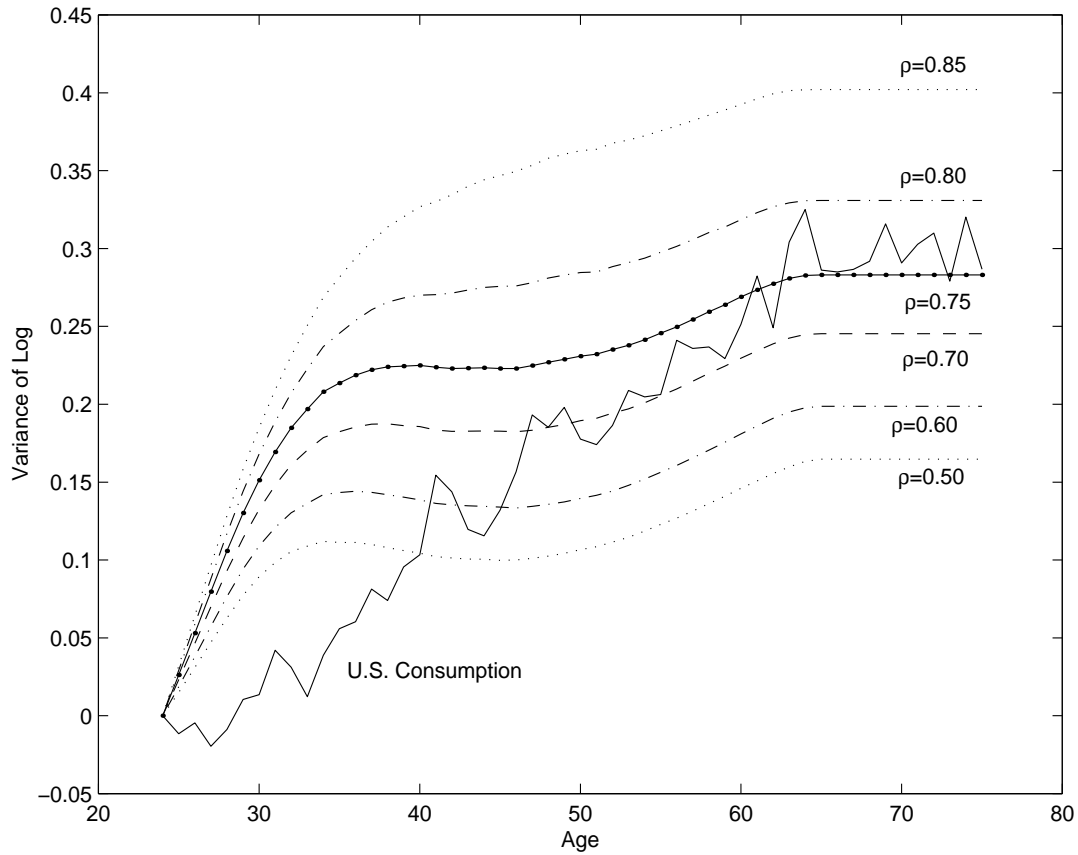
This graph provides examples which substantiate point B of section 4.3: highly autocorrelated shocks are necessary to account for the increase in U.S. consumption dispersion. The solid line represents sample moments and the others are from economies with varying degrees of persistence. In order to focus on the increase, the loci are all indexed to unity at age 23. Each line represents the answer to the question, "what is the maximum increase in consumption dispersion that can be generated by shocks of a given level of persistence?" In order to generate the maximum, we set the fixed-effect shocks and the transitory shocks to zero. The conditional variance of the persistent shocks is then chosen to match the average level of earnings dispersion in the data (*i.e.*, the average of the solid line in Figure 3).

Figure 6
Effect of Persistence in Conjunction with Heteroskedastic Shocks
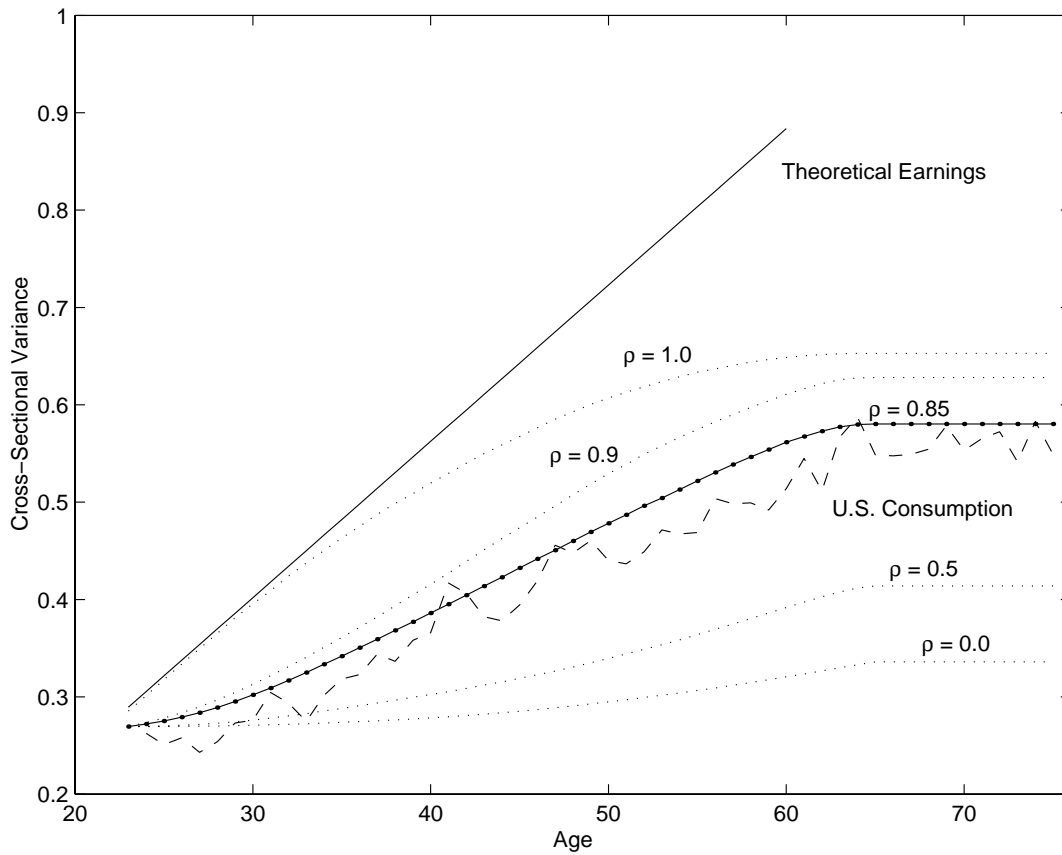
The dashed line represents sample moments on consumption dispersion. Each of the dotted lines represent consumption dispersion from a theoretical economy with persistence as noted in the graph, and with age-dependent heteroskedasticity in the conditional variance chosen so that earnings inequality matches the solid line (which also matches that in our benchmark economy). All the economies, therefore, have an identical pattern of earnings inequality, but different patterns in consumption inequality. The economy with $\rho = 0.85$ most closely matches the data. The heteroskedastic pattern which gives rise to this features older workers experiencing idiosyncratic shocks which are more volatile than those faced by younger workers. See Section 4.4 for a discussion.

Figure 7
Effect of Lower Persistence
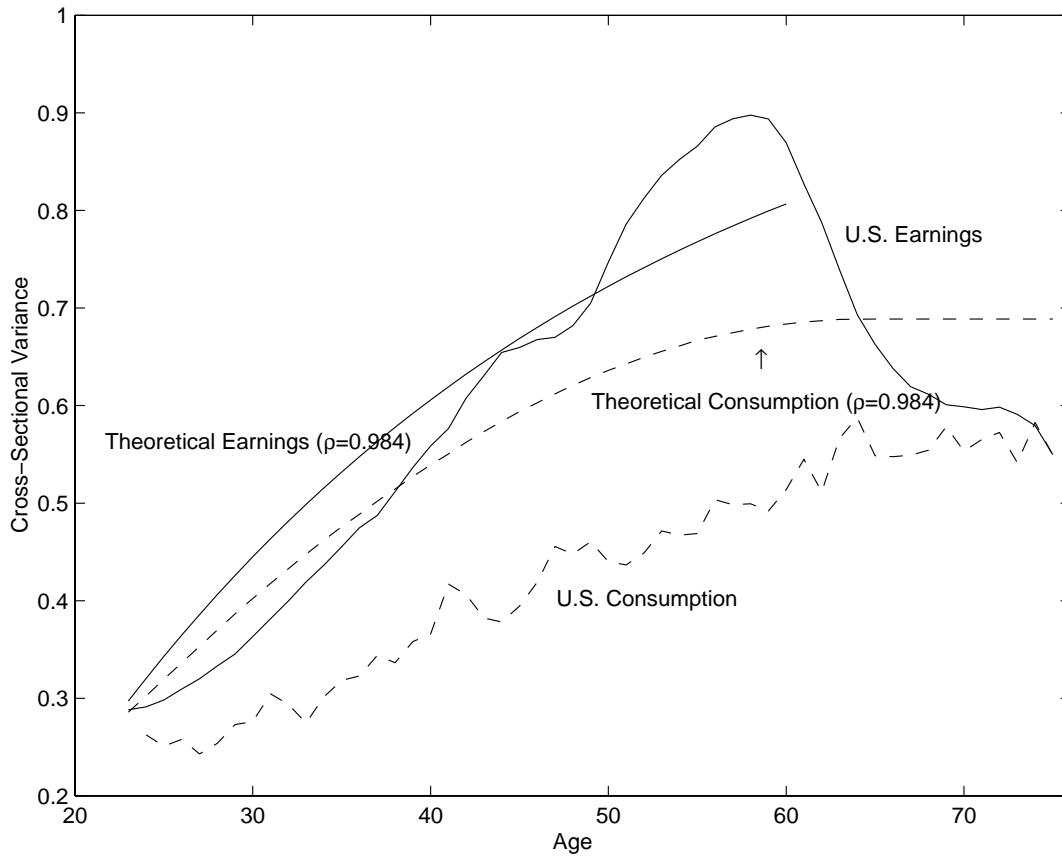
This graph mimics Figure 3, but represents shocks with an autocorrelation of 0.984, chosen to be representative of the estimates in Table 1. See point (a), Section 5 for a discussion.