Asset Pricing with Idiosyncratic Risk and Overlapping Generations*

Kjetil Storesletten,† Chris Telmer,‡ and Amir Yaron§

July 1999
(first draft: June 1996)

Abstract

A number of existing studies have concluded that risk sharing allocations supported by competitive, incomplete markets equilibria are quantitatively close to first-best. Equilibrium asset prices in these models have been difficult to distinguish from those associated with a complete markets model, the counterfactual features of which have been widely documented. This paper asks if life cycle considerations, in conjunction with persistent idiosyncratic shocks which become more volatile during aggregate downturns, can reconcile the quantitative properties of the competitive asset pricing framework with those of observed asset returns. We begin by arguing that data from the Panel Study on Income Dynamics support the plausibility of such a shock process. Our estimates suggest a high degree of persistence as well as a substantial increase in idiosyncratic conditional volatility coincident with periods of low growth in U.S. GNP. When these factors are incorporated in a stationary overlapping generations framework, the implications for the returns on risky assets are substantial. Plausible parameterizations of our economy are able to generate Sharpe ratios which match those observed in U.S. data. Our economy cannot, however, account for the level of variability of stock returns, owing in large part to the specification of its production technology.

*A predecessor of this paper circulated as "Persistent Idiosyncratic Shocks and Incomplete Markets," first presented at the NBER summer institute, July 1996. In addition to numerous participants at seminars and conferences, we thank Dave Backus, Geert Bekaert, David Bowman, Martin Browning, Wouter den Haan, Mick Devereux, Ron Gallant, René Garcia, John Heaton, Mark Huggett, Narayana Kocherlakota, Per Krusell, Deborah Lucas, Pierre Mella-Barral, Bob Miller, Michael Parkin, Victor Ríos-Rull, Amlan Roy, Tony Smith, Paul Willen, Steve Zeldes and Stan Zin for helpful comments and suggestions. Valuable research assistance was provided by Xiaorong Dong.

†Institute for International Economic Studies, Stockholm University; storeslk@iies.su.se
‡Graduate School of Industrial Administration, Carnegie Mellon University; telmerc@andrew.cmu.edu
§Wharton School, University of Pennsylvania; yarona@savage.wharton.upenn.edu

1 Introduction

A number of recent papers have focused on the extent to which departures from the classical Arrow-Debreu model of securities markets can help reconcile intertemporal asset pricing theory with historical observations on the joint behavior of U.S. consumption and asset returns. One popular approach has been to posit a model in which incomplete markets inhibit the complete pooling of agent-specific income risks. For example, papers by Aiyagari (1994), Aiyagari and Gertler (1991), Alvarez and Jermann (1997), den Haan (1994), Heaton and Lucas (1996), Huggett (1993), Lucas (1994), Mankiw (1986), Marcet and Singleton (1991), Ríos-Rull (1994), Krusell and Smith (1997), Telmer (1993), Weil (1992), and Zhang (1997) examine the extent to which agents can hedge transitory income shocks by trading in competitive markets for a very limited number of financial assets. By and large, these studies have concluded that the incomplete markets allocations, and therefore the supporting price processes, are fairly close to those which would obtain under complete markets. Quantitatively, therefore, these models inherit many of the well known drawbacks associated with the Mehra and Prescott (1985) representative agent economy. The imposition of frictions above and beyond market incompleteness, borrowing constraints and transactions costs for instance, can obviously overturn these results for some specifications of the frictions. However, in many cases the degree of restrictiveness required to achieve a substantial deviation from first-best risk sharing has seemed unrealistic.

The underlying theme of most of the above papers is the notion that idiosyncratic variation has aggregate consequences, both in terms of quantities and prices. Our paper continues along this line of inquiry, motivated by a large body of evidence suggesting that actual households bear substantial amounts of idiosyncratic risk. Our working hypothesis is that this risk is relevant for individual savings and portfolio decisions, and that these decisions are important determinants of asset prices, at least when measured at a suitably low frequency. The question, then, is what is missing in existing models, which find the aggregate implications of idiosyncratic variation to be quantitatively small?

Our interpretation of what is missing is based on two simple observations, each of which is formalized in the framework of Constantinides and Duffie (1996). First, a successful model must feature allocations which represent a substantial departure from full risk sharing. Because many existing studies have not had this property, we are led to question the specification of the processes for idiosyncratic risk and/or the set of frictions which inhibit societal risk sharing. Second, for models in this class to have interesting asset pricing properties, there must exist some sort of dependence between idiosyncratic and aggregate sources of uncertainty. The reason is simple. The 'equity premium puzzle' of Mehra and Prescott (1985) is essentially a quantitative statement about the relationship between asset prices and the consumption allocations of a complete markets model. If the only difference between incomplete and complete market allocations is *i.i.d.*, idiosyncratic variation, then the impact on asset prices — which are driven by covariation — is likely to be minimal. We are therefore led to carefully examine the aggregate-idiosyncratic relationship underlying previous work.

The central theme of our paper is that life cycle effects are important for addressing each of these issues. In regard to the first — risk sharing — we interpret previous work as suggesting that the infinite horizon abstraction endows theoretical agents with far greater risk sharing possibilities than those seemingly faced by actual agents. In other words, because they have a long horizon over which to implement it, a simple strategy of contingent borrowing and lending allows an infinitely lived agent to eliminate most of the variation attributable to idiosyncratic shocks, even those which are quite persistent. Constantinides and Duffie (1996) demonstrate that this can be overcome by modeling idiosyncratic shocks as being permanent (*i.e.*, unit root processes), but at the cost of a nonstationary cross sectional distribution or the assumption of dynasties with probabilistic death. The attractiveness of the life cycle model, in this context, is that it replaces infinitely lived shocks with finitely lived agents and can, conceivably, achieve the same end in a stationary, recursive framework. Moreover, much of the baggage which accompanies a life cycle model plays a welcome role in constraining risk sharing further, and in a manner which seems quite natural. There is a tension, for instance, between life cycle and precautionary savings motives, one which can limit the extent to which a given shock can be offset by borrowing or, equivalently, dissaving. There is also an important role played by the life cycle distribution of wealth — the extent to which the young are born with relatively few assets — in that a strategy of contingent savings and dissavings is difficult to implement if one does not possess a buffer stock of assets to begin with. Portfolio constraints turn out to play an important role along this dimension.

Life cycle effects can also play an important role in what is critical for asset pricing: the interaction between aggregate and idiosyncratic shocks. A hallmark of our approach is that idiosyncratic risk arises in the labor market. An important implication is that, necessarily, there exists a life cycle pattern in the distribution of idiosyncratic risk across age cohorts: young agents face more than older agents. The lynchpin for asset pricing, in our model, is that this life cycle pattern in idiosyncratic risk manifests itself as a life cycle pattern in portfolio choice — younger agents hold far less of the risky asset than older agents — which in turn results in a subset of agents being marginal in the sense of price determination. If idiosyncratic risk causes these agents to dislike risky assets, then a relatively high equity premium will result, higher than if equity ownership were more evenly distributed. The issue, then, is the manner in which idiosyncratic risk is related to asset returns. We follow Constantinides and Duffie (1996) and Mankiw (1986) in formulating this relation as an increase in idiosyncratic risk during aggregate downturns, something which we label 'countercyclical cross sectional variation' (CCV). What is novel in our setting is a life cycle effect: an interaction between asset prices and the (endogenous) distribution of this CCV risk over the generations of agents who populate our model. We find the quantitative implications of this effect to be important.

A more specific description of the distinguishing characteristics of our model is as follows. We study a recursive life cycle model populated by 78 generations of agents, where each generation consists of many heterogeneous individuals. Roughly two thirds of a given agent's life is spent working, with the remainder spent in retirement. Each agent faces both idiosyncratic and aggregate shocks, the latter being attributable to an aggregate production technology. Idiosyncratic shocks are *i.i.d.* across agents and generations with a variance which depends on the aggregate state of the world. There are two assets traded: capital and bonds.

Our model is, admittedly, complex. However, a good case can be made that each ingredient plays a necessary role in addressing our question. The large number of generations, for instance, is important for quantitative questions which, ultimately, are the essence of the 'equity premium puzzle.' The *combination* of life cycle effects and persistent, countercyclically heteroskedastic shocks are, as we've argued above, critical for generating two necessary characteristics for models of our class: incomplete risk sharing and idiosyncratic-aggregate dependence. We find either element of this combination, in isolation, to be ineffective. Complexity, therefore, seems necessary to some extent. In spite of this, our framework yields a relatively straightforward, intuitive punch-line; a life cycle pattern in countercyclical, idiosyncratic risk gives rise to a life cycle pattern in portfolio choice, which concentrates aggregate risk on a subset of agents, who demand to be compensated for bearing it. This last feature — the asset pricing implications of an intergenerational concentration of aggregate risk — is also an important feature of related work by Constantinides, Donaldson, and Mehra (1997). Our results, however, are driven by starkly different economic forces than theirs, forces which generate strong, testable restrictions. We find it constructive to defer elaboration and further discussion to the concluding section.

By now it is apparent that our story, both in a qualitative and a quantitative sense, hinges on the statistical properties of idiosyncratic shocks. What we require is that idiosyncratic shocks be highly volatile relative to aggregate shocks, that they be quite persistent, and that they display countercyclical increases in variance. We follow Heaton and Lucas (1996) in using data from the Panel Study on Income Dynamics (PSID) in order to assess the plausibility of such processes, as well as to be precise about various quantitative magnitudes. Our methodology is distinct in a number of ways, each of which is motivated by our use of the OLG model as a window through which to view the data. We construct our panel, for instance, so as to maintain a stable demographic structure over its time dimension. Doing so seems important for questions of how idiosyncratic and aggregate shocks are related to one another. We also interpret the data as being generated by a class of finite processes, thereby avoiding (at a cost) well known issues associated with highly autocorrelated, possibly nonstationary, time series. This in turn allows us to address an important problem associated with using panel data to examine aggregate-idiosyncratic interactions, the relatively short amount of cyclical information spanned by most available panel data sets. Our methodology, in spite of using panel data spanning only the years 1968-1991, incorporates information on aggregate shocks dating back to 1910. What we find, which stands somewhat in contrast to Heaton and Lucas (1996), is supportive of the importance of idiosyncratic risk for asset pricing. Our estimates suggest substantial persistence in idiosyncratic shocks — something not inconsistent with the literature on labor market dynamics — with a unit root not entirely outside the realm of possibilities. We also find strong evidence of countercyclical heteroskedasticity. Our estimates suggest that idiosyncratic, conditional volatility more than doubles during aggregate downturns.

Our findings are as follows. Our baseline economy generates a Sharpe ratio of just over 8 percent, thereby accounting for roughly 20 percent of the Sharpe ratio associated with the U.S. stock market. This represents a substantial improvement over existing, production-based models, in most cases by several orders of magnitude. With a modest increase in risk aversion — a coefficient of 4.5 as opposed to 2 — our economy generates a Sharpe ratio which matches that of the U.S. stock market. When compared with a calibrated version of the Constantinides and Duffie (1996) model, our model generates slightly higher Sharpe ratios, but in an environment with non-degenerate trade and a plausible degree of risk sharing. We attribute this primarily to life cycle effects. Where our framework falls short is a common stumbling block for production-based models: asset return volatility. In our baseline case the standard deviation of the excess return portfolio is slightly more than an order of magnitude smaller than its U.S. counterpart.

In addition to the papers mentioned above, our work is related to previous efforts as follows. Krusell and Smith (1997) study an infinite horizon environment which has much in common with our study. Their results serve as an important benchmark with which the life cycle effects we emphasize can be compared and contrasted, something which we emphasize throughout the text. Our computational methodology is indebted to developments in both Krusell and Smith (1997), its predecessor, Krusell and Smith (1998), as well as den Haan (1994). Heaton and Lucas (1996) serves as an important benchmark, both in terms of their asset pricing explorations (they also examine the effects of persistence and countercyclical heteroskedasticity) and their use of PSID data in an asset pricing context. A large literature, including Abowd and Card (1989), Altonji, Hayashi, and Kotlikoff (1991), Altug and Miller (1990), Attanasio and Davis (1996), Deaton (1991), Deaton and Paxson (1994), Hubbard, Skinner, and Zeldes (1994), MaCurdy (1982) and Mrkaic (1997), has used panel data to examine the time series properties of idiosyncratic income risk and how they relate to various risk sharing issues. Our study is distinguished in terms of the composition of our panel, various aspects of our statistical methodology, and most importantly, our use of a formal, quantitative general equilibrium model. Finally, the stationary OLG framework we use owes much to Ríos-Rull (1994) and subsequent work by Huggett (1996) and Storesletten (1999).

The remainder of the paper is organized as follows. In section 2 we describe our model, its equilibrium and our solution technology. In section 3 we describe the sampling and statistical methodology with which we use PSID data to obtain estimates of the parameters of our model. Section 4 uses these estimates to examine the quantitative properties of our model, section 5 examines some alternative parameterizations and compares our results to a calibrated version of Constantinides and Duffie (1996), and section 6 offers conclusions and suggestions for future work.

2 The Overlapping Generations Economy

Our formulation of a stationary OLG environment incorporates idiosyncratic risk into the framework of Ríos-Rull (1994). Agents are indexed by their age, h, where $h \in \mathcal{H} = \{1, 2, \ldots, H\}$. Each of the H age cohorts consists of a large number of atomistic agents who face uncertain lifetimes with maximum length of H years. Each year a new cohort of agents are born and some positive fraction of each existing cohort dies. 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 a preference ordering over consumption distributions, an endowment process and an asset market position. Preferences for a unborn agent are represented by,

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

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.

We adopt the following process for the logarithm of hours worked,

$$\log n_h = \kappa_h + z_h + \varepsilon_h \quad , \tag{2}$$

where ε_h is *i.i.n.d.* with mean zero and variance σ_ε^2,

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

Z is an aggregate productivity shock and κ_h is a parameter used to characterize the cross sectional distribution of mean income across age cohorts. This parameterization is chosen on both theoretical and empirical grounds. Empirically, the extensive literature on income dynamics, from which we draw guidance in our own econometric work, finds it a useful decomposition between persistent and transitory sources of income variation. Its theoretical relevance derives from previous work suggesting that persistent and transitory shocks are likely to have very different impacts on equilibrium outcomes. Consequently, the fraction of the variation in n_h attributable to each source of variation will be an important question in the measurement exercise in the next section.

An equally important measurement issue is the functional relationship between the conditional variance of the persistent process, z_h, and the aggregate shock process, Z. As Constantinides and Duffie (1996) and Mankiw (1986) have suggested, an inverse relationship is likely to lead to higher risk premia being attached to an asset with payoff related to aggregate productivity. We choose the following simple specification.

$$\begin{aligned}\sigma_\eta^2(Z) &= \sigma_H^2 \text{ if } Z \geq E(Z) \\ \sigma_\eta^2(Z) &= \sigma_L^2 \text{ if } Z < E(Z) \ .\end{aligned}$$

The notion of 'countercyclical cross sectional variation' is simply the condition $\sigma_H < \sigma_L$. We impose no such *a priori* restriction at this point, choosing instead to estimate the magnitudes of σ_H and σ_L in the next section. In the theoretical exercise which follows, we will implement the joint process for conditional variance and the aggregate shock as a (restricted) discrete state Markov chain.

Turning to the financial market structure, agents can trade in two assets: one-period, riskless bonds which are in zero net supply and shares of ownership in the risky aggregate technology. We refer to the latter as 'capital holdings.' Each agent's choice problem, therefore, amounts to a consumption savings decision and a portfolio allocation decision. We use the notation b_h and k_h to denote beginning-of-period bond and capital holdings, respectively, of an agent who is h years old. The assets (or liabilities) of the fraction of agents who die each year are assumed to vanish, but *are* incorporated into the conditions which define market clearing. Altering this last assumption — allowing for lump sum redistribution, for instance — is not difficult but would have only minor qualitative effects on our results.

Output is produced by an aggregate technology to which individuals rent their labor services and capital. The production function takes the form,

$$Y = Zf(K, N) \ , \tag{3}$$

where K and N represent per capita capital and labor, respectively, Y represents per capita output and Z is a technology shock restricted to lie in a finite set, $\mathcal{Z}$. Given aggregate consumption, C, and the rate of depreciation on aggregate capital, δ, the law of motion for aggregate capital can be written,

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

This completes the description of the physical environment. We can now represent the state of the economy as a pair, (Z, μ), where μ is a measure defined over an appropriate family of subsets of $S = (\mathcal{H} \times \tilde{\mathcal{Z}} \times \mathcal{A})$, $\tilde{\mathcal{Z}}$ is the product space containing all possible idiosyncratic shocks (permanent and transitory), and $\mathcal{A}$ denotes the set of possible beginning-of-period wealth realizations. In words, μ is simply a distribution of agents across ages, idiosyncratic shocks and wealth. The aspect of μ which is somewhat non-standard is that, because of the aggregate uncertainty in our economy, it must evolve stochastically over time (*i.e.*, μ belongs to some family of distributions over which there is defined yet another probability measure). We therefore use G to denote the law of motion of μ, the cross sectional distribution of the economy, and write,

$$\mu' = G(\mu, Z, Z') \quad .$$

This characterization of the state of the economy allows us to express prices as $P(\mu, Z)$, $R(\mu, Z)$ and $W(\mu, Z)$, where P, R and W denote the bond price, the market clearing rate of return on capital and the wage rate, respectively. Our timing convention is that portfolio decisions are made at the end of the current period, and capital market returns are paid the following period at the realized capital rental rate. The decisions of an agent of age h are therefore constrained by,

$$\begin{aligned} c_h + k'_{h+1} + b'_{h+1} P(\mu, Z) &\leq a_h + n_h W(\mu, Z) \\ a_h &= k_h R(\mu, Z) + b_h \\ k'_{h+1} &\geq \underline{k} \\ b'_{h+1} &\geq \underline{b} \end{aligned} \tag{4}$$

where a_h denotes beginning-of-period wealth, k_h and b_h are beginning-of-period capital and bond holdings, and k'_{h+1} and b'_{h+1} are end-of-period holdings. Portfolio constraints are denoted by $\underline{k}$ and $\underline{b}$ and, although we do not make the dependence explicit, can be both state and agent-specific. For example, our computational framework allows for borrowing constraints which are expressed as a fixed fraction of individual wealth, where this fraction varies across age cohorts. In addition, we impose the terminal conditions, $k'_{H+1} \geq 0$ and $b'_{H+1} \geq 0$.

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

$$\begin{aligned} V_h(\mu, Z, z_h, a_h) \quad = \quad & \max_{k'_{h+1}, b'_{h+1}} \Big\{ u(c_h) + \\ \beta \frac{\phi_{h+1}}{\phi_h} E\left[V'_{h+1}(G(\mu, Z, Z'), Z', z'_{h+1}, k'_{h+1} R(G(\mu, Z, Z'), Z') + b'_{h+1})\right] & \Big\} \end{aligned} \tag{5}$$

subject to equations (4).

2.1 Equilibrium

An equilibrium consists of stationary price functions, $P(\mu, Z)$, $R(\mu, Z)$ and $W(\mu, Z)$, a set of cohort-specific value functions and decision rules, $\{V_h, k'_{h+1}, b'_{h+1}\}_{h=1}^{H}$, and a law of motion for μ, $\mu' = G(\mu, Z, Z')$, such that the firm's profit maximization problem is satisfied,

$$\begin{aligned} R(\mu, Z) &= Zf_1(K, N) - \delta + 1 \\ W(\mu, Z) &= Zf_2(K, N) \ , \end{aligned}$$

the bond market clears,

$$\int_S b' \, d\mu = 0$$

aggregate quantities result from individual decisions,

$$\begin{aligned} K &= \int_S k \, d\mu \\ N &= \int_S n \, d\mu \ , \end{aligned}$$

agents' optimization problems are satisfied given the law of motion for (Z, μ) (so that $\{V_h, k'_{h+1}, b'_{h+1}\}_{h=1}^{H}$ satisfy problem (5)), and the law of motion, G, is consistent with individual behavior.

2.2 Computation

The use of computational solution algorithms has by now become commonplace in the study of dynamic equilibrium models. In the quantitative analysis which follows, we use a fairly standard discretization-based approach to solve the dynamic programming problem (5). One feature of our problem which is somewhat non-standard is that we must characterize the law of motion for a distribution (μ), in addition to the optimal policy and value functions. We make use of techniques developed in the context of infinite-horizon economies by Castañeda, Díaz-Giménez, and Ríos-Rull (1994), den Haan (1994), Krusell and Smith (1998) and, in particular, Krusell and Smith (1997) which was, to the best of our knowledge, the first paper featuring aggregate shocks, many agents and more than one asset.

The basic idea is to approximate μ with a finite number of moments (or other statistics), and then characterize the law of motion for this set of moments. Having done so, we are left with a finite dynamic programming problem at the individual agent level which is relatively manageable. Solving the latter leads to a distribution of agent-specific capital and bonds, based on the individual policy functions, which can be compared to the conjectured functional form for that distribution. The process is iterated until the realized and perceived processes for the distribution coincide. The actual implementation, which requires a number of intermediate level decisions on the part of the researcher, is described in more detail in appendix B.

3 Measuring Household Specific Risk

The driving force in our theory is the exogenous stochastic process for each individual agent's endowment of labor hours, equation (2). The primary parameters of interest are σ_ε, which will govern the relative magnitude of the persistent and transitory shocks, σ_H and σ_L, which measure the extent to which variation in the cross sectional distribution is countercyclical, and ρ, the persistence parameter. In this section we use panel data on U.S. households from the Panel Study on Income Dynamics (PSID) to estimate these parameters.

The PSID quantity which we will interpret as a household's 'endowment' is the labor market earnings of all adult household members, plus any transfers received such as unemployment insurance, workers compensation, transfers from non-household family members, and so on. Several important considerations underly this definition. First, we include transfers because our model abstracts from the implicit insurance mechanisms which these payments often represent. That is, we wish to measure the amount of income variation which impinges on household financial decisions *net* of risks which are insured against by programs such as unemployment insurance. In a similar vein, we study the household as a single unit in order to measure household risk net of things like substitution in labor supply between household members in response to some shock. Finally, we focus on household income in spite of the fact that our model's exogenous process is hours worked. Doing so seems appropriate, primarily because our model abstracts from the indivisibility in labor supply which is so evident in data on hours worked. In addition, by measuring idiosyncratic risk via household income we allow for a more straightforward incorporation of the various types of transfers discussed above. In section 4 we verify that the statistical properties of the endogenous process for labor income in our model are very similar to those of hours worked (since our theoretical wage process is relatively stable), thereby providing a sense in which we actually do calibrate theoretical income to PSID income.

Turning to the specifics of our PSID extraction, we depart from the common approach of constructing a longitudinal panel with an equal number of time series observations on a fixed cross section of households. For our purposes, a longitudinal panel is problematic for a variety of reasons. It is likely to exhibit a kind of survivorship bias — the bias introduced by incorporating only households which report income in every survey year — as well as contain a relatively small cross section, in particular if the time dimension is large. The main problem, however, is that average age in a longitudinal panel necessarily increases by one year for each annual cross section. For example, in the longitudinal PSID panel which we analyze (for comparison's sake) in appendix A, mean age increases from 39 to 62 between the years 1968 and 1991. This is problematic for us because a primary issue is the relationship between aggregate shocks and the variability of idiosyncratic shocks. In a longitudinal panel a large fraction of household heads will have been retired, or at least been in their late earning years, during the last 2 of only 5 business cycles witnessed during the period 1968-1991. It seems likely, therefore, that a longitudinal analysis will understate any existent relationship between aggregate and idiosyncratic shocks. This, of course, will compound upon the fact that one is trying to uncover a fundamental relationship based on, in a sense, only 5 cyclical observations. The methodology we propose addresses each of these issues. It achieves a very stable cross sectional distribution of age across the time dimension of the panel and, as we explain below, exploits data on cyclical fluctuations dating back as far as 1910.

The specifics of our methodology are as follows. We use annual data from each of the surveys dated 1969 through 1992. Since each survey pertains to household data from the previous year, we refer to the time dimension of our panel as being 1968 through 1991. For each of these years we construct a three year panel consisting of households which reported strictly positive total household earnings (inclusive of transfers) for the given year and the next 2 consecutive years in the survey. For example, our 1970 panel is essentially a longitudinal panel on 1,663 households over the years 1970, 1971 and 1972. This results in a sequence of 22 overlapping panels, where the last one contains data from 1989, 1990 and 1991. As is shown below, these overlapping panels of three time periods are sufficient to identify the parameters of our time series process (essentially, the first two autocovariances), while at the same time mitigating survivorship bias and generating a stable cross sectional distribution of household age.

As is commonplace in PSID studies, we apply a number of additional filters with the goal of obtaining a more stable panel. Households are restricted to be those which report a male head and which have not reported any change in structure during the 3 years corresponding to the particular sub-panel, with the exception of an increase or decrease in the number of children. Any household which reported an earnings growth rate of greater than twenty-fold, followed by a decline greater than twenty-fold, is deleted from the sample on the grounds of extreme measurement error. We also follow previous studies in attempting to control for PSID oversampling of poorer U.S. households by excluding those which were originally included as part of the Survey of Economic Opportunity. These restrictions are helpful in that they negate, for instance, the need to incorporate new families and keep track of families which split-up, both of which raise difficult issues in estimating a time series model. The drawback, obviously, is that our results will not directly incorporate a number of idiosyncratic sources of variation — divorce for instance — which may be important, uninsurable determinants of household savings and portfolio choice. Enhancing the PSID sampling criteria to include such effects is, in our minds, an important avenue for future work.

Two final transformations we apply are to deflate nominal income using the CPI and, in order to incorporate differing family size, to divide total household earnings by the number of household members. The end result is 22 overlapping panels, each with a time dimension of 3 years. The cross sectional distribution of age is quite stable over each of the panels; the mean and standard deviation of the average age in each panel is 44.2 and 1.1, respectively. The number of households is substantially larger than would be possible in a longitudinal sample, with a mean and standard deviation (across panels) of 2045 and 228 observations, respectively.

Further details on the exact composition of our panel are available in appendix A, where we also report results based on a longitudinal panel of PSID households.

3.1 Summary Statistics

Prior to estimating a formal time series model for individual income, we find it informative to begin by highlighting several basic cross sectional properties of our panel. We focus on two dimensions which will be important for our questions: age and time. The former will have implications for the dynamic properties of our idiosyncratic income processes, whereas the latter will be informative for what Constantinides and Duffie (1996) and Mankiw (1986) have suggested is an important moment for asset pricing: the relationship between aggregate shocks and cross sectional variation in the earnings distribution. For the sake of transparency we choose not to pre-filter our data whatsoever for this informal first-pass. In subsequent sections we control for aggregate shocks and several demographic variables in order to focus more specifically on unexpected, idiosyncratic variation in the cross section.

Figures 1 and 2 report the first four sample moments associated with the age-dependent, cross section in our panel, where the time series' on a particular age cohort are pooled over the 24 years of data in the panel. Figure 1 reports moments for the raw data, whereas Figure 2 reports moments for logarithms of income, the idea being that the cross sectional distribution is more closely associated with log-normality than normality (all subsequent analysis will involve log-income). We see that the means demonstrate a fairly strong life cycle pattern, as do the standard deviations which increase over most of a household's earning years and then diminish sharply over the retirement years. The higher moments of the logarithmic data (the raw data are clearly highly non-normal) are imprecisely estimated, but seem to indicate that negative skewness and excess kurtosis are more strongly associated with the peak earning years that either the early or retirement years.

These moments, the first two in particular, have important implications for our theory. The age-dependent variation in the cross sectional mean will be helpful in calibrating our model, which has a fairly rich demographic structure. Just as important is the way in which cross sectional dispersion increases with age, suggesting some combination of persistence and/or heteroskedasticity at the individual time series level. For instance, if idiosyncratic shocks are homoskedastic and follow a unit root process, the cross sectional variance will increase linearly: a feature which is not qualitatively at odds with Figure 2. The identification of the relative magnitudes of persistence and age-dependent heteroskedasticity — something which is critical for our economic questions — will be an important task required of the time series model to be estimated in the next section. The bulk of the existing literature suggests that heteroskedastic shocks with low persistence and homoskedastic shocks with high persistence have very different implications for risk sharing and asset pricing.

In Figure 3 we cut the data by time instead of age. Figure 3 reports annual observations on the first four sample moments of the date-specific cross sectional distribution of log-income, 1968-1991. The means trend upwards at an average rate of 1.02 percent with cyclical properties which are quite similar to U.S. NIPA data; the correlation between the deviations from trend in our cross sectional means and the deviations from a linear trend in real U.S GNP is 0.72. The relatively low average growth rate of mean income — 1.02 percent as opposed to 2.13 percent in U.S. GNP — is consistent with the notion that the PSID oversamples poorer U.S. households.

Our primary interest in Figure 3 involves the time series relationship between the cross sectional mean and standard deviation. In the first two graphs we plot the raw mean followed by deviations from a linear trend as well as the growth rate in the mean. The third graph reports the coefficient of variation, which is motivated by the fact that the standard deviation grows slightly alongside the mean (the qualitative implications are not changed if we use the unscaled standard deviation). Casual observation indicates an inverse relationship between either measure of the first moment and the level of cross sectional dispersion. The correlation coefficient between the deviations from trend and the coefficient of variation is -0.85. The correlation between the growth rate and the coefficient of variation is -0.33 (the analogous values associated with the standard deviation — instead of the coefficient of variation — are -0.67 and -0.34, respectively). In either case, this first-pass look at the data is supportive of the notion of 'countercyclical cross sectional variation,' something which will drive our theory's asset pricing properties and which has been a focal point of previous work (*e.g.*, Constantinides and Duffie (1996), Mankiw (1986)). We now turn to an explicit time series model in order to give these casual observations more precision and a well defined statistical foundation.

3.2 Time Series Model

We denote the natural logarithm of household i's endowment at time t as y_{it}, and the per capita aggregate endowment as y_t. Our theory requires that household-specific endowment processes are comprised of an aggregate and an idiosyncratic component. We perform the following decomposition of y_{it} with this in mind,

$$y_{it} = g_{it}(y_t) + u_{it} \quad . \tag{6}$$

The component $g_{it}(y_t)$ is comprised of aggregate shocks as well as deterministic components of household-specific earnings such as unobservable 'fixed effects' and deterministic variation attributable to household age, education level and so on (for simplicity, our notation omits dependence on a deterministic vector of time-varying household characteristics). The component u_{it} is therefore the random component of a household's endowment which is idiosyncratic. The condition which identifies u_{it} is,

$$\tilde{E}_t(u_{it}) = 0 \ , \ \forall t$$

where the notation $\tilde{E}_t$ denotes expectation with respect to the date t, cross sectional probability measure. Allowing for time variation in the cross sectional distribution is important for both our theory and our interpretation of the PSID, because we wish to allow for aggregate shocks.

From a theoretical perspective, it is the statistical properties of u_{it} which are critical in determining how effectively individuals can pool risk using asset markets. This is perhaps most clearly demonstrated in the autarkic economies of Constantinides and Duffie (1996), where u_{it} is chosen so as to generate a unit root in equilibrium marginal utility. In this sense, the specification of $g_{it}(y_t)$ is not critical. Loosely speaking, we expect to see similar types of risk sharing behavior in economies with aggregate income dynamics which are quite different.

The sense in which $g_{it}(y_t)$ is critical is that it affects how one uses data to infer an appropriate process for u_{it}. We follow much of the literature on income and employment dynamics (*e.g.*, Abowd and Card (1989), MaCurdy (1982), Hubbard, Skinner, and Zeldes (1994)) and specify $g_{it}(y_t)$ as a time-dependent intercept term (*i.e.*, a set of annual dummy variables) in addition to a quadratic function of age and a linear function of education:

$$g_{it}(y_t) = a_0 + d_t a_1 + h_{it} a_2 + h_{it}^2 a_3 + e_{it} a_4 + \text{residuals} \ , \tag{7}$$

where d_t is a vector of time-dependent dummy variables, h_{it} is the age of the household head at time t and e_{it} is the number of years of education undertaken by this person as of date t. This specification is chosen to incorporate two sources of variation: aggregate variation and deterministic cross sectional variation. It is, of course, arbitrary, but is motivated by several considerations which are relevant for our questions. In regard to aggregate variation, the cross sectional dimension of our data allows for a relatively non-parametric approach in the use of time dummies. Note that, in the absence of age and education terms, time dummies are (roughly) equivalent to specifying u_{it} as the logarithm of household $i's$ share of date t aggregate income, the approach taken by Heaton and Lucas (1996). The only difference involves a term related to Jensen's inequality which turns out to be unimportant. Details are provided in appendix C. We also experimented with linear and quadratic time-trends, and found no major qualitative differences.

The extent to which equation (7) does an adequate job in incorporating deterministic, household-specific variation, is, in our view, on less firm ground. Education is but one (albeit an important one) in a long list of variables, including unobserved 'fixed-effects,' which the labor and income dynamics literature has focused upon as being important determinants of where a household lies in the cross sectional distribution. Our approach is to account for some fraction of this cross sectional dispersion in a manner which is parsimonious in terms of the list of variables, economical in terms of parameters, and which does not stray too far from the economic forces at play in our model. While we did not venture beyond the inclusion of education, we did experiment with higher order terms, functions of education intended to capture employment potential (*c.f.* Gottschalk and Moffit (1992)) as well as first-differencing and quasi-differencing. The alternative functions of education had very little effect on our results. Differencing the data, the main target being unobserved 'fixed-effects,' affected our estimates in a quantitatively important manner but had only minor qualitative implications, especially as they pertain to the economics of our model. This is discussed further in appendix A. Overall, we feel that equation (7) does a suitable job for our particular question, which is not to account for the rich complexity in the cross sectional income distribution, but to obtain economically plausible estimates for the parameters of the simple time series process which drives our model.

We estimate the coefficients in equation (7) using least-squares and obtain an estimated time series for u_{it} as the residuals from this regression (all standard errors incorporate the sampling uncertainty associated with this first-stage regression). A fundamental aspect of our estimation methodology is that it conditions on household age. We therefore append our notation accordingly, denoting u_{it}^h as the idiosyncratic shock observed for household i of age h at time t. We then estimate the following time series model for u_{it}^h, which is almost identical to the theoretical process in equation (2):

$$\begin{aligned} u_{it}^h &= z_{it}^h + \varepsilon_{it} \\ z_{it}^h &= \rho z_{i,t-1}^{h-1} + \eta_{it} \ , \end{aligned} \tag{8}$$

where household age, h, is made explicit only when the conditional distribution of a variable depends upon it. We assume that $\varepsilon_{it} \sim \text{Niid}(0, \sigma_\varepsilon^2)$, $\eta_{it} \sim \text{Niid}(0, \sigma_\eta^2(Y_t))$ and,

$$\begin{aligned} \sigma_\eta^2(Y_t) &= \sigma_H^2 \ \text{ if aggregate expansion at date } t \\ &= \sigma_L^2 \ \text{ if aggregate contraction at date } t \ , \end{aligned}$$

where Y_t denotes aggregate income. Our methodology will admit any definition of what constitutes an aggregate expansion and contraction. For the results we present, we will use National Income and Product Account (NIPA) data, and define an expansion (contraction) as a year in which growth in real U.S. GNP per capita is above (below) its average over our sample.

The process (8) is attractive in that it follows a number of studies on labor and income dynamics (*e.g.*, Hubbard, Skinner, and Zeldes (1994)) in admitting both a transitory component, ε_{it}, and a persistent component, z_{it}^h. Doing so incorporates, among other things, *i.i.d.* measurement error, thereby mitigating the overstatement of variation attributable to persistent shocks and reducing the downward bias which measurement error exerts on the estimate of the parameter ρ. We also find (8) attractive from a theoretical perspective in that, when incorporated into our theory, it allows us to examine the allocational effects of the interaction between persistent and transitory idiosyncratic shocks in an environment where 'aggregation' in the sense of Constantinides and Duffie (1996) does not apply.

To estimate the parameters of the process (8) we exploit two important features of our dataset: information on household age and information on the macroeconomic history which a given household experienced while working. The interaction between these two variables — or, alternatively, the interaction between cross sectional variation in age and in macroeconomic history — is essentially what allows for the identification of σ_H and σ_L. It also allows us to exploit information on aggregate shocks dating back to 1910, not just the four or five business cycles which occurred during our PSID sample period. The following simple example demonstrates how this is accomplished.

Consider an individual (we omit the i subscript for now) who has been working for two years, at dates t, and $t+1$. The moving average representation of their idiosyncratic shock process is,

$$u_{t+1}^2 = \rho^2 z_{t-1}^0 + \rho\eta_t + \eta_{t+1} + \varepsilon_{t+1} \tag{9}$$

Assuming that their initial condition, z_{t-1}^0, is zero, (we discuss alternative assumptions and results below and in the appendix), our distributional assumptions imply that, conditional on knowledge of the history of aggregate shocks, the variance of u_{t+1}^2 is,

$$Var(u_{t+1}^2) = \rho^2[I_t\sigma_H^2 + (1-I_t)\sigma_L^2] + [I_{t+1}\sigma_H^2 + (1-I_{t+1})\sigma_L^2] + \sigma_\varepsilon^2 \ , \tag{10}$$

where $I_t = 1$ if the aggregate economy at date t is in an expansion and $I_t = 0$ otherwise. Moments analogous to equation (10) form the basis of the GMM estimator we employ. This estimator has several noteworthy advantages. First, by interpreting our data as a collection of finite processes, we avoid the well known plethora of issues related to non-stationary time series'; moments such as (10) are well defined even for values of ρ greater than unity. The costs, of course, are assumptions regarding initial conditions and/or the possibility of differencing stationary time series (discussed below). Given that our model is distinguished by finite time series, however, we feel that bearing such costs is advantageous.

The second — and perhaps more important — aspect of expressions analogous to (10) is that they capture an interaction between persistence, aggregate variation and age which identifies a crucial aspect of our analysis: countercyclical cross sectional variation. Simple inspection of (10) reveals that for an agent who has been working for h years there will be h terms in the distributed lag of indicator functions, each multiplied by a power of ρ. It is this aspect of our GMM estimator, in addition to the use of NIPA data, which incorporates aggregate shock information dating back to the year 1910 (corresponding to the year in which the oldest individual in our 1968 panel attained the assumed initial working age of 22). As an illustrative example, consider the 70 year old cohort in the first year of our panel, 1968. Given that these individuals were of working age through a greater number of contraction years than the 70 year old individuals in the 1991 panel (the former were of (adult) working age during the Great Depression for instance), and given that idiosyncratic shocks are highly persistent and more volatile during downturns, one would expect to see greater cross sectional variation amongst the 1968 group than the 1991 group. This interaction between age, macroeconomic history and properties of the cross sectional distribution is implicit in the moment conditions underlying our estimator and is important in distinguishing our approach from previous work.

The general version of equation (9), associated with a household of age h, is,

$$u_{it}^{h} = \sum_{j=0}^{h-1} \rho^{j} \eta_{i,t-j} + \rho^{h} z_{t-h}^{0} + \varepsilon_{it} \quad . \tag{11}$$

The following moment conditions are the means with which we estimate the three parameters, ρ, σ_{ε} and σ_{η}.

$$\begin{aligned}
\tilde{E}_t \left[(u_{it}^{h})^2 - \sigma_{\varepsilon}^2 - \sum_{j=0}^{h-1} \rho^{2j} (I_{t-j}\sigma_H^2 + [1 - I_{t-j}]\sigma_L^2) \right] &= 0 \\
\tilde{E}_t \left[u_{it}^{h} u_{it-1}^{h-1} - \rho \sum_{j=1}^{h-1} \rho^{2(j-1)} (I_{t-j}\sigma_H^2 + [1 - I_{t-j}]\sigma_L^2) \right] &= 0 \\
\tilde{E}_t \left[u_{it}^{h} u_{it-2}^{h-2} - \rho^2 \sum_{j=2}^{h-1} \rho^{2(j-2)} (I_{t-j}\sigma_H^2 + [1 - I_{t-j}]\sigma_L^2) \right] &= 0
\end{aligned} \tag{12}$$

Again, an important assumption underlying the above conditions is that $z_{it-h}^{0} = 0$, or that all households start with the population mean (recall, however, that cross sectional dispersion which is related to age and educational levels has been controlled for). In Appendix A we discuss some robustness checks along these lines (*e.g.*, differencing and quasi-differencing the data) and find the qualitative nature of our results to be relatively stable.

Table 1 reports parameter estimates obtained using GMM in conjunction with the moments in equations (12). In Panel A we report estimates which constrain $\sigma_H = \sigma_L$ and in Panel B we relax this restriction. In either case we obtain a precise estimate of the autocorrelation parameter, ρ, greater than 0.90. Our estimate is 0.94 for the case with homoskedastic innovations and 0.92 for heteroskedastic innovations. The fraction of the conditional variance attributable to persistent innovations is substantially larger than transitory innovations, with the standard deviation being roughly twice as large in either specification (based on the average of σ_H and σ_L in Panel B). Finally, our estimate of the degree to which idiosyncratic shocks are more volatile when the aggregate economy is contracting is striking. Our estimate of σ_H is 0.43 and our estimate of σ_L is 0.19, which represents an increase of 126% from expansion to contraction. This finding stands in stark contrast to previous work. Heaton and Lucas (1996), for instance, found an analogous increase in volatility of roughly 27%. We attribute the differences mainly to our use of information on many more cyclical fluctuations, our explicit conditioning on age, and our alternative panel which maintains a stable age structure. These issues, as well as several others, are discussed in appendix A where we replicate Heaton and Lucas's (1996) results using a longitudinal panel in addition to providing some robustness checks on the results in Table 1.

The findings of several related papers provide a useful frame of reference for our results. The estimates of the magnitude of idiosyncratic shocks (the conditional variances of η_{it} and ε_{it}) are somewhat larger than several previous authors. Where we find a conditional standard deviation (assuming homoskedastic innovations) for the persistent shocks of 0.25, Heaton and Lucas (1996) and Hubbard, Skinner, and Zeldes (1994) report 0.24 and 0.18, respectively. This is not surprising given that our overlapping panel admits a wider variety of households and, to a certain extent, mitigates survivorship bias. We also find the fraction of the innovation variance attributable to transitory shocks to be smaller than previous studies. Hubbard, Skinner, and Zeldes (1994), for example, estimate that the transitory and persistent variances are roughly of the same magnitude, whereas we find the former to be approximately half the size of the latter.

Our estimates of the autocorrelation parameter, ρ, are substantially larger than those of Heaton and Lucas (1996) (who obtain an estimate of 0.53), but are quite similar to those of a number of other papers, including Abowd and Card (1989), Hubbard, Skinner, and Zeldes (1994) and MaCurdy (1982). Moreover, our summary statistics indicate (Figure 2), as does the more exhaustive work of Deaton and Paxson (1994), that cross sectional dispersion in income increases at close to a linear rate over the life cycle. This is what one would expect, should idiosyncratic shocks follow a unit root process and be independent across households. Our estimates incorporate this cross sectional evidence, but only implicitly (via age-dependence in our moments). Our guess is that a more formal treatment of how inequality increases with age would generate an even higher value for the parameter ρ than we report in Table 1 (Storesletten, Telmer, and Yaron (1997) investigate this further).

Finally, providing a frame of reference for our results on heteroskedastic idiosyncratic shocks over the business cycle is more problematic; aside from Heaton and Lucas (1996) we are not aware of comparable studies. In the next subsection we attempt to provide some corroborating evidence by examining the manner in which overall, cross sectional dispersion — something we feel relatively confident in being able to measure accurately — is likely to change, given the magnitude of the heteroskedasticity reported in Table 1.

3.3 Cross Sectional Dispersion

The upper right panel of Figure 3 provides the coefficient of variation, taken across our entire panel, for each of the years 1968-1991. The largest increase associated with an economic downturn is roughly 12%, which is associated with the recession in the early

1980's (the same answer is obtained using the cross sectional standard deviation). At first blush, this might seem inconsistent with the values reported in Table 1, where we estimate the increase in the conditional standard deviation of the persistent shock, coincident with a downturn, to be on the order of 126%. What's going on, however, is that the larger number is associated with the conditional distribution of a given agent's innovation, whereas the smaller number is much more closely associated with the unconditional distribution. The two are, of course, related. Understanding this relationship, something we achieve through simulation, is critical for understanding that relatively large increases in the conditional variance of idiosyncratic shocks can be consistent with relatively small movements in the overall level of cross sectional dispersion.

To see this, consider two extreme cases. First, suppose that our theoretical economy had been in an expansion (*i.e.*, the good aggregate shock) for many years, so that all existing age cohorts had only received persistent idiosyncratic shocks with variance σ_H^2. A cross sectional law of large numbers implies that the cross sectional variance amongst the oldest generations will be, approximately,

$$\sigma_\varepsilon^2 + \frac{\sigma_H^2}{1-\rho^2} \ ,$$

where the approximation arises from ignoring finiteness in any age cohort's history of shocks. Along similar lines, cross sectional variance amongst the youngest generation will simply be $\sigma_\varepsilon^2 + \sigma_H^2$. The overall cross sectional variance, that which includes agents of all age cohorts, will be a simple weighted average of these terms and terms associated with intermediate cohorts.

In Figure 4 we represent this extreme case — the stationary, population cross sectional standard deviation associated with the σ_H^2 distribution — as the lower dashed line. We also represent the other extremity — an analogous economy in which shocks are drawn from the distribution with variance σ_L^2 — as the upper dashed line. The solid line in the middle is simply the cross sectional standard deviation associated with one particular realization of our economy, where idiosyncratic shocks follow the heteroskedastic process, (8). The length of the simulation, 24 periods, is chosen to match the time series dimension of our panel.

The point of Figure 4 is that the overall cross sectional variance will not immediately move to the long run level associated with the conditional distribution, should a 'regime shift' occur. Rather, the cross sectional variance will be a moving average of the two extreme points, where the average will depend on the degree of persistence in both aggregate and idiosyncratic shocks, as well as the demographic structure of the population. The avenue through which aggregate persistence works is simply the frequency with which switches occur in the conditional distribution. Idiosyncratic persistence plays an important role, as is obvious if one considers the case of $\rho = 0$; the cross sectional variance associated with each cohort will be identical and will oscillate between high and low levels. Finally, demographics clearly matter in that, conditional on $\rho \neq 0$, they determine both the weights applied to the cross sectional variance associated with each age cohort and the number of terms in the moving average (which equals the number of working generations).

Figure 4 is intended for illustrative purposes and, consequently, reports data from just a single realization of our economy. We also conducted a more comprehensive Monte Carlo experiment, based on 1000 independent replications of a 24 year history of our economy. In each realization we computed ratio of the maximum to the minimum of the overall, cross sectional coefficient of variation of the log of labor income. This ratio corresponds, roughly, to the upper right panel of Figure 3, where the maximum occurs in 1981, the minimum occurs in 1973, and the ratio is just under 14 percent. The fraction of the realizations of our Monte Carlo experiment in which this ratio was 14 percent or less was 34 percent. This suggests that our estimates of σ_H and σ_L may overstate the extent to which idiosyncratic volatility increases in recessions, but that they are not outside the realm of possibilities. In other words, the amount of variation in the cross sectional variance of labor income — something we feel relatively confident in being able to measure accurately — is not entirely inconsistent with the relatively large amount of variation which we attribute to the agent-specific, conditional variance.

4 Quantitative Implications of the Theory

Our interpretation of the PSID evidence is as follows. First, a strong case can be made that the idiosyncratic component of a typical household's 'endowment' (defined above) contains a highly persistent component which accounts for roughly 88% of the overall conditional volatility. Second, an equally strong case can be made that, in the context of our model, the conditional variance of these shocks is countercyclical with the standard deviation increasing by as much as 120% during aggregate downturns. We now turn back to our theory and investigate the implications of these findings for asset prices and individual choices.

4.1 Parameterization: Baseline Economy

In this section we describe what we view as an appropriate 'baseline' economy: a set of parameter values which we feel are plausible. In subsequent sections we conduct a sensitivity analysis and more fully explore the properties of our economy under alternative parameterizations.

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

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

Following much of the business cycle literature, 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 technology shocks, Z, follow a first-order Markov chain with parameter values chosen so that theoretical aggregate *consumption* matches several important features of observed, aggregate U.S. consumption. We focus on consumption — as opposed to output — as the anchor for calibrating our aggregate economy because of our focus on asset pricing. Were we to follow the business cycle tradition of matching the variability in Solow residuals, our theoretical consumption process would be too smooth by a factor of two, excessively smooth consumption being a common feature of business cycle models. Given the central role of consumption dynamics in asset pricing, we find it natural to turn this 'puzzle' on its head and recast it as one of excessive output variability instead of excessive consumption smoothness (see Table 2).

Operationally, the secular growth rate in GNP per capita, by which we normalize all quantities in our model, is chosen to be 1.5% per year. The transition probabilities for Z are chosen so that the expected duration of a 'business cycle' is 6 years, whereas the possible realizations of Z are chosen to match the variability in theoretical and observed aggregate consumption. The end result is a two state Markov chain for the aggregate shock with $Z \in \{-0.057, 0.057\}$ and probability of remaining in the current state of 2/3.

In Table 2 we demonstrate that this parameterization for technological shocks generates realistic behavior for various (endogenous) aggregate quantities in our model. Doing so is important for our quantitative questions as it makes for meaningful comparisons with full insurance economies in which aggregate quantities are relevant for pricing. Table 2 shows that the variability in theoretical aggregate consumption matches that of the data, but that aggregate output is too volatile by a factor of 2.72. The same holds for investment, but by a factor of 1.65. The contemporaneous correlation of aggregate output and investment with consumption match those of the data, but their autocorrelations are somewhat smaller. Relatively low autocorrelation, however, is a necessary implication of our two-state aggregate shock process averaging one 'cycle' every 6 years. In other words, given our simple process for aggregate variation, we can match either autocorrelation or some notion of 'cyclicality,' but not both. We choose the latter, largely because of the methodology used in the last section to measure countercyclical, cross sectional variation.

Turning to the characteristics of individual agents, preferences are identical (up to age-dependent mortality risk) and are described by equation (1). We parameterize the period utility function with the standard isoelastic specification,

$$u(c) = \frac{c^{1-\alpha} - 1}{1 - \alpha} \quad .$$

For the time being, we set the risk aversion parameter, α, to 2 and the utility discount factor, β, to 0.95. The latter is chosen in order to generate an aggregate capital to output ratio of 3.3, which is close to the estimate reported in Cooley and Prescott (1995). The choice of α is arbitrary, but seems a sensible starting place in addition to representing a rough consensus of what is typical in the literature. In section 5 we experiment with alternative values for α and find that our model performs substantially better for relatively minor perturbations.

The demographic structure of our economy is calibrated to correspond to several simple properties of the U.S. work force. 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 and having to finance consumption from an existing stock of assets. Mortality rates are chosen to match those of U.S. females in 1991 and population growth is set to 1.0%.

The process for idiosyncratic labor supply, equation (2), is implemented as a discrete approximation to the autoregressive time series model and is parameterized using our point estimates from Table 1. The age dependent intercept terms, κ_h, are chosen so that, on average, our theoretical age-earnings profile matches that of the PSID. The transitory shocks, ε_h, follow a two-state binomial process with equally likely probabilities and a standard deviation of 0.160. This results in $\varepsilon_h \in \{-0.160, 0.160\}$. The persistent process, a regime-shifting autoregression for z_h with $\sigma_H = 0.425$, $\sigma_L = 0.192$ and $\rho = 0.916$, is approximated with an 11-state Markov chain where the conditional heteroskedasticity is generated by variation in the transition probabilities, as opposed to the potential realizations of z_h. Further details are provided in appendix B.

This calibration of our theoretical income process, equation (2), suffers from two potential sources of discrepancy relative to the PSID-based estimates from Table 1. The first is the error induced by approximating an infinite-state autoregression with a finite-state Markov chain. The second derives from the fact that labor income is an endogenous process in our model: it is the product of the exogenous supply of hours worked and the endogenously determined wage rate. Strictly speaking, we have calibrated the process for hours worked to data on income received. We find, however, that the combined effects of these sources of discrepancy are of negligible importance for the moments we focus on. Descriptive statistics — the mean, standard deviation and autocorrelation for example — are very similar when comparing population moments from our model with their sample analogs. More to the point, we find that if we use our theoretical economy to generate an arbitrarily long sequence of 3 year panel data sets (which corresponds to our PSID sampling method), the application of our GMM estimator (section 3.2) to these data yields estimates which closely match those from Table 1; our simulated point estimates are 0.902, 0.029, 0.172, and 0.028 for ρ, σ_H^2, σ_L^2 and σ_ε^2, respectively. The overall implication is that the population moments for our model's labor income process closely match the sample moments underlying the estimates presented in Table 1.

The only remaining items are the portfolio constraints. For our baseline economy we disallow short-selling of the risky technology — thereby setting $\underline{k} = 0$ — and restrict borrowing to be less than 50% of expected GNP per capita. That is, $\underline{b} = 0.5\, E(Y)$.

4.2 Results: Baseline Economy

Table 3 reports a variety of population moments for asset prices associated with the baseline economy we have described. In addition, Figure 5 plots the Sharpe ratio — the ratio of the average excess return on the risky technology to its standard deviation — from a sequence of economies which differ according to both the degree of persistence in idiosyncratic shocks as well as the degree to which the shock variance is countercyclical. We focus on the Sharpe ratio because the level of variability in the risky asset return in our economies is roughly two orders of magnitude smaller than that observed on the (levered) U.S. stock market. The low level of volatility in our theory is a byproduct of using a production economy, as opposed to the more common (but less restrictive) endowment economy used in most asset pricing studies. In our production economy, variation in the return on the risky asset is almost synonymous with variation in technology shocks, the only difference being variation in the level of aggregate capital. Since the latter is small (due to its aggregate nature), and since variation in our technology shocks is pinned down by variation in aggregate consumption — another low variance process — it is *un fait accompli* that the theoretical return to holding capital will be far less variable than that of the observed return to holding the U.S. stock market portfolio. Our feeling, therefore, is that an informative manner in which to compare theoretical and observed stock return data is to scale the two into equivalent volatility units and focus on what is commonly referred to as the 'price of risk.'[1]

The main message of Table 3 and Figure 5 is that our baseline economy represents a substantial departure from complete markets — a feature which stands in contrast to a number of existing studies — and that the combination of persistent and countercyclically heteroskedastic shocks can have a substantial impact on risk premia. In our baseline case, the Sharpe ratio is 8 percent, a factor of 8 larger than either the complete markets economy, or the economies with homoskedastic, idiosyncratic shocks. This effect is amplified substantially by unit root shocks, something we demonstrate in Figure 5 but defer discussing until we address alternative parameterizations in section 5. Relative to U.S. data, our baseline model accounts for roughly 20 percent of the Sharpe ratio paid to U.S. equity holders over the period 1956-1996 (the same holds over a longer horizon: see Mehra and Prescott (1985)).

It is informative to be somewhat more explicit about the contrast between our baseline results and those of existing studies on production-based, general equilibrium asset pricing models. Rouwenhorst (1995), for instance, considers a complete markets economy with equity claims being levered by roughly 40 percent. In spite of explicitly incorporating leverage (which we abstract from), he reports risk premia (and therefore Sharpe ratios) which are identical to zero, given the level of rounding in his Table 10.3. Ríos-Rull (1994) also reports values (see his Table 3) which make it impossible to distinguish the risk premium — either scaled or unscaled — from zero. Finally, and most informatively, Krusell and Smith (1997), find substantially smaller Sharpe ratios in a class of incomplete markets economies which are quite similar to ours along many dimensions (*e.g.*, the quantitative magnitude of idiosyncratic shocks, asset trading opportunities, etc.). In their Table 2 they report an equity premium which, assuming a level of aggregate variability comparable to ours, implies a Sharpe ratio two orders of magnitude smaller than that in our Table 3.

[1]An alternative would be to consider the return on a highly levered claim on our aggregate technology, and leave the volatility scale alone. Note that, unlike several previous papers in the asset pricing literature, the moment-matching exercise here would be that of matching equity return volatility using a leverage ratio, not the equity premium level.

The overall implication is that, although we are clearly talking about small numbers, our baseline calibration generates a return to bearing risk far in excess of most similar models. As we argue below, we attribute the difference to two inter-related forces: life cycle effects and countercyclically heteroskedastic idiosyncratic shocks.

The remainder of this section provides details geared towards understanding the economic structure of our baseline model and interpreting the quantitative results in Table 3. The discussion is organized in a hierarchical nature. We begin by describing consumption allocations or, equivalently, the mapping between the risk which agents are endowed with and that which they are (implicitly) unable to pool amongst themselves. Next we consider the flip side of the consumption decision — the savings decision — and describe the interaction between precautionary and life cycle savings motives. Finally we describe how agents allocate their savings between stocks and bonds, how the portfolio constraints affect their decisions and, as a result, market clearing asset prices. The role played by the countercyclical cross sectional variation will be of particular importance for these last items.

4.2.1 Risk Sharing

Asset prices in our model will be explicit functions of consumption allocations. The extent to which consumption allocations differ from those which would obtain under complete markets will govern the extent to which our model's driving force — idiosyncratic risk — can account for asset return anomalies where other approaches have failed. We therefore begin by exploring how well our model's asset markets facilitate the smoothing of consumption relative to income. A natural starting point is the unconditional variance of consumption relative to that of labor market income, reported in Figure 6 for different age cohorts. We choose to report the variance, as opposed to the standard deviation, because it is associated with a useful benchmark; if idiosyncratic shocks follow a unit root process, then the cross sectional variance of wage income will increase linearly with age.

Each point in Figure 6 represents the cross sectional variance, of income or consumption, associated with a specific age cohort. Each point also represents the variance of the distribution from which an unborn agent views their future labor income (or consumption) as being drawn. These graphs illustrate several simple, yet important, aspects of our economy. First, an unborn agent faces an increasing amount of labor market uncertainty the farther ahead they look; the standard deviation of labor income increases by a factor of 2.4 over their working life. Older agents, on the other hand, face less (conditional) uncertainty, a simple manifestation of finite lives (or, equivalently, retirement). For, example, the standard deviation of labor income received at age 65, from the perspective of an unborn agent, is roughly 0.90. The standard deviation of the same income realization, conditional on being 10 years away from retirement, is 0.81. This 'life cycle pattern' in labor market uncertainty will turn out to play a critical role in the portfolio choices made by agents of differing ages and, largely due to the CCV effect, in the determination of equilibrium asset prices.

Figure 6 also demonstrates that the allocations in our economy are characterized by partial risk sharing; agents are able to use decentralized asset markets to partially insure consumption outcomes against idiosyncratic income shocks. Quantitatively, the variance of consumption is, on average, roughly 2/3 that of labor income. To put this in context, autarky — interpreted as an agent being forced to consume their labor income (and abstracting from retirement) — would imply that consumption and income variances coincide. Complete markets, at the other end of the spectrum, implies a consumption standard deviation roughly 30 times less (on average) than our baseline case.[2] The upshot is that our economies generate far less risk sharing than many previous studies, in which allocations have been quantitatively close to first best, but far more than the autarkic abstraction of Constantinides and Duffie (1996). In a related paper, Storesletten, Telmer, and Yaron (1997), we argue that the magnitude of risk sharing inherent in our economy — as measured by the relationship between cross sectional measures of income and consumption dispersion — is consistent with U.S. panel data on income, consumption and wealth.

The information in Figure 6 relates to risk sharing outcomes defined in an unconditional (or age zero) sense. More important for asset pricing is the age-dependent, conditional distribution, which impacts directly on agent-specific Euler equations. The conditional distribution also provides more explicit information about how effectively agents are able 'self-insure' — dynamically trade in financial markets in order to partially replicate missing insurance markets — at different stages of the life cycle. Figure 7 plots what we find to be a good approximation of the conditional standard deviation of consumption: the unconditional standard deviation of the age-specific consumption growth rate. The figure shows results from our baseline economy and, in order to provide some useful benchmarks, results from an autarkic economy, a complete markets economy and economy with unit root idiosyncratic shocks.

Any complete markets allocation in our environment will have the characteristic feature that simple functions of consumption growth — marginal rates of substitution adjusted by mortality factors — will be equated between all agents, both within and across generations. Cross sectional dispersion in consumption growth will therefore be invariant across age. This is evident in the lower, dashed locus in Figure 7. In contrast, the incomplete markets economies, in which marginal rates of substitution are not equated, feature both higher overall consumption growth variability as well as a great deal of variation across age cohorts. This affirms what was apparent in Figure 6; agents are able to self insure to a certain extent — the age-averaged standard deviation of consumption growth is just less than half its autarkic counterpart — but risk sharing in our economy is far from first best. This is particularly true for the young, for whom consumption growth is more volatile than it would be under complete markets by a factor of just over 11. Averaging across all age cohorts, this difference is a factor of 8. Interestingly, after retirement — at which point an agent no longer

[2] By 'complete markets' we refer to an situation in which (a) the allocation is isomorphic to one in which all idiosyncratic shocks are zero and (b) each period there exists a full set of contingent claims for all aggregate states of nature in the subsequent period. The former can be implemented by a competitive financial intermediary who can observe agent-specific outcomes whereas the latter can be thought of in the standard 'dynamic spanning' sense (Arrow (1970), Kreps (1979)).

faces idiosyncratic risk — changes in consumption are *less* variable than in the complete markets economy. The main reason is the borrowing constraint. As we'll see shortly, a complete market allocation features retired agents holding a levered position in equity, a simple manifestation of intergenerational risk sharing (*i.e.*, working aged agents face more aggregate risk that retired agents because wages are more volatile than capital returns). In the incomplete market world, retired agents are borrowing constrained and are, therefore, unable to achieve the same degree of leverage as their complete market counterparts.

Just as important as the overall level of the graphs in Figure 7, is their shape. The baseline graph, for which idiosyncratic shocks die out over time, is hump shaped, peaking at age 28 and then monotonically diminishing. The reason is related to two offsetting forces: first, how the persistence of a given shock changes over the life cycle, and, second, how the composition of total wealth changes over the life cycle. The first effect is manifest in how an agent uses financial markets to self-insure. As the agent ages, any given shock looks increasingly persistent, as long as $0 < \rho < 1$. Age, therefore, dampens the extent to which agents will attempt to smooth out current period shocks through contingent savings and dissavings, tending to make consumption growth more volatile. The second, countervailing, effect is the composition of total wealth. For a young agent almost all their wealth is held as 'human capital:' the value of their future wage receipts. As an agent ages, financial capital is accumulated, both in order to finance retirement consumption and for use as a precautionary buffer. As an increasing fraction of total wealth is represented by financial capital, the marginal effect of an idiosyncratic shock diminishes. The result, as Figure 7 shows, is a declining pattern of consumption growth volatility, once the wealth composition effect dominates the self-insurance effect.

This interpretation of the hump-shaped pattern in Figure 7 is substantiated by the monotonically decreasing shape of the locus which corresponds to the unit root economy. In this case the first effect mentioned above — the effect of changing persistence on the motive to self-insure — is entirely absent. One would expect, therefore, to see variability in consumption growth decline over the life cycle as an increasing share of income comes from the ownership and physical capital. This is precisely what Figure 7 shows.

The unit root economy turns out to be particularly informative for understanding savings behavior and self-insurance in our economy. It is unencumbered by dynamics in the idiosyncratic shock process, which makes for a simpler environment as well as an informative contrast with a known, analytical solution: the Constantinides and Duffie (1996) model. Our unit root economy has several key features which are critical in this regard. First, agents are net savers throughout their working years, a simple reflection of the fact that they are born with zero financial wealth and must provide for retirement consumption. This pattern of endowments is admittedly extreme, but it does capture the salient feature of the world that young people are less wealthy than old people and that most people save for retirement. Second, the marginal propensity to save out of income is large, relative to economies with lower persistence in idiosyncratic shocks. In simple terms, the more uncertain the future path of labor income is, the larger is the precautionary motive to accumulate financial wealth. Third, provided that financial wealth is positive, the marginal propensity to save out of current income is increasing in the level of current income but decreasing in the level of wealth. That is, holding wealth fixed, an agent will save proportionally more out of current income, the larger is the income shock. In contrast, holding income fixed, a poor agent will save proportionally less than a rich agent. The implication, therefore, is that in spite of being characterized by permanent shocks, our unit root economy displays a type of contingent, self-insurance behavior. What is critical is that this behavior be viewed in the context of the life cycle evolution of wealth. In other words, although an agent who receives a bad idiosyncratic shock will save — a decision which might seem to contradict 'self-insurance' behavior — they will save less, proportionally, than an agent who receives a good shock. Net of what is governed by the life cycle savings motive, therefore, this represents dissavings upon receiving a bad income shock, which is precisely what we mean by 'contingent, self-insurance behavior.'

Our model, then, provides an interesting contrast to the autarkic, unit root economy of Constantinides and Duffie (1996). Agents in our model change their financial behavior in response to a permanent shock, whereas a Constantinides-Duffie agent is content to do nothing. The critical difference — a point which has been made in a different context by Krusell and Smith (1997) — concerns the distribution of wealth. A necessary feature of the Constantinides and Duffie (1996) autarkic equilibria is that the distribution be degenerate across agents and identical to a particular level which is dictated by their specification for endowments. Consider, for example, an extreme version of their setup (which we adapt from an example in Krusell and Smith (1997)). The environment is an infinite horizon endowment economy with the only asset market being riskless borrowing and lending. Endowments are $y_{it} = exp(z_{it})$ and $z_{it} = z_{i,t-1} + \varepsilon_{it}$, where $\varepsilon_{it} \sim \mathrm{N}(-\sigma^2/2, \sigma^2)$. If the initial distribution of bonds is degenerate, with every agent holding zero bonds, then, with our class of preferences, an equilibrium is $c_{it} = y_{it}$ where the bond price is $\beta \exp(\alpha(1+\alpha)\sigma^2/2)$. The critical point is that this only works for this specific initial distribution of bonds. Should some agents start with a short position, while others start with a long position, the economy will converge to the autarkic steady state, but interim savings behavior will (we conjecture) display something akin to the 'risk sharing' behavior outlined above. The link to our OLG framework is that, because lives are finite and savings behavior has a strong life cycle component, any autarkic steady state, which may be important in an infinite horizon framework, is essentially irrelevant.

4.2.2 Wealth Accumulation

The flip side of the consumption decision is, of course, the savings decision. Probably the most immediate impact of idiosyncratic risk in our economy, an effect studied at length by many previous authors (*e.g.*, Aiyagari (1994), Huggett (1996)), is its effect on the motive to save and accumulate capital. In Figure 8 we report average financial wealth — capital plus bonds — for our baseline economy as well as the complete market and unit root economies which have proven useful contrasts to this point. The complete market economy shows a fairly standard life cycle savings pattern. Because average labor market productivity, and therefore the average wage rate, increases until just before retirement (recall that this is a feature of our calibration), the desire to smooth consumption over the life cycle results in an agent holding negative financial wealth during their early working years. This effect is quickly offset by the need to smooth consumption over the retirement years as well, and the average agent is a net saver by age 39.

Imperfect sharing of idiosyncratic risk changes all this. For young agents, the desire to accumulate a precautionary buffer of assets works against the standard life cycle, consumption smoothing motive. Young agents are far more unlikely to increase consumption by borrowing against wage income to be earned during their high productivity years. This is seen most strongly in the unit root economy, where the average agent holds positive wealth even during their first working year of life. The effect is mitigated slightly for our baseline case, where average wealth is still zero or negative for the first 6 years of life.

Idiosyncratic risk also has an appreciable effect on the overall level of savings and capital accumulation. As is apparent in Figure 8, the buffer stock savings motive drives up individual, and therefore aggregate, savings. In terms of asset pricing, the net effect is by now well known in the literature; the precautionary demand for assets drives down the level of interest rates and the expected rates of return on risky assets. In our production economy this takes the form of a higher level of aggregate capital leading to a lower marginal product of capital and therefore a lower average return to holding capital.

4.2.3 Portfolio Behavior and Asset Prices

We've seen that by adding idiosyncratic risk to an agent's income the life cycle savings motive is swamped by a precautionary savings motive. We now investigate how individuals choose to allocate these savings between risky and riskless assets and the implications for asset prices.

Figures 9 and 10 characterize life cycle portfolio behavior in our economy. In Figure 9 we characterize the portfolio choice of the median agent of each age cohort, as well as the fraction of agents who are constrained in each of the two asset markets. The latter is a loose guide to how representative the behavior of the median agent is for the entire distribution. In Figure 10, we plot a typical set of theoretical portfolio decision rules, the understanding of which is critical for understanding the behavior displayed in Figure 9.

Starting with Figure 9, the salient features of portfolio choice are as follows. First, almost all the young agents — those between age 23 (newborn) and, roughly, age 30 — hold all their wealth in the form of bonds. Most of these agents would actually like to hold a levered position in bonds, but cannot because of the stock market short-selling constraint. Second, almost all the old agents — essentially those aged 60 and above — are constrained in the bond market and hold a levered position in the stock market. Finally, most of the agents who are at an interior position in both asset markets belong to intermediate age cohorts: the 'prime age workers.' It is the behavior of these agents, therefore, that is critical for understanding the equity premium in our environment.

These patterns of portfolio choice mimic several salient features of U.S. data which seem important for a story linking the life cycle, idiosyncratic risk, and asset pricing. Our economy features a substantial fraction of agents who choose not to hold stock, one of the striking features of U.S. household portfolio composition (*c.f.*, Mankiw and Zeldes (1991)). As we'll argue further, this serves to 'concentrate the aggregate risk' and increases the premium paid to those who hold it. Our economy also has the feature that the lion's share of the stock market is owned by older agents, agents who face little, if any, idiosyncratic risk. While this stands in stark contrast to 'folklore' portfolio theory — the notion that stockholding should decrease with age — it is consistent with some aspects of U.S. data. Heaton and Lucas (1998), for instance, find that the households which hold the most substantial fraction of their wealth as stocks are both rich and elderly, a striking characteristic of our economy. They also find evidence suggesting that a life cycle pattern in idiosyncratic risk — middle aged households having relatively more exposure to risks associated with proprietary business ownership — is important for portfolio choice, another key aspect of our model. While Heaton and Lucas's notion of idiosyncratic income risk is clearly richer than ours — the distinction between labor market and privately owned business risk plays a key role in their study — we still think the comparison is relevant. Idiosyncratic income risk has, almost by definition, an important life cycle aspect to it. This 'life cycle of idiosyncratic risk' manifests itself in portfolio composition — both in the data and in our model — and must therefore be relevant for the determination of prices. This simple relationship lies at the heart of our asset pricing model.

There are three critical forces governing the portfolio choices represented in Figure 9. The first is the composition of an investor's wealth which, as we've seen in Figure 8, exhibits a strong life cycle pattern. Newborn agents begin life with zero financial assets: a stark abstraction but one which captures the fact that younger households hold the lion's share of their wealth in terms of human capital (claims to future wage receipts, in our case). As they age they must save, both for precautionary and retirement related motives. The fraction of their total wealth attributable to human capital is therefore monotonically decreasing, a fact of life which has been emphasized by many previous authors in studies of life cycle portfolio behavior (*c.f.*, Bodie, Merton, and Samuelson (1992), Jagannathan and Kocherlakota (1996) and references therein). In our setup, and in U.S. data, the returns to human and physical capital — wages and 'dividends,' respectively — are positively correlated, with the variability of the wages exceeding that of dividends. Agents with the greatest degree of wage risk will therefore seek to short-sell physical capital and invest the proceeds in bonds. Since short-selling is prohibited, the outcome we observe, which is displayed in Figure 9, is that young agents hold all their wealth in bonds and then gradually shift into stocks as they age.

Our economy therefore has the simple feature that stock holding is increasing in wealth, a feature shared by most related papers (see, for instance, Krusell and Smith (1997), Figure

3). This behavior is documented in Figure 10, where we graph the decision rules which govern portfolio choice for a newborn agent. The graph shows that when an agent is relatively poor, they choose to hold only bonds. For a small range of intermediate wealth levels both bonds and stock are held, and for a relatively high wealth level a levered position in stock is chosen. As an agent ages, the shapes of the decision rules change somewhat. Most importantly, the range of wealth levels associated with an interior choice in both markets increases, a manifestation of mature workers — those who face the most aggregate wage risk — seeking a compromise between higher returns and diversification against aggregate shocks.

The second force impinging on Figure 9 relates to intergenerational heterogeneity and how agents of different ages face different amounts of aggregate uncertainty. To best understand this, the complete markets allocation — an allocation which efficiently distributes aggregate risk across generations — is quite informative. Presuming that idiosyncratic risk is eliminated, perhaps via a centralized financial intermediary, a complete markets allocation can be supported by unconstrained trade in two assets: stocks and bonds. In this case we find portfolio behavior to be qualitatively similar to the incomplete markets case represented in Figure 9: the young hold bonds issued by the old. The reason is, again, related to the composition of wealth and the income that it generates. Young agents derive most of their income from the labor market. Since wage income is more volatile than capital income in our economy (labor's share of aggregate income is larger), a complete markets allocation features young agents short-selling stock and investing the proceeds in bonds. Retired agents, on the other hand, hold levered positions in stock, effectively assuming more aggregate risk than they would face otherwise. A small proportion of older working agents hold positive amounts of both stocks and bonds, something akin to the common notion of a diversified portfolio. The net result is a redistribution of aggregate risk from those who are endowed with the most of it — the young — to those who are endowed with the least — the old. Interestingly, the life cycle pattern in wealth is crucial for this outcome. If, hypothetically, we consider a young worker and a prime age worker agent with equal wealth, then the complete markets allocation will feature the younger worker holding *more* stock than the older worker, the reason being that labor market productivity is increasing with age.

The degree of exposure to aggregate shocks, therefore, works in the same direction as the composition of wealth: it makes older workers better equipped in terms of bearing aggregate risk. A third factor which is important for Figure 9, one which tends to strengthen these effects, turns out to be the lynchpin for our asset pricing results: countercyclically heteroskedastic income shocks (CCV). Both Constantinides and Duffie (1996) and Mankiw (1986) have persuasively shown that agents will demand a premium to hold stocks which pay relatively little when idiosyncratic shock volatility increases. More precisely, the conditional risk premium — the familiar covariance term which is the bread and butter of asset pricing theory — involves the covariance of stock returns with the (time varying) cross sectional variation in the consumption distribution. In our baseline economy this covariance is strongly negative. The average of the correlation between the variance of the *innovation* in consumption and the risky asset return, across age all cohorts, is -0.42. The average across working agents is -0.89 and the average across retirees is 0.13.

The CCV which we impart into income, therefore, becomes manifest in consumption. The crucial link in terms of Figure 9 — as well as the determination of prices — involves which group of agents are effectively 'pricing' the excess return portfolio and how much of the CCV risk they bear in equilibrium. Figure 9 shows that these agents are the prime age workers: those, roughly, between ages 35 and 60. These agents face less CCV risk than the very young, owing to the fact that they derive a substantial fraction of their income from capital. In this sense they are 'better equipped' to bear CCV risk, which results in the young choosing to avoid stock altogether. On the other hand, prime age workers are the most productive, something which aggravates their attitude towards CCV risk. In any case, it is the prime age workers who hold the interior portfolio in our economy, and therefore 'set the price.' These workers bear an appreciable amount of the aggregate, CCV risk — some of which they implicitly take off of the hands of younger workers — and demand to be compensated for doing so. We find that this effect is the central one in terms of driving risk premia in our economy. Were the equilibrium characterized, for instance, by the old holding all the stock, CCV risk would not be 'priced' in equilibrium and our economy would generate much smaller Sharpe ratios.

This completes the discussion of the economics behind portfolio choice. Good computational economics, however, should verify economic interpretations with supporting experiments. A natural question in this regard asks what happens to equilibrium portfolio behavior and asset returns if we alter the magnitude of the conditional variation (the CCV) in the income shocks. Are the CCV shocks important for portfolio choice, pricing, or both? We find that the answer is 'pricing,' as long as stocks don't represent a *hedge* against increases in conditional variance. More specifically, in a homoskedastic economy (represented by the unity point, CCV=1, in Figure 5), portfolio choices do not change, qualitatively, relative to the baseline case of CCV=2.2. That is, young agents still avoid stock by lending to the old, who hold a levered position in equity. What does change is that the reward paid to those who do bear the stock market risk drops by a factor of 8 (see Figure 5). In contrast, in an economy with *procyclical* cross sectional variation — one in which stocks pay off more when cross sectional variation increases — young agents view stocks as a hedge against increases in idiosyncratic risk and choose to hold all their wealth as stock. One would expect this to lead to a further reduction in the excess stock market return, as retirees become the marginal agents (holding all the bonds), but do not face any CCV risk. This is confirmed by Figure 5. These experiments suggest, therefore, that our interpretation of the baseline case is valid. The three effects we've described all work in the same direction; they tend to make the young hold bonds and the old hold stock. Removing one effect — the CCV — does not change this, but it does change pricing in a predictable manner. Altering the sign of the CCV risk, on the other hand, does change the 'comparative advantage' of the young in terms of bearing aggregate risk and, if large enough in magnitude, changes portfolio behavior in a manner which is consistent with our analysis.

To summarize, the equity premium in our economy is driven by a subset of the working population being averse to the poor hedge which stocks provide against times when idiosyncratic risk increases. Relative to the working population, these agents bear a disproportionate share of society's aggregate risk. They don't do this, of course, out of benevolence, but because they are better equipped to bear aggregate risk and because they receive compensation for doing so. This comparative advantage in risk bearing is driven by an inescapable aspect of labor market risk: individuals have less of it the older they are. In this sense our asset pricing theory is driven by a natural link between a life cycle pattern in idiosyncratic risk and a life cycle pattern in portfolio choice.

5 Interpretation and Alternative Parameterizations

Our baseline economy delivers Sharpe ratios which, while being significantly larger than those of related studies, remain deficient when compared with U.S. data. A natural question, then, is 'what does one have to do, in our framework, to generate risk premia of a realistic magnitude?' We address this question through several alternative parameterizations. We also provide quantitative results from a calibrated version of the Constantinides and Duffie (1996) economy, which we find to be an informative benchmark in terms of what is driving our results. Additional sensitivity analyses, designed primarily to assess the robustness of our computational methods, are discussed in appendix B.

Figures 5, 7 and 8 have already provided information on how our economy behaves under different values for the parameter ρ, the autocorrelation of the idiosyncratic shocks. We see that increases in ρ tend to increase savings, change the life cycle pattern in the variability of consumption growth, and increase the Sharpe ratio on the risky asset. The latter is particularly important in understanding how our results differ from those of others. Heaton and Lucas (1996), for instance, find the asset pricing effects of large increases in CCV to be small. Our results are consistent in the sense that relatively low autocorrelation — the $\rho = 0.5$ locus in Figure 5 corresponds to their assumed value — leads to small Sharpe ratios, irrespective of the magnitude of CCV.

In Exhibit 1 we demonstrate the effect of increasing the risk aversion coefficient, α, from 2 to 4.5. The latter is chosen because it generates a Sharpe ratio which is close to the U.S. sample moment of 0.41. Exhibit 1 also reports comparable values for the Constantinides and Duffie (1996) model and — to provide a complete markets benchmark — the Mehra and Prescott (1985) model. In both cases the models are calibrated so as to generate a meaningful comparison; details are provided in appendix C. The economies associated with our model feature unit root idiosyncratic shocks, something which is motivated by both computational considerations (unit root economies are less costly to compute) and a desire to provide a close connection to the Constantinides-Duffie framework (which is distinguished by unit root shocks).

Exhibit 1: Effects of Higher Risk Aversion

Risk Aversion	CCV	*Sharpe ratios in percent* MP	CD	STY
$\alpha = 2$	homoskedastic	3.04	3.04	0.88
	120% increase	–	11.39	16.10
$\alpha = 4.5$	homoskedastic	5.55	5.55	3.01
	120% increase	–	33.60	39.02

Entries correspond to Sharpe ratios, expressed as percentages, associated with the Mehra and Prescott (1985) (MP) model, the Constantinides and Duffie (1996) (CD) model, and our model (STY), with unit root idiosyncratic shocks. In each case the rate of time preference is chosen so as to match the expected return on the risky asset. The particular value for the risk aversion coefficient we report, $\alpha = 4.5$ is chosen so that our economy generates a Sharpe ratio which is approximately equal to that observed in U.S. data over a long horizon, 41 percent. More explicit details of how we calibrate each model are provided in appendix C.

Exhibit 1 makes several important points. First, an increase in risk aversion has a substantially greater impact upon the price of risk in the heterogeneous agent economies than in the representative agent economy. In the Mehra-Prescott environment, a 125% increase in the value of risk aversion coefficient increases the Sharpe ratio by just over 80%, whereas the increase is 140% and almost 200% for our economy and the Constantinides-Duffie economy, respectively. This disproportionate effect is, in fact, strongly suggested by the form of the Constantinides-Duffie representative agent. In their model, the mapping between α, the risk aversion coefficient of actual agents, and $\hat{\alpha}$, the risk aversion coefficient of the representative agent, is:

$$\hat{\alpha} = \alpha - \frac{\alpha(\alpha+1)}{2} b \ ,$$

where b is a parameter which relates to the amount of CCV inherent in the economy (see appendix C for details). For values of b less than zero — which *defines* what we mean by CCV — increases in the 'effective' risk aversion coefficient, $\hat{\alpha}$, are increasing in the level of α. For instance, a value of $\alpha = 2$ corresponds to a value for $\hat{\alpha}$ of 11.4, whereas $\alpha = 4.5$ generates $\hat{\alpha} = 43.2$. The fact that our economy shares this feature — increases in risk aversion having disproportionately large effects on risk premia — is encouraging. It serves as one check on our computational solution and, just as importantly, suggests that our economy can generate sizable risk premia (per unit of volatility) while maintaining moderate attitudes towards risk amongst its inhabitants.

A second important feature of Exhibit 1 relates to the absolute magnitudes of the Sharpe ratios. For a comparable amount of heterogeneity and CCV, our economy generates larger Sharpe ratios than the Constantinides-Duffie model, and at the same time does so in what is certainly a less extreme environment (*e.g.*, our economy features non-degenerate trade in asset markets and, as we argue in Storesletten, Telmer, and Yaron (1997), a realistic degree of imperfect risk sharing). We attribute the differences to the distinguishing characteristic of model, the life cycle. In an admittedly loose sense, our interpretation is as follows.

An important characteristic of the allocations in our model is partial risk sharing; agents are able to use asset markets to eliminate roughly 1/3 of their labor market risk, even in the presence of unit root shocks. *Ceteris paribus*, this will reduce risk premia in our model, relative to those in the autarkic Constantinides-Duffie environment. On the other hand, we've argued above that risk premia are likely to be driven up by life cycle considerations, in particular the fact that the agents who are most exposed to CCV risk — the prime age workers — must hold a disproportionate share of it. The fact that Sharpe ratios in our model exceed those of our Constantinides-Duffie calibration suggests that this second effect dominates the first. In other words, the positive effect of the 'life cycle of idiosyncratic risk' on risk premia outweighs the negative effect of partial risk sharing. In this sense we interpret the OLG aspect of our economy as being a quantitatively important ingredient in understanding the interaction of idiosyncratic risk and asset pricing.

One final alternative we have investigated involves a large increase in aggregate volatility. As we've argued above, a key restriction of a model with our type of aggregate production technology — a restriction which is absent in an endowment economy — is a strong link between the volatility of aggregate quantities and the volatility of risky asset returns. Risky asset returns are equal to the marginal product of aggregate capital. Their volatility, therefore, is driven by variation in technology shocks and variation in aggregate capital. Aggregate capital is, by its very nature, a low volatility process. Volatility in technology shocks is strongly restricted by volatility in aggregate output, consumption, investment, and so on. Models such as ours, therefore, cannot generate high volatility in asset returns without generating unrealistically high volatility in aggregate quantities.

In order to get a feel for the impact of higher return volatility on risk premia, our final experiment (results not reported) ignores restrictions on aggregates. Specifically, we increase the volatility of the technology shock process, Z, until the standard deviation of the risky asset return matches that of the S&P 500 (roughly 16 percent per annum). We find that the excess return on the risky asset is roughly 4 percent per annum, with a Sharpe ratio of 25 percent. This compares somewhat favorably with a similar exercise conducted in a representative agent context. Using our calibration of the Mehra-Prescott economy, but where variation in the endowment process is chosen to match the standard deviation of the S&P 500, we compute an equity premium of 3.2 percent and a Sharpe ratio of 15 percent. The overall implication is that the form of heterogeneity we study is incrementally helpful for understanding asset returns, even if the 'puzzle' of excessive volatility in stock returns is abstracted from.

6 Conclusions

Our line of inquiry is motivated by the supposition that individual households face a number of important sources of idiosyncratic risk, that they are unable (or unwilling) to eliminate them from their consumption choices, and that these risks have an important effect on their savings and investment behavior. This is a statement which, we think, is fairly uncontroversial in that it is supported by a large empirical literature, a strong body of economic theory (*e.g.*, models which study agency relationships) as well as casual introspection and plain common sense. The asset pricing literature is concerned with a somewhat deeper issue: whether or not these idiosyncratic sources of income and consumption variation have anything to do with observed levels of interest rates and return differentials between risky assets. The challenge then, is to formalize our supposition in a model and examine the pricing implications.

A focal point of our paper is the notion that the OLG formalization is an attractive alternative to what has become more commonplace in the literature: the infinite horizon model, with or without dynastic mortality risk. To understand this, it is useful to consider recent experiences with infinite horizon models. Our reading of the literature suggests that the intuition one gets from the 'permanent income hypothesis' carries forth to varying degrees in an equilibrium model where savings and investment behavior are disciplined by both the (implicit) need to find another agent to trade with, and the changing nature of the price process which governs this trade. In other words, the nature of competitive equilibrium in these models is that agents consume permanent shocks and diversify away transitory ones. Any departures from this statement are almost exclusively driven by frictions of one form or another, such as portfolio constraints and transactions costs.

This leads one to the thought that some combination of 'permanent shocks' and frictions must characterize a model in which idiosyncratic risk is to have an appreciable impact. The OLG framework is attractive in this sense, in that (a) permanent shocks are easily modeled in a tractable, recursive framework (indeed, all shocks are 'permanent' to some degree), and (b) a number of natural constraints on savings and investment behavior arise as a result of both finiteness and life cycle patterns in wealth and income. An obvious example, one which is starkly presented in our model, is that the young are relatively asset-poor and are unlikely to be able to self-insure until a buffer stock has been accumulated. In addition, the 'constraint' that one must typically save for retirement mitigates the extent to which any given buffer of assets is useful for eliminating even transitory sources of income variation. We find the summation of these effects to be quantitatively important. Our model has little difficulty satisfying a necessary condition that has hampered previous work, the condition that allocations be sufficiently far from first-best.

That the OLG model may generate imperfect risk sharing outcomes in a plausible way is one thing. In addition, the framework seems well suited for asset pricing questions which, in our class of models, hinge on the dependence between idiosyncratic and aggregate sources of variation. This dependence cannot be in the first moment — having a mean of zero is what defines the term 'idiosyncratic' — so we follow Mankiw (1986) and Constantinides and Duffie (1996) in incorporating it in the second moment, something we label 'countercyclical cross sectional variation,' or 'CCV.' What drives our results turns out to be the manner in which CCV interacts with the life cycle. Because we model idiosyncratic risk as being persistent and as arising in the labor market, a distinguishing feature of our economy is something which seems quite natural: young agents face more idiosyncratic risk than old agents. Age, in other words, affects the way in which an agent views the risk which the CCV dependence represents. This in turn affects portfolio choice — the young choose to hold only the riskless asset — and forces a greater than per-capita share of the aggregate risk onto risky asset holders. The latter, in our equilibrium, turn out to be high productivity working agents who dislike CCV risk, demand a premium in order to bear it, and thereby drive up the excess return paid to equity. It is in this sense that the life cycle in our model is an integral part of the mapping between idiosyncratic risk and asset returns. A *natural* life cycle pattern in idiosyncratic risk leads to a life cycle pattern in portfolio choice which leads to a life cycle pattern in the characteristics of the group of investors who are relevant for price determination. Again, we find the quantitative implications to be substantial. Our economy generates far greater Sharpe ratios than comparable, production-based models, and can match the Sharpe ratio on the U.S. stock market with only a moderate increase in risk aversion, relative to conventional values.

The importance of life cycle portfolio behavior for asset prices has also been emphasized in recent work by Constantinides, Donaldson, and Mehra (1997) (CDM). They stress the impact of young agents not holding any stock and of the economy's aggregate risk, therefore, being concentrated on older agents. The same feature is important in our model, but for fundamentally different reasons. Aside from a number of important quantitative issues, the main characteristic which distinguishes our approach is idiosyncratic risk *within* generations, something which we've argued is fundamental for our results. The implications for individual behavior — and therefore for testable restrictions — are strong. In our model, the decision of a young agent to avoid equity is very much a portfolio allocation decision: equity is too risky, so they choose not to hold any. In the CDM framework, at least the version of their model which generates sizable equity premia, the driving force is consumption smoothing and how it interacts with borrowing constraints. Young agents receive a relatively meager endowment, cannot borrow or short sell equity, and therefore choose not to hold any assets whatsoever. These are starkly different interpretations of why one might see a young household choose not to hold equity. The testable restrictions are related to overall savings behavior and how important the precautionary motive is. Our framework is consistent with the *average* young household (poor young households in our model do face binding borrowing constraints) being a net saver during the first third of their lives. That is, the precautionary motive dominates the life cycle motive, and the decision to avoid equity is driven by *risk*, in our case CCV risk. The CDM framework is consistent with the same average, young household not accumulating any assets and, in contrast, viewing equity (in a shadow value sense) as an attractive investment. Which of these interpretations is more important — it seems clear to us that the world features aspects of each of them — is something we leave to future work.

We choose to address quantitative issues in our model by exploiting information from the PSID. A reasonable criticism is that the PSID does not contain a representative share of stock-holding households. Heaton and Lucas (1998) implicitly make this point in their study of household portfolio behavior based on a wealthier set of households. Our interest in the PSID, in spite of this limitation, is best understood as follows. What we are primarily interested in are low-frequency movements in asset returns and their relationship to the real economy. The crux of our approach is that, in addition to aggregate variables, cross sectional variables play an important role this relationship. The manner in which the cross sectional distribution matters can be interpreted in (at least) two ways. One interpretation is that, in order to make decisions, each agent requires information on every other agent's wealth, labor market status, and so on. An alternative interpretation, one which we prefer, is that a lower dimensional representation exists. That is, what matters for forecasting future prices is not the entire distribution, but just a small set of its characteristics. This is the sense in which Krusell and Smith (1998), Telmer and Zin (1995) and others refer to a model as admitting an 'approximate aggregation.' Given this interpretation — that our interest in the cross section remains as a *macroeconomic* variable — our feeling is that there is no better dataset than the PSID for measuring several of its simple properties, most specifically how its variance changes over the business cycle. Moreover, our feeling is that the simple statistics we obtain from the PSID, most importantly autocorrelation and CCV, are likely to move in our favor were we able to obtain a more representative dataset. For instance, our guess is that 'entrepreneurial risk' (to borrow a phrase from Heaton and Lucas (1998) and Polkovnichenko (1999)) is likely to be even more persistent than the labor market risk inherent in the PSID, in addition to having rich cyclical dynamics.

Many previous papers on idiosyncratic risk and asset pricing have contributed by providing negative results: demonstrating what cannot solve the equity premium puzzle. Our results are encouraging in that they are positive in nature, suggesting several factors which may be quantitatively important in terms of resolving the puzzle. Plausible parameterizations of our model, both in terms of preferences and PSID data, can generate Sharpe ratios which match U.S. data, even in the absence of frictions such as transactions costs and the like. While positive results are, in a sense, more constructive, they are also more easily challenged. The most glaring weakness in our set-up is simple: volatility. Our model clearly stands no chance of accounting for the *level* of the equity premium (as opposed to the Sharpe ratio) because risky asset returns are pinned down by our aggregate technology, which is in turn calibrated to macroeconomic information on the U.S. capital stock. In other words, we are unable to address the issue of why stock returns are so variable relative to returns on physical investment. First on our list of alternative models, chosen to address this issue, are models with adjustment costs to physical capital or irreversibilities. In our current model investment in completely reversible, implying that the price of capital is always unity. Adjustment costs would drive the price from unity and result in a model in which the word 'stock' is more closely associated with the common stocks underlying the U.S. data we attempt to explain.

References

Abowd, J. and D. Card, (1989), On the covariance structure of earnings and hours changes, *Econometrica* 57, 411–455.

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

Aiyagari, S. R. and M. Gertler, (1991), Asset returns with transactions costs and uninsured individual risk, *Journal of Monetary Economics* 27, 311–331.

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

Altug, S. and R. A. Miller, (1990), Households choices in equilibrium, *Econometrica* 58, 543–70.

Alvarez, F. and U. Jermann, (1997), Asset pricing when risk sharing is limited by default, Unpublished working paper, University of Chicago.

Arrow, K., (1970), *Essays in the Theory of Risk Bearing*, North-Holland, London.

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

Bodie, Z., R. C. Merton, and W. F. Samuelson, (1992), Labor supply flexibility and portfolio choice in a life cycle model, *Journal of Economic Dynamics and Control* 16, 427–49.

Castañeda, J., J. Díaz-Giménez, and J. V. Ríos-Rull, (1994), Unemployment dynamics, aggregate fluctuations, and income distribution, Working paper, University of Pennsylvania.

Constantinides, G. M., J. B. Donaldson, and R. Mehra, (1997), Junior can't borrow: a new perspective on the equity premium puzzle, Working paper, University of Chicago, GSB.

Constantinides, G. M. and D. Duffie, (1996), Asset pricing with heterogeneous consumers, *Journal of Political Economy* 104, 219–240.

Cooley, T. F. and E. C. Prescott, (1995), *Economic Growth and Business Cycles*, chapter 1, 1–38, Princeton University Press, New Jersey, In *Frontiers of Business Cycle Research*, T. Cooley, editor.

Deaton, A., (1991), Savings and liquidity constraints, *Econometrica* 59, 1221–1248.

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

den Haan, W., (1994), Heterogeneity, aggregate uncertainty and the short term interest rate: a case study of two solution techniques, Working paper, University of California at San Diego.

den Haan, W., (1997), Solving dynamic models with aggregate shocks and heterogeneous agents, *Macroeconomic Dynamics* 1, 355–386.

den Haan, W. and A. Marcet, (1994), Accuracy in simulations, *Review of Economic Studies* 61, 3–17.

Gordon, R. J., ed., (1986), *The American Business Cycle: Continuity and Change*, University of Chicago Press.

Gottschalk, P. and R. Moffit, (1992), Earnings and wage distributions in the NLS, CPS and PSID, Manuscript, Brown University.

Heaton, J. and D. J. Lucas, (1996), Evaluating the effects of incomplete markets on risk sharing and asset pricing, *Journal of Political Economy* 104, 443–487.

Heaton, J. and D. J. Lucas, (1998), Portfolio choice and asset prices; the importance of entrepreneurial risk, Working paper, Northwestern University.

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.

Jagannathan, R. and N. Kocherlakota, (1996), Why should older people invest less in stocks than younger people?, *Federal Reserve Bank of Minneapolis, Quarterly Review* 20, 11–23.

Kreps, D., (1979), Three essays on capital markets, Technical Report 298, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Krusell, P. and A. A. Smith, (1997), Income and wealth heterogeneity, portfolio choice, and equilibrium asset returns, *Macroeconomic Dynamics* 1, 387–422.

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

Lucas, D. J., (1994), Asset pricing with undiversifiable risk and short sales constraints: Deepening the equity premium puzzle, *Journal of Monetary Economics* 34, 325–341.

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

Mankiw, N. and S. Zeldes, (1991), The consumption of stockholders and non-stockholders, *Journal of Financial Economics* 29, 97–112.

Mankiw, N. G., (1986), The equity premium and the concentration of aggregate shocks, *Journal of Financial Economics* 17, 211–219.

Marcet, A. and K. J. Singleton, (1991), Equilibrium assets prices and savings of heterogeneous agents in the presence of portfolio constraints, forthcoming, *Macroeconomic Dynamics.*

Mehra, R. and E. Prescott, (1985), The equity puzzle, *Journal of Monetary Economics* 15, 145–61.

Mrkaic, M., (1997), Estimation of incomplete markets models with panel data – the case of the permanent income hypothesis, Working Paper, Carnegie Mellon University.

Polkovnichenko, V., (1999), Heterogeneity and proprietary income risk: implications for stock market participation and asset prices, Manuscript, Northwestern University.

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

Rouwenhorst, K. G., (1995), *Asset pricing implications of equilibrium business cycle models*, chapter 10, 294–330, Princeton University Press, New Jersey, In *Frontiers of Business Cycle Research*, T. Cooley, editor.

Storesletten, K., (1999), Sustaining fiscal policy through immigration, forthcoming, *Journal of Political Economy.*

Storesletten, K., C. I. Telmer, and A. Yaron, (1997), Consumption and risk sharing over the life cycle, GSIA working paper no. 1997-E228, Carnegie Mellon University.

Telmer, C. I., (1993), Asset pricing puzzles and incomplete markets, *Journal of Finance* 48, 1803–1832.

Telmer, C. I. and S. E. Zin, (1995), Approximate aggregation with incomplete markets, GSIA working paper no. 1994-E5.

Weil, P., (1992), Equilibrium asset prices with undiversifiable labor income risk, *Journal of Economic Dynamics and Control* 16, 769–790.

White, H., (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity, *Econometrica* 48, 817–38.

Zhang, H., (1997), Endogenous borrowing constraints with incomplete markets, forthcoming, *Journal of Finance.*

A Data Appendix

Our data sources are 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:

- The head of the household is male.
- There are no changes in family structure except for the number of children.
- 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.

The net result is a sequence of panels in which average age is quite stable — the mean and standard deviation, across the panels, of average age is 44.2 and 1.1, respectively — and the number of households per panel is quite large. The average number of households, across the panels, is 2,045 and the standard deviation is 228. This compares with the 610 households who populate the associated longitudinal panel (see below).

Aggregate Data

Aggregate GNP data, used to construct the indicator functions in equation (12), are an amalgamation of annual data from Gordon (1986) for the years 1910-1958 and annual data from CITIBASE for the years 1959-1992. The data in Gordon (1986) are reported as 1972 dollars. The CITIBASE data were deflated using the CPI deflator and were converted to per-capita values using CITIBASE data on the non-military U.S. population.

Longitudinal Panel

For comparison's sake we also constructed a 24 year longitudinal panel using the same selection criteria as above. The result is a panel on 610 households where the average age is roughly 39 in 1968 and, therefore, 62 in 1991. The standard deviation of age in each annual cross section is 10.54.

When we apply the methodology of Heaton and Lucas (1996) — estimating household-specific parameters and averaging across the estimates — we replicate the qualitative features of their results. The estimate of the autocorrelation parameter, ρ, is 0.64 and 0.50 using their sample and our sample, respectively. When we use our methodology we obtain an estimate of 0.931, which is essentially identical to that from the overlapping panels in Table 1. The differences in our approaches are as follows. We detrend the data somewhat differently (including life cycle and education effects). The parameters in our specification are constrained to be the same across all households. The moment conditions underlying our estimates condition on household age, something which we find can have substantial effects. Finally, we do not explicitly allow for household-specific fixed-effects as they do, by estimating an intercept parameter, per-household. This last item turns out to be particularly important. We find that taking out a household specific mean, or 'fixed effect,' reduces our methodology's estimate of ρ to 0.75.

Individual fixed effects, then, can have an important effect on one's view of persistence. Our feeling is that there is no easy answer here. Our approach economizes on parameters (involving 610 fewer) by modeling deterministic cross sectional variation as being related to education levels alone. The cost is the strong likelihood of mistaking shocks for deterministic (at birth) differences. One common approach is to eliminate the fixed effects by differencing or quasi-differencing the data (*i.e.*, basing moment conditions on, respectively, $u_{i,t+1} - u_{it}$ or $u_{i,t+1} - \rho u_{it}$). We find that our inference of high persistence is relatively robust in this sense. Based on the longitudinal panel, our estimate of ρ is never less than 0.83 and in most cases is roughly 0.90 (depending on the particular specification and sample).

A.1 Estimation

Our estimation procedure has two distinct steps. In the first stage we estimate equation (7). In the second stage we estimate the system given in (12).

Let the parameters of the first stage (the coefficients of the year dummies, age, age squared, and education) be summarized by θ_1. Also, let the parameters of the system in (12) be denoted by θ_2 (that is σ_H, σ_L, σ_ϵ, ρ). The joint system we estimate can be written compactly as

$$E\left[\begin{array}{c}\psi_1(y_{it}^h,\theta_1)\\ \psi_2(y_{it}^h,Y_t,\theta_1,\theta_2)\end{array}\right]=0 \tag{13}$$

where ψ_1 and ψ_2 are the moment conditions corresponding to (7) and (12) respectively. The triangular structure of the moment condition allows us to get consistent estimates of θ_1 using only ψ_1. We then estimate θ_2 using moment conditions ψ_2. This second step incorporates the standard errors in estimating θ_1 using a standard two-step GMM procedure. The additional complication that arises in our set-up is due to the overlapping structure of our repeated panels. Since these panels overlap each other by 2 years an MA(2) correction is added to the estimate of the covariance matrix associated with moment conditions ψ_2.

Specifically define the moment conditions ψ_2 to be

$$\begin{aligned}
\psi_{2,i,t}^{h,0} &\equiv [(u_{i,t}^{h,t})^2 - E(\sigma_\varepsilon^2 + \sum_{j=0}^{h-1}\rho^{2j}(I_{t-j}\sigma_H^2 + [1-I_{t-j}]\sigma_L^2))]\\
\psi_{2,i,t}^{h,1} &\equiv [u_{i,t}^{h,t}u_{i,t+1}^{h+1,t} - E(\rho\sum_{j=1}^{h-1}\rho^{2(j-1)}(I_{t-j}\sigma_H^2 + [1-I_{t-j}]\sigma_L^2))]\\
\psi_{2,i,t}^{h,2} &\equiv [u_{i,t}^{h,t}u_{i,t+2}^{h+2,t} - E(\rho^2\sum_{j=1}^{h-1}\rho^{2(j-2)}(I_{t-j}\sigma_H^2 + [1-I_{t-j}]\sigma_L^2))]
\end{aligned}$$

where the superscript t in u denotes the base year of the panel from which this agent is selected. By assumption $\psi_{2,i,t}^{h,j}$ is not correlated with $\psi_{2,i,t+k}^{h,l}$ $\forall k \neq 0$, and $j,l = 0,1,2$. It can be easily shown, however, that due to the overlap of the sample, for each t, $\psi_{2,i,t}^{h,0}$, $\psi_{2,i,t}^{h,1}$ and $\psi_{2,i,t}^{h,2}$ are correlated. We stack the repeated 3 moment conditions and use sample counterparts to estimate these covariance terms – where the covariance matrix is block-diagonal and each 3×3 block has non-empty off-diagonal elements.

Our results are robust to selecting fewer moment conditions, using subsets of the years above, and to a system in which the parameters of interest are exactly identified. For the exactly identified case, we also experimented with repeated panels of 4 years and the additional moment condition

$$\psi_{2,i,t}^{h,3} \equiv [u_{i,t}^{h,t}u_{i,t+3}^{h+2,t} - E(\rho^3\sum_{j=1}^{h-2}\rho^{2(j-3)}(I_{t-j}\sigma_H^2 + [1-I_{t-j}]\sigma_L^2))],$$

and then use the four sample counter-parts $\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}\psi_{2,i,t}^{h,j}$ for $j=0,1,2,3$, as the moment conditions.

B Computational Appendix

Our general solution strategy follows the work of den Haan (1994), den Haan (1997) and, in particular, Krusell and Smith (1997) and Krusell and Smith (1998). The crucial step is the specification of a finite dimensional vector to represent the law of motion for μ. Given this, each individual faces a finite horizon dynamic programming problem. The essence of the fixed point problem is the consistency of the law of motion for μ with the law of motion implied by individual decisions. More specifically, our algorithm involves the following steps.

1. Approximate the distribution of agents, μ, with a finite number of moments or statistics, μ_m. Following Krusell and Smith (1997), we began with just the first moment, aggregate capital, $\bar{k}$. Having found this to be inadequate for our problem, we added a second, the conditional expected equity premium, ξ. Krusell and Smith (1997), for comparison, use aggregate capital and the bond price.

   The conditional expected equity premium (CEEP) is defined as $\xi_t \equiv E_t\{R_{t+1}\} - q_t^{-1}$, where R_{t+1} is the return on equity in period $t+1$ and q_t is the period t price of a claim that pays one unit of the consumption good in period $t+1$. Note that, given ξ_t and conditional expectations over the future states of the world, the implicit bond price is $q_t = (E_t\{R_{t+1}\} - \xi_t)^{-1}$.

   Conceptually, μ is an infinite dimensional vector and can depend on almost any conceivable state variable. Our use of ξ (a "price") as a state variable may seem odd, but should be thought of as a convenient way to summarize the information which agents need for their savings and portfolio allocation decisions. The alternative and perhaps more direct route would have been to specify moments of bond holdings by age and wealth and possible higher order moments of capital. In this sense our two-moment formulation represents a relatively efficient use of state variables that must be part of an agent's information set.

2. Approximate the agents' expectations of the law of motion for μ'_m with a linear function of μ_m, Z and Z':

$$\begin{aligned} \log(\bar{k}') &= a_0(Z,Z') + a_1(Z,Z')\log(\bar{k}) + a_2(Z,Z')\xi \\ \xi' &= b_0(Z,Z') + b_1(Z,Z')\log(\bar{k}) + b_2(Z,Z')\xi \end{aligned} \tag{14}$$

   The aggregate shock Z can only take on two values, $Z \in \{\underline{Z}, \bar{Z}\}$, so each of the coefficients above can take on four different values. Assume a particular set of values for

   $\{a_0(Z,Z'), a_1(Z,Z'), a_2(Z,Z'), b_0(Z,Z'), b_1(Z,Z'), b_2(Z,Z')\}_{Z,Z' \in \{\underline{Z},\bar{Z}\}}$.

3. Given a set of expectations over μ'_m, now defined as $\mu'_m = \hat{G}(\mu_m, Z, Z')$, solve an appropriately modified version of (5):

$$\begin{aligned} \hat{V}_h(\mu_m, Z, z, \epsilon, a) &= \max_{b'_{h+1}, k'_{h+1}} \{u(c_h) + \\ & \beta \frac{\phi_{h+1}}{\phi_h} E\left[\hat{V}'_{h+1}(\hat{G}(\mu, Z, Z'), Z', z', \epsilon', k'_{h+1} R(\hat{G}(\mu, Z, Z'), Z') + b'_{h+1})\right]\Big\} \end{aligned} \tag{15}$$

subject to (4) and (14).

The implementation of this is described below.

4. Assume an initial distribution of a large, but finite, number of agents, μ, across wealth, idiosyncratic shocks and age (we use 400 agents in each age cohort). Using the decision rules obtained in (15), simulate a long sequence of the economy (2100 periods) and discard the first 100 periods from this sequence. Note that, for each period in time, the CEEP ξ must be set so that the bond market clears. That is, find a ξ^* such that

$$\int b'_h(\bar{k}, \xi^*, Z, z, \epsilon, a) d\mu = 0 .$$

This is the sense in which ξ is an 'endogenous moment.'

5. Update $\hat{G}$ by running a linear regression of μ'_m on μ_m from the realized sequence in Step 4. If the coefficients change, use the updated $\hat{G}$ and return to Step 3. Continue this process until convergence.

6. Evaluate the ability of $\hat{G}$ to forecast μ'_m. If the goodness of fit is not satisfactory, return to Step 1 and increase the number of moments or change the functional form of $\hat{G}$.

Dynamic Programming Problem

We now turn to how the dynamic programming problem in (15) is solved. We solve the decision rules backward, starting at the terminal age.

1. First, we choose a grid for the continuous variables in the state space. That is, we pick a set of values for $\bar{k}$, ξ, and a. The grid points are typically chosen to lie in the stationary region of the state variables and in addition, for wealth, near the borrowing constraint and far in excess of the maximum observed wealth holdings (conditional on age). We pick 5 points for aggregate capital, 5 points for the conditional expected equity premium, and 25 points for individual wealth at each age.

2. Second, we make piecewise linear approximations to the decision rules by solving for portfolio holdings on the grid and iterating on the Euler equations.

This is done in the following way. Given the terminal condition associated with (4) and the form of utility function, we know that the decision rules of the oldest agents (H years old) must be $b'_{H+1} = k'_{H+1} = 0$, in any state of the world. That is, the agent consumes all their wealth.

Knowing c_H, we can in turn compute b'_H and k'_H at each grid point using Euler equations of an $H-1$ year old agent (and imposing the borrowing constraints and Kuhn-Tucker conditions):

$$\begin{aligned} u'_{H-1}(c_{H-1}) &\geq E\{u'_H(c'_H)R' \mid \mu_m, Z, z, \epsilon\} \\ qu'_{H-1}(c_{H-1}) &\geq E\{u'_H(c'_H) \mid \mu_m, Z, z, \epsilon\} \end{aligned} \tag{16}$$

Knowing b'_H and k'_H at each grid point, we then make a piecewise linear approximation of the decision rules by linear interpolation (outside the grid we do linear extrapolation). Given this approximated decision rule, we can, for any state variable, compute c_{H-1}. Using this we can again compute b'_{H-1} and k'_{H-1} at each grid point by solving for the associated Euler equations, (16), for $H-2$ year old agents. This process is iterated backward until $h = 1$.

Note that no further iterations are needed; given the (imperfect) expectations $\hat{G}$ and the decision rules for $h+1$ years old agents, the piecewise approximations are found in one single step for h years old agents.

Accuracy

We now discuss accuracy issues concerning our solution method.

Euler equations

The solutions on the grid points are exact by construction. To evaluate whether the interpolation between grid points gives rise to systematic Euler equation pricing errors we follow den Haan and Marcet (1994) and use simulation to construct the following moment conditions:

$$g(c, Z, z, R, q) \equiv \frac{1}{T}\sum_{t=1}^{T}\sum_{h=1}^{H}\frac{1}{N_h^*}\sum_{i=1}^{N_h^*}\left[\left(\beta\left(\frac{c_{i,t+1}^{h+1}}{c_{i,t}^{h}}\right)^{-\alpha}\begin{bmatrix} R_{t+1} & \frac{1}{q_t}\end{bmatrix} - 1\right) \otimes z_{i,t}^h\right] \tag{17}$$

where T is the number of periods in the simulation, N_h^* is the number of *unconstrained* agents within age cohort h, and the instruments are $z_{i,t}^h = \{1, a_{i,t}^h, R_t\}$. The p-values corresponding to the χ^2 statistic based on moment conditions $g(\cdot)$ and the corresponding covariance matrix did not exceed .12 – indicating very small pricing errors.

Law of motion of μ

We now describe in more detail the $\hat{G}$ function which our algorithm converges to. It turns out that we can forecast μ'_m very well with only two moments – aggregate capital and CEEP. The coefficient of variation of the forecasting error is very small, less than 0.7% for ξ' and less than 0.2% for $\bar{k}$. The R^2 of equation (14) is more than 0.98 for forecasting $\bar{k}$ in either combination of $\{Z, Z'\}$, and more than 0.94 for forecasting ξ. Our simulations produce a sequence of realized moments of μ, which we denote $\hat{G}_{t+1}$. The difference between the realized moments for μ and the forecasting formula for these moments which agents use in their decisions, $\hat{G}_{m,t+1}$ must be orthogonal to other variables available at time t. Formally, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\hat{G}_{t+1}-\hat{G}_{m,t+1})\otimes z_t\right]S_w^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[(\hat{G}_{t+1}-\hat{G}_{m,t+1})\otimes z_t\right]'$ is asymptotically χ^2_{df} where df equals the dimension of z times the number of moments in G.

The moments we use in z include the variance of asset holdings for workers, the number of unconstrained agents in the bond market, the wealth holdings of the 10% richest agents, and the variance of the wealth holdings of the 10% richest agents, the third and fourth non central moments of wealth. We do not reject the hypothesis that these z's cannot help further predict $\hat{G}$. In particular, the p-values for this statistic are no larger than 0.07 and 0.10 for $\bar{k}$ and ξ respectively for the two different Z shocks.

Markov chain for Persistent Income Process

The persistent process, which its variance depends on the aggregate state variable Z, is approximated with an 11-state Markov chain. The elements of the process η are $\eta \in \{-3.1238, -2.6082, -2.0925, -1.5769, -1.0612, -0.5455, -0.0299, 0.4858, 1.0014, 1.5171, 2.0327\}$. We account for the dependence of this process on the aggregate shock Z by having *two* distinct 11×11 transition matrices, each corresponding to one of the shocks Z can take.

C Calibration Appendix

In this section we describe how we calibrate the Constantinides and Duffie (1996) and Mehra and Prescott (1985) economies which appear in Exhibit 1. We also demonstrate the sense in which our specification for countercyclical cross sectional variation (CCV) — heteroskedasticity in the innovations to the idiosyncratic component of log income — is consistent with the approach used by previous authors (*e.g.*, Heaton and Lucas (1996), Constantinides and Duffie (1996)). In each case the cross sectional variance which matters turns out to be the variance of the change in the log of an individual's share of income and/or consumption.

Our calibration of both the Constantinides-Duffie and the Mehra-Prescott models is driven by a two state Markov chain for aggregate consumption growth. In order to make the comparison in Exhibit 1 as meaningful as possible, we calibrate the parameters of the Markov chain to match the mean, standard deviation and autocorrelation of aggregate consumption growth from our baseline economy. The differences in the values we use, relative to those used in Mehra and Prescott (1985), are minor, owing largely to the fact that our aggregate economy is calibrated to HP filtered U.S. data from the outset. Specifically, whereas Mehra-Prescott base their calibration on a mean, standard deviation and autocorrelation (of aggregate consumption growth) of 0.018, 0.036 and -0.14, respectively, our calibration is based on values of 0.015, 0.023 and -0.23.

We calibrate the Constantinides and Duffie (1996) model via a simple reinterpretation of the preference parameters of the Mehra and Prescott (1985) representative agent. Recall that we use β and α to denote an *individual* agent's utility discount factor and risk aversion parameters, respectively. Constantinides and Duffie (1996) construct a representative agent (their equation (16)) whose rate of time preference and coefficient of relative risk aversion are (using our notation),

$$-\log\hat{\beta} = -\log(\beta) - \frac{\alpha(\alpha+1)}{2}a \quad , \tag{18}$$

and

$$\hat{\alpha} = \alpha - \frac{\alpha(\alpha+1)}{2}b \quad , \tag{19}$$

respectively. In these formulae, the parameters a and b relate the cross sectional variance in the change of the log of individual i's share of aggregate consumption (y_{t+1}^2, using Constantinides-Duffie's notation) to the growth rate of aggregate consumption, as follows.

$$Var(\log\frac{c_{i,t+1}/c_{t+1}}{c_{it}/c_t}) = a + b\log\frac{c_{t+1}}{c_t} \quad . \tag{20}$$

All that we require, therefore, are the numerical values for a and b which are implied by our PSID-based estimates in Table 1.

Our estimates are based on individual income, y_{it}. In the context of the Constantinides-Duffie model, where the equilibrium outcome is autarkic, we interpret these estimates as pertaining to individual consumption, c_{it}. The alternative — using individual consumption data to calibrate the model — is both more problematic (given the quality of the available consumption data) and less likely to generate sizable asset pricing effects. Using income data, in this sense, gives the Constantinides-Duffie model a leg up. In addition, the objective here is more relative than absolute. Asset prices in our OLG model are functions of consumption allocations, which are, in turn, functions of our estimates of idiosyncratic *income* risk parameters. What Exhibit 1 asks is, what would the Constantinides-Duffie economy look like, were its agents to be endowed with idiosyncratic risk of a similar magnitude? Moreover, how does our model measure up, in spite of its non-degenerate (and more realistic) risk sharing technology? Using income data seems appropriate in this context. Notationally, for the remainder of this appendix, we therefore set $c_{it} = y_{it}$.

Next, we need to establish the relationship between our specification for idiosyncratic shocks and the individual shares of aggregate consumption which appear in equation (20). Denote individual $i's$ share at time t as γ_{it}, so that,

$$\log \gamma_{it} \equiv \log c_{it} - \log \tilde{E}_t c_{it} \quad ,$$

where the notation $\tilde{E}_t(\cdot)$ denotes the cross sectional mean at date t, so that $\tilde{E}_t c_{it}$ is date t, per capita aggregate consumption. For our specification, if we ignore the transitory shocks, ε_{it}, as well as the terms which capture cross sectional variation due to age and education (see equation (7) in section 3.2), then our estimation in Table 1 boils down to a time series model of the residuals from a regression involving only year-dummy variables. In a large cross section this will be,

$$z_{it} = \log c_{it} - \tilde{E}_t \log c_{it} \quad ,$$

which have a cross sectional mean of zero, by construction, and a sample mean of zero, by least squares. The difference between our specification and the log-share specification is, therefore,

$$\begin{aligned} \log \gamma_{it} - z_{it} &= \tilde{E}_t \log c_{it} - \log \tilde{E}_t c_{it} \\ &= \tilde{E}_t \log \gamma_{it} - \log \tilde{E}_t \gamma_{it} \quad . \end{aligned}$$

The share, γ_{it}, is *defined* so that its cross sectional mean is always unity. The second term is therefore zero. For the first term, note that in both our economy and the statistical model underlying our estimates, the cross sectional distribution is log normal, *conditional* on knowledge of current and past aggregate shocks. If some random variable x is log normal and $E(x) = 1$, then $E(\log x) = -Var(\log x)/2$. As a result,

$$\log \gamma_{it} - z_{it} \quad = \quad -\frac{1}{2} \tilde{V}_t(\log \gamma_{it}) \quad ,$$

where $\tilde{V}_t$ denotes the cross sectional variance operator. It is important to note that, because lives are finite in our model, and because we interpret data as being generated by finite processes, this cross sectional variance will always be well defined, irrespective whether or not the shocks are unit root processes.

The quantity of interest in equation (20) can now be written as,

$$
\begin{aligned}
\log \frac{c_{i,t+1}/c_{t+1}}{c_{it}/c_t} &\equiv \log \gamma_{i,t+1} - \log \gamma_{it} \\
&= z_{i,t+1} - z_{it} - \frac{1}{2}\left(\tilde{V}_{t+1}(\log \gamma_{i,t+1}) - \tilde{V}_t(\log \gamma_{it})\right) \qquad (21)
\end{aligned}
$$

The term in parentheses — the difference in the variances — does not vary in the cross section. Consequently, application of the cross sectional variance operator to both sides of equation (21) implies,

$$
\tilde{V}_{t+1}\left(\log \frac{c_{i,t+1}/c_{t+1}}{c_{it}/c_t}\right) = \tilde{V}_{t+1}\left(z_{i,t+1} - z_{it}\right) \quad .
$$

Ignoring the transitory shocks, the process underlying the estimates in Table 1 is:

$$
z_{i,t+1} - z_{it} = (1-\rho)z_{it} + \eta_{i,t+1} \quad ,
$$

where the variance of $\eta_{i,t+1}$ depends on the aggregate shock. For values of ρ close to unity the variance of changes in z_{it} is approximately equal to the variance of $\eta_{i,t+1}$. The left side of equation (20) is, therefore, approximately equal to the variance of innovations, $\eta_{i,t+1}$,

$$
\tilde{V}_{t+1}\left(\log \frac{c_{i,t+1}/c_{t+1}}{c_{it}/c_t}\right) \approx \tilde{V}_{t+1}\left(\eta_{i,t+1}\right) \quad .
$$

In this sense, the estimates of σ_H and σ_L in Table 1 provide estimates of what is necessary to calibrate the Constantinides-Duffie model.

All that remains are to map our estimates into numerical values for a and b from equation (20). Since aggregate consumption growth — the variable on the right side of equation (20) — takes on only two values (3.8 percent and -0.8 percent), computing the parameters a and b simply involves two linear equations:

$$
\begin{aligned}
0.037 &= a + 0.038b \\
0.181 &= a - 0.008b \ ,
\end{aligned}
$$

where the values on the left are the cross sectional variances from Table 1. The resulting values are $a = 0.156$ and $b = -3.130$.

To summarize, the Sharpe ratios reported in Exhibit 1 for the Mehra-Prescott economy are simply those associated with their model, given the slightly different calibration for the aggregate consumption process, discussed above. The values for the Constantinides-Duffie model correspond to the same Mehra-Prescott economy, but where the parameter values are re-interpreted as is dictated by equations (18) and (19). In each case, the value for α — the risk aversion coefficient associated with an *actual agent* — is set according to the entry in the table whereas the value for β is chosen to match the theoretical expected return on the risk asset to its sample counterpart of just under 7 percent per annum.

Table 1
Idiosyncratic Endowment Process: Parameter Estimates

	ρ	σ_η^2	σ_H^2	σ_L^2	σ_ε^2
		A. Homoskedastic Innovations			
Estimate	.935	.061	–	–	.017
Standard Error	.008	.004	–	–	.005
		B. Heteroskedastic Innovations			
Estimate	.916	–	.037	.181	.025
Standard Error	.009	–	.007	.033	.007

Entries describe GMM estimates, based on the age-dependent moments (12), for the idiosyncratic endowment process described in the text:

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

where $\varepsilon_{it} \sim \text{Niid}(0, \sigma_\varepsilon^2)$, $\eta_{it} \sim \text{Niid}(0, \sigma_\eta^2(Y_t))$ and

$$\begin{aligned} \sigma_\eta^2(Y_t) &= \sigma_H^2 \ \text{ if } \ y_t - y_{t-1} \geq \sum_{j=2}^{T} (y_j - y_{j-1})/(T-1) \\ &= \sigma_L^2 \ \text{ if } \ y_t - y_{t-1} < \sum_{j=2}^{T} (y_j - y_{j-1})/(T-1) \ . \end{aligned}$$

We use u_{it}^h to denote the idiosyncratic component of the i'th household's endowment at time t, where the household head is of age h. Our annual panel, obtained from the Panel Study on Income Dynamics (PSID), spans the years 1968-1991 and is fully described in the text and in appendix A. Observations on u_{it}^h are obtained as the residuals from the first-stage regression, equation (7) in the text. We use y_t to denote the logarithm of per capita income, Y_t, which we obtain from the U.S. National Income and Product Accounts (NIPA) for the years 1910-1991. The longer time series on aggregate income is necessary (and helpful) because of the heteroskedastic nature of the innovations, η_{it} (further details are provided in the text). Standard errors are computed using the White (1980) estimator and incorporate sampling uncertainty from the first-stage regression.

Table 2
Aggregate Moments: Baseline Economy

Panel A: Population Moments, Baseline Theoretical Economy

	Std Dev	Autocorrelation	Correlation with Output
Output	0.060	0.44	1.00
Investment	0.141	0.20	0.86
Consumption	0.018	0.67	0.89

Panel B: Sample Moments, Detrended U.S. Economy, 1955-1997

	Std Dev	Autocorrelation	Correlation with Output
Output	0.022	0.52	1.00
Investment	0.085	0.36	0.85
Consumption	0.018	0.62	0.91

U.S. sample moments are based on annual NIPA data obtained from the Bureau of Economic Analysis, 1955-1997. Each series was detrended by applying the Hodrick-Prescott filter with a smoothing parameter of 100 to the natural logarithm of the deflated series. Theoretical moments are also based on logarithms and are computed as sample averages of a long simulated time series.

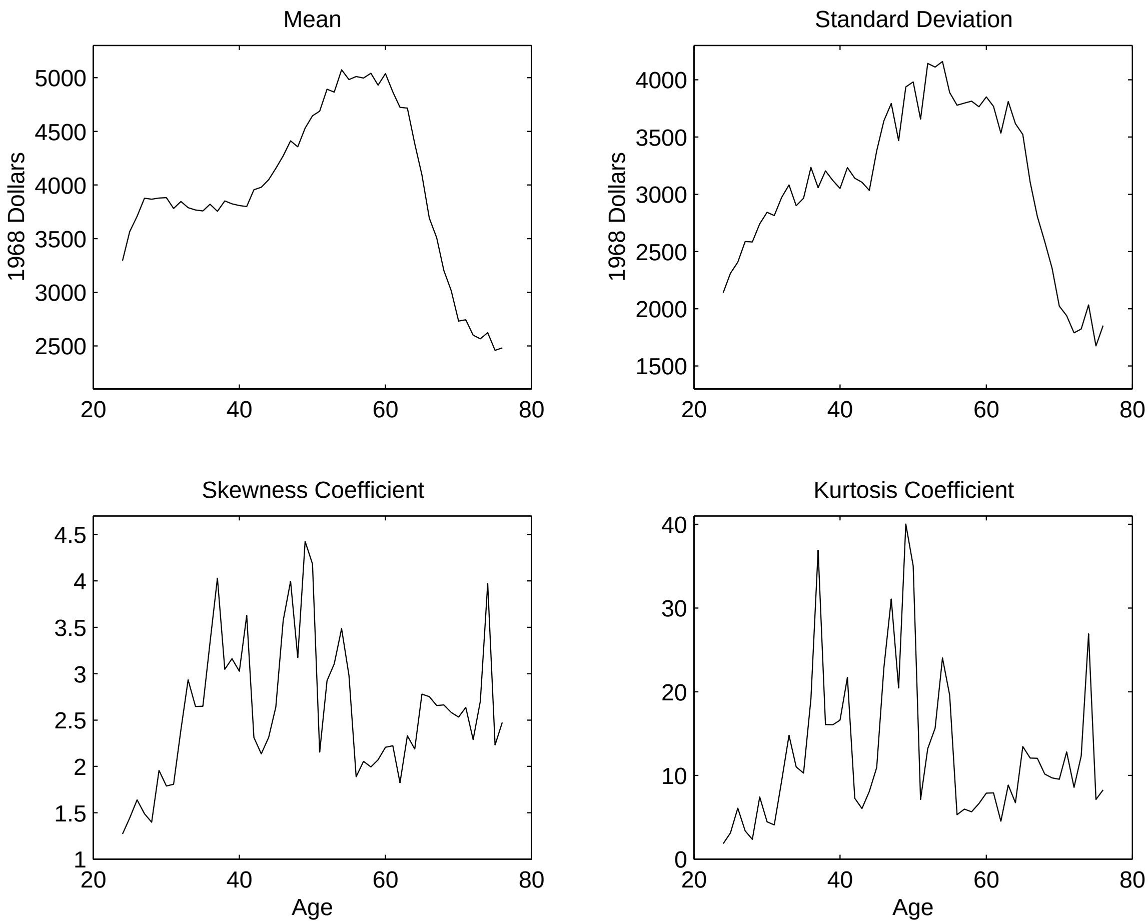
Table 3
Population Moments: Asset Prices

	Mean	Std Dev	Autocorrelation
	Panel A: Baseline Theoretical Economy		
Bond return	4.61	0.18	0.257
Return on capital	4.65	0.69	0.222
Excess return on capital	0.054	0.67	0.003
Sharpe ratio, capital	8.02		
	Panel B: Complete Markets Theoretical Economy		
Bond return	10.01	0.30	0.907
Return on capital	10.02	0.52	0.413
Excess return on capital	0.0043	0.42	0.027
Sharpe ratio, capital	1.03		
	Panel C: Annual Sample Moments, U.S. Economy		
Bond return	1.30	1.88	0.75
Return on equity	8.15	16.67	-0.18
Excess return on equity	6.85	16.64	-0.23
Sharpe ratio, equity	41.17		

U.S. sample moments for equity and short term bonds are computed using non-overlapping annual returns, (end of) January-over-January, 1956-1996. Estimates of means and standard deviations are qualitatively similar using annual data beginning from 1927, or a monthly series of overlapping annual returns. Equity data correspond to the annual return on the CRSP value weighted index, inclusive of distributions. Each annual short term bond return corresponds to an end of January investment in the one month U.S. treasury bill, with the proceeds rolled over in the one month bill for the subsequent year. Nominal returns are deflated using the GDP deflator.

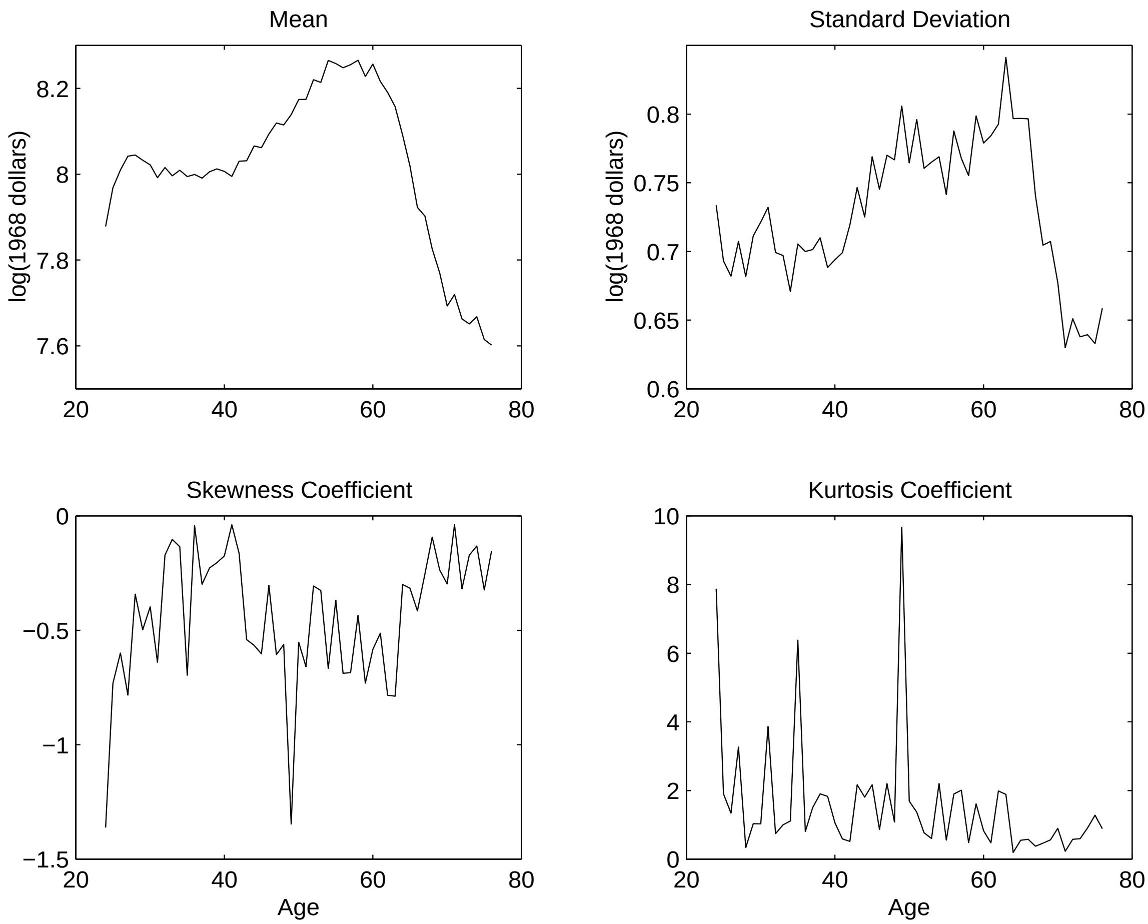
All return moments are expressed as annual percentages. The Sharpe ratio, expressed here in percentage, is the ratio of the unconditional mean to the unconditional standard deviation. Theoretical moments are computed as the sample averages of a long simulation.

Figure 1
Level of Income: Cross Sectional Moments by Age



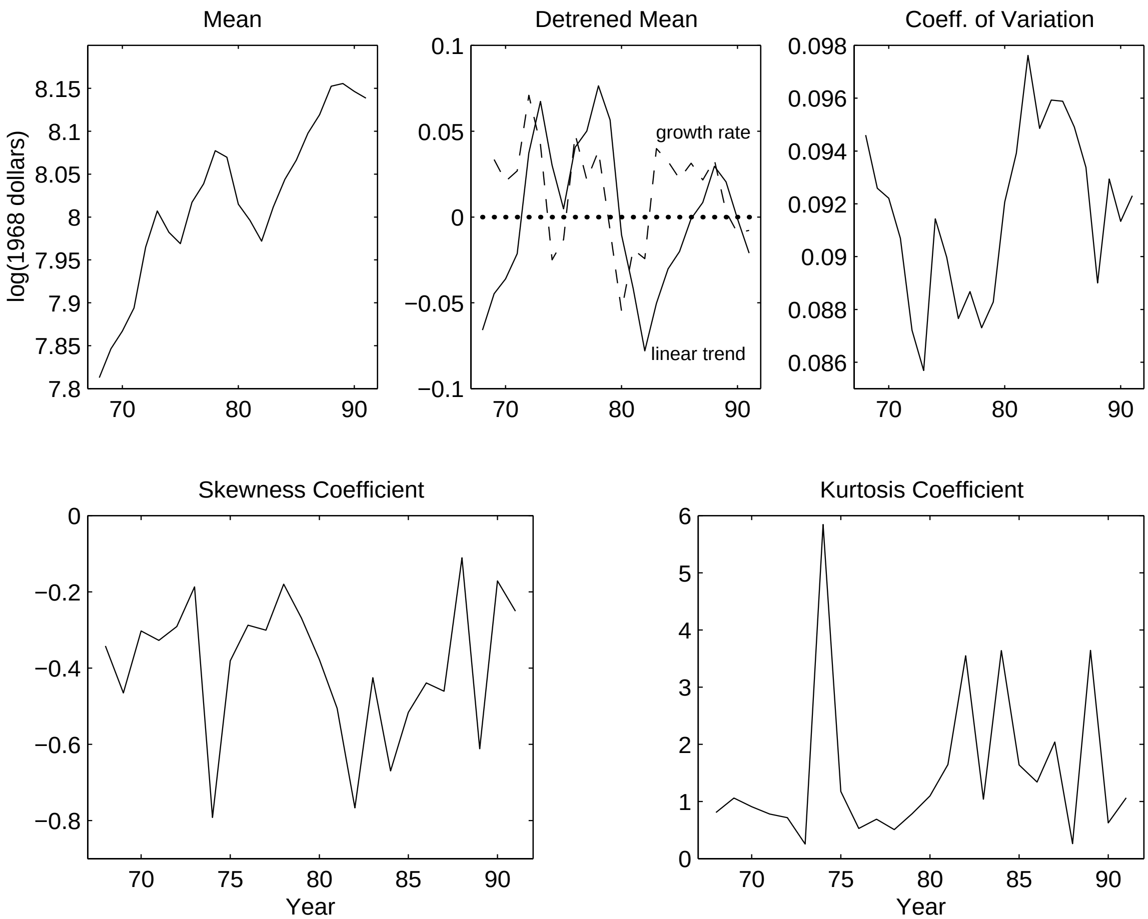
Sample moments are estimated using a sequence of three year, overlapping panels from the Panel Study on Income Dynamics (PSID), spanning the years 1968-1991. Appendix A provides a detailed description of our sampling procedure and income definition. Each point on a given graph is computed by pooling time series observations on the associated age cohort and computing the relevant cross sectional sample moment(s). Skewness and kurtosis coefficients are scaled in the standard manner so as to make deviations from zero represent departures from normality.

Figure 2
Logarithm of Income: Cross Sectional Moments by Age



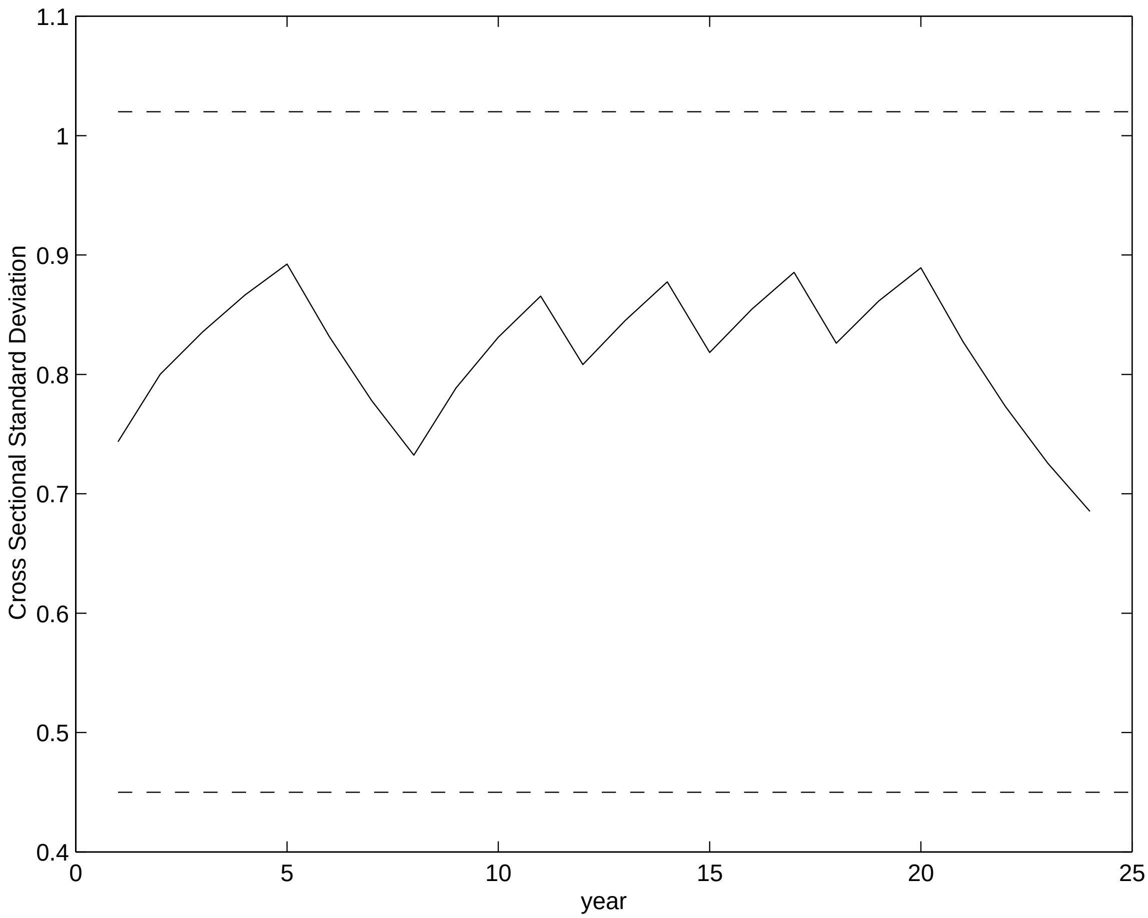
Sample moments are estimated using a sequence of three year, overlapping panels from the Panel Study on Income Dynamics (PSID), spanning the years 1968-1991. Appendix A provides a detailed description of our sampling procedure and income definition. Each point on a given graph is computed by pooling logarithms of time series observations on the associated age cohort and computing the relevant cross sectional sample moment(s). Skewness and kurtosis coefficients are scaled in the standard manner so as to make deviations from zero represent departures from normality.

Figure 3
Logarithm of Income: Cross Sectional Moments by Time



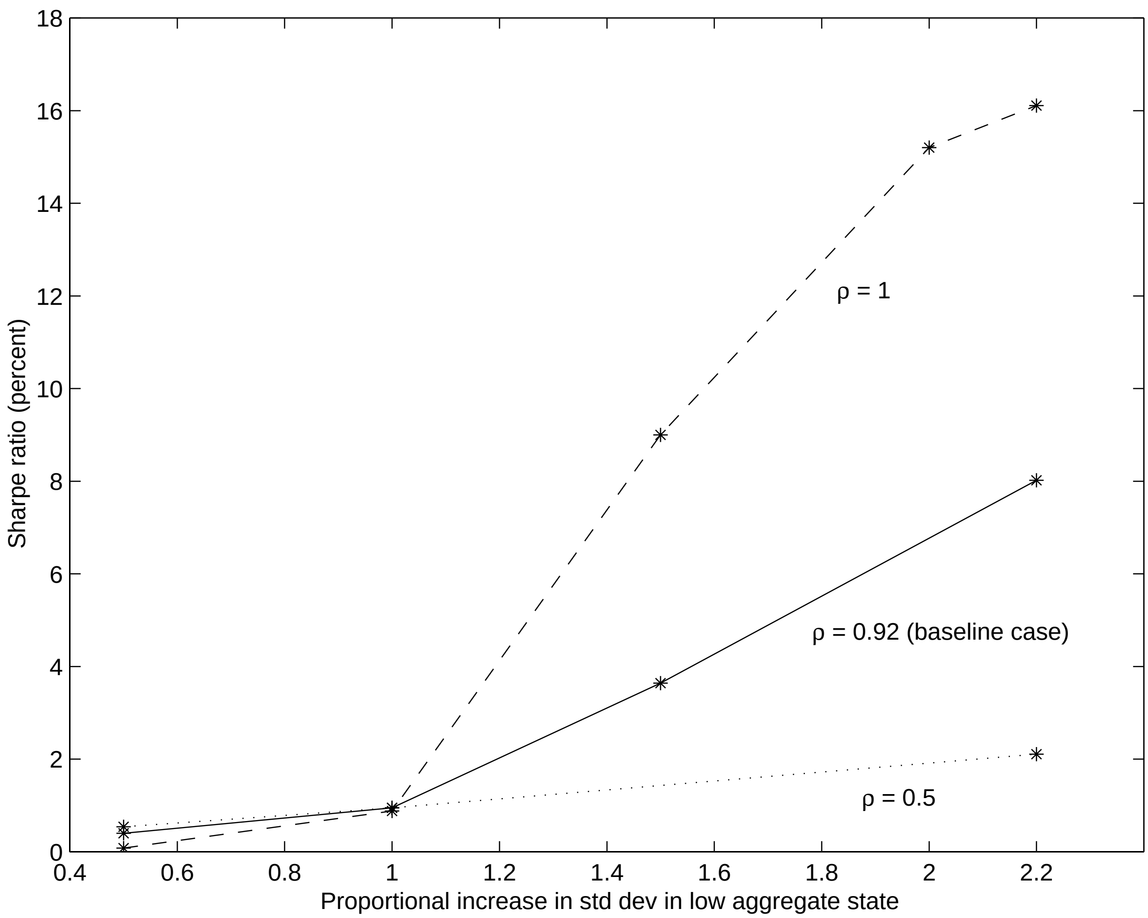
Sample moments are estimated using a sequence of three year, overlapping panels from the Panel Study on Income Dynamics (PSID), spanning the years 1968-1991. Appendix A provides a detailed description of our sampling procedure and income definition. Each point on a given graph represents the associated sample moment of the cross sectional distribution at a given point in time. The solid line in the 'detrended mean' graph represents deviations around a linear trend fit through the cross sectional mean (the upper left graph). The dashed line represents the growth rate in the cross sectional mean. The correlation between the deviations from the mean and the coefficient of variation is -0.85, whereas the correlation between the growth rate and the coefficient of variation is -0.33. Skewness and kurtosis coefficients are scaled in the standard manner so as to make deviations from zero represent departures from normality.

Figure 4
Simulated Cross Sectional Standard Deviation



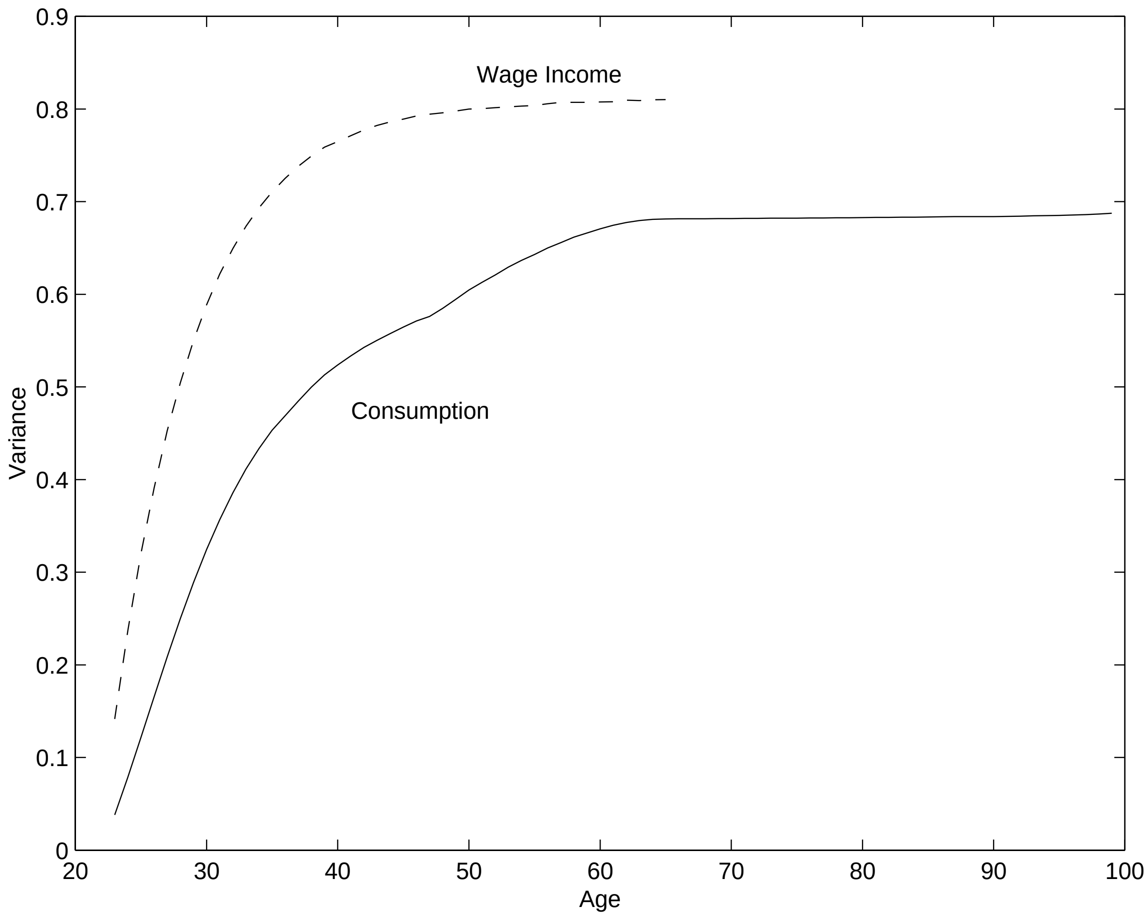
The solid line represents one particular time series realization — over 24 years — of the overall (*i.e.*, across all generations) cross sectional standard deviation of labor market income from our theoretical model. The upper, dashed line, corresponds to the limit point, should the idiosyncratic shocks be drawn from the high variance (recessionary) distribution, for a long sequence of consecutive periods. The lower dashed line represents the analogous limit point, but where the shocks are drawn from the low variance (expansionary) distribution. Further details are provided in section (3.3) of the text.

Figure 5
Sharpe Ratio on Excess Return Portfolio: Effect of Heteroskedastic Conditional Variance



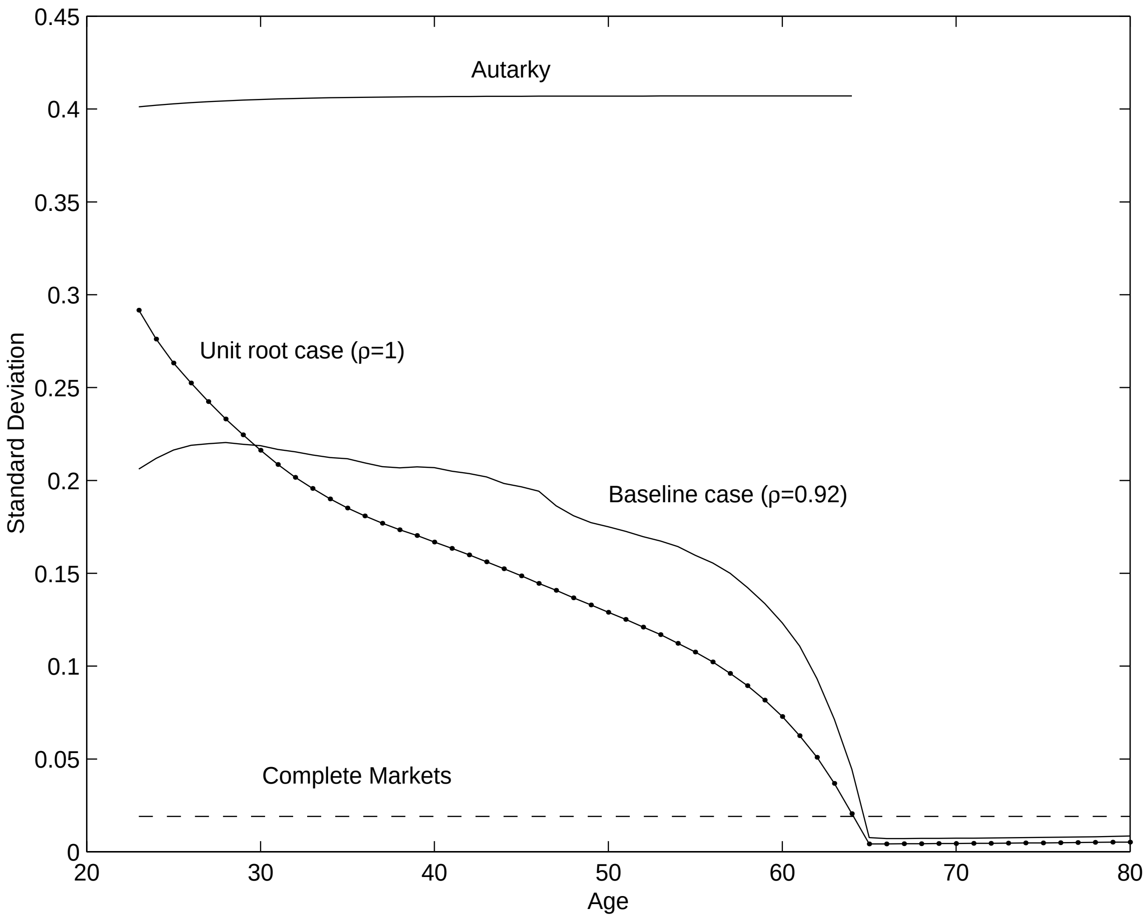
Points on the graph represent the Sharpe ratio — the ratio of the unconditional mean to unconditional standard deviation — on a portfolio which pays the excess return on the risky aggregate technology in our two-asset economy. Each locus represents a set of economies with a different value for ρ, the persistence parameter for idiosyncratic shocks. The horizontal axis represents the proportional increase in the conditional standard deviation of the idiosyncratic shock process as our economy moves from the high productivity aggregate state to the low productivity aggregate state. For instance, the point corresponding to a 'Proportional increase in conditional std dev in low aggregate state' of unity is simply the homoskedastic economy. The point at 2.0 represents an economy where the conditional standard deviation doubles when moving from the high productivity state to the low productivity state. Theoretical moments are computed as the sample averages of a long simulation.

Figure 6
Variability of Consumption and Labor Income by Age



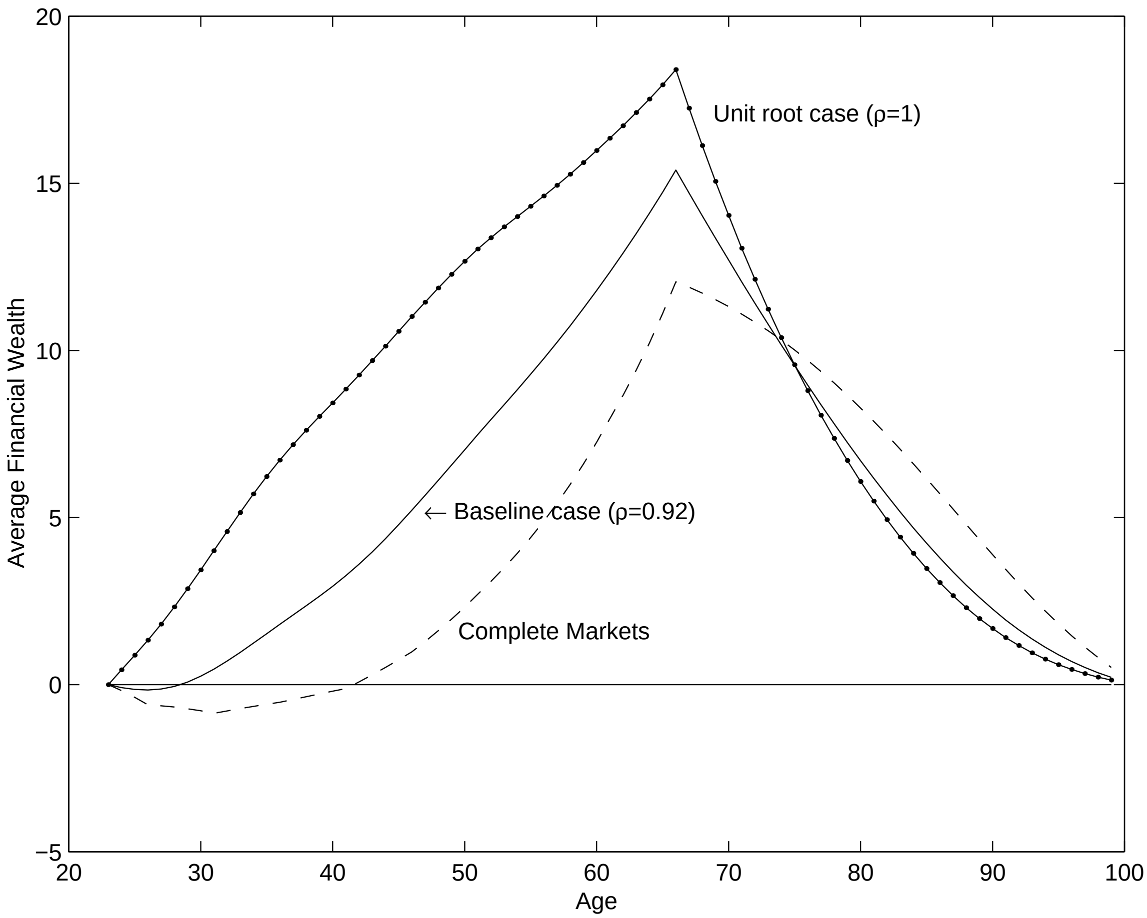
This graph corresponds to the baseline economy described in section 4.1. The dashed line represents the cross sectional variance (which coincides with the unconditional time-series variance for an unborn agent) of the logarithm of wage income for each age cohort, prior to retirement. The solid line represents the age-specific, cross sectional variance of the logarithm of individual consumption. As a reference point, the variance of the logarithm of *per capita* consumption is 0.0003 (for a standard deviation of 0.018). Theoretical moments are computed as the sample averages of a long simulation.

Figure 7
Variability of Consumption Growth by Age



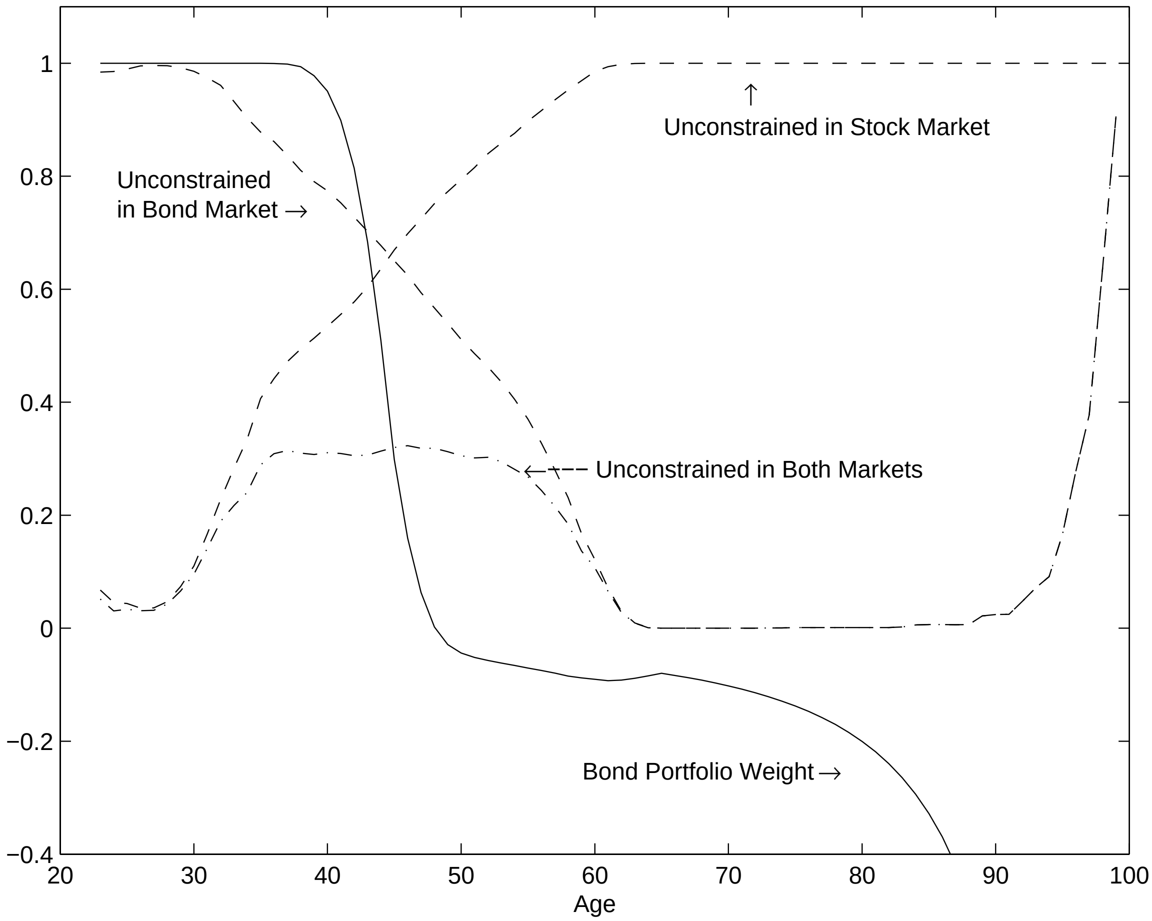
Each line represents the coefficient of variation of the consumption growth rate for agents in a given age cohort. The solid line corresponds to the baseline economy, whereas the dashed line corresponds to the complete markets counterpart discussed in the text. The population moments in this graph were computed by averaging over realizations from a simulation of 2000 time periods with 400 agents in each age cohort.

Figure 8
Average Financial Wealth by Age



Each line represents the average value of financial wealth — capital plus bonds — across agents in a given age cohort. The solid line corresponds to the baseline economy, whereas the dashed line corresponds to the complete markets counterpart discussed in the text. The population moments in this graph were computed by averaging over realizations from simulation of 2000 time periods with 400 agents in each age cohort.

Figure 9
Riskfree Asset Portfolio Weight, by Age



FIX THIS. The graph gives the fraction of financial wealth invested in the riskfree asset for the median agent of a given age cohort. The stars correspond to age cohorts for which the median agent (by bondholdings) is constrained in the bond market. Similarly, the circles correspond to age cohorts for which the median stockholder is constrained (at zero). The population moments in this graph were computed by averaging over realizations from simulation of 2000 time periods with 400 agents in each age cohort. This graph corresponds to an economy with unit root labor income.

Figure 10
Theoretical Decision Rules