Macroeconomic Dynamics, **1**, 1997, 387–422. Printed in the United States of America.

INCOME AND WEALTH HETEROGENEITY, PORTFOLIO CHOICE, AND EQUILIBRIUM ASSET RETURNS

PER KRUSELL
University of Rochester

ANTHONY A. SMITH, JR.
Graduate School of Industrial Administration,
Carnegie Mellon University

We derive asset-pricing and portfolio-choice implications of a dynamic incomplete-markets model in which consumers are heterogeneous in several respects: labor income, asset wealth, and preferences. In contrast to earlier papers, we insist on at least roughly matching the model's implications for heterogeneity—notably, the equilibrium distributions of income and wealth—with those in U.S. data. This approach seems natural: Models that rely critically on heterogeneity for explaining asset prices are not convincing unless the heterogeneity is quantitatively reasonable. We find that the class of models we consider here is very far from success in explaining the equity premium when parameters are restricted to produce reasonable equilibrium heterogeneity. We express the equity premium as a product of two factors: the standard deviation of the excess return and the market price of risk. The first factor, as expected, is much too low in the model. The size of the market price of risk depends crucially on the constraints on borrowing. If substantial borrowing is allowed, the market price of risk is about one one-hundredth of what it is in the data (and about 15% higher than in the representative-agent model). However, under the most severe borrowing constraints that we consider, the market price of risk is quite close to the observed value.

Keywords: Incomplete Markets, Heterogeneity, Equity Premium, Market Price of Risk

1. INTRODUCTION

A recent strand of literature, starting with papers by Marcet and Singleton (1990), Telmer (1993), Lucas (1995), Den Haan (1996a), and Heaton and Lucas (1996), has explored whether incomplete asset markets and heterogeneity among consumers

For their helpful comments, we would like to thank Ravi Bansal, Wouter den Haan, Lars Hansen, John Heaton, Martin Lettau, Kjetil Storesletten, Tom Tallarini, Chris Telmer, Harald Uhlig, Amir Yaron, Harold Zhang, an anonymous referee, and seminar participants at Cornell University, Queen's University, and at the conference on Computation and Estimation in Economics and Finance, held in September 1995 at Washington University in St. Louis. Address correspondence to: Anthony A. Smith, Jr., Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213, USA; e-mail: smithaa@andrew.cmu.edu.

 1365-1005/97 $9.00 + .10

can help explain asset pricing anomalies such as the equity premium puzzle [see, e.g., Mehra and Prescott (1985) or the recent survey by Campbell (1997)]. One way of summarizing the results from the earlier part of this literature is as follows:

1. It is, at least in principle, possible to generate larger equity premia with this set of models.
2. For large equity premia to result, however, significant equilibrium heterogeneity in consumption (and intertemporal marginal rates of substitution) is necessary.
3. Only one asset (a riskless asset or a market portfolio) goes a long way in terms of providing reasonable insurance for agents, thus preventing such heterogeneity from occurring in equilibrium.
4. As a result, this framework does not appear capable of explaining the asset pricing puzzles without relying on (unreasonably) high idiosyncratic income variance, (unreasonably) tight borrowing constraints, or transactions costs (which can be analyzed without heterogeneity).

From the perspective of these papers, then, the success in explaining asset prices with this endeavor has been partial at best.

A much more optimistic perspective, however, is offered in the recent paper by Constantinides and Duffie (1996), who are able to construct an analytical framework with which they illustrate how a low risk-free rate and a large risk premium are possible as equilibrium phenomena in an incomplete-markets heterogeneous-agent model. The key ingredient in their model is the assumption that (log) labor income follows a random walk (with drift). This assumption and its implications for individual consumption processes, the authors argue, are at least not at apparent odds with empirical studies of individuals.

In the present paper, we construct a model along the lines of the above studies: We consider a dynamic model with heterogeneous agents in which idiosyncratic risk is only insurable partially and indirectly by holding aggregate assets (a riskless bond and aggregate capital). The principal purpose of this undertaking is to evaluate the equilibrium asset prices and portfolio choices that obtain under these assumptions. Our paper adds value to the existing literature, first, by pointing to the necessity of restricting the set of incomplete-markets models to those with reasonable implications for heterogeneity (a point to be elaborated below). Second, we explore the implications for the equity premium and the market price of risk for the class of models we thus are restricted to.

At this stage of the incomplete-markets asset pricing research program it seems important to subject the models' implications for heterogeneity to more stringent tests: The most recent literature has results in which asset prices look more empirically reasonable, but the question is whether these results rely on assuming quantitatively unreasonable heterogeneity. Earlier papers [e.g., Telmer (1993), Den Haan (1994, 1995, 1996a), Heaton and Lucas (1995, 1996), and Lucas (1995)], which are less successful in generating realistic asset prices, are not different in their basic approach, but they differ in some of the modeling details: They build on a model framework with two infinitely lived agents [as opposed to the continuum of agents of Constantinides and Duffie (1996)], and they use different processes for the idiosyncratic income shocks.[1] Our present framework—one with a large number of agents—is designed to replicate some of the key features of real-world heterogeneity. This framework allows us to derive an alternative set of implications for asset prices and to discuss the role of the specific assumptions used in the existing literature.

Although the two-agent models have been parameterized in as reasonable a way as possible, their population structure makes them hard to confront with cross-sectional data on individuals. Such data, for example, reveal significant dispersion in consumption growth rates across agents, and the assumption of two agents forces any heterogeneity in the model to have direct aggregate consequences, or direct consequences for average consumption variability. This is potentially an important restriction; in a continuum-of-agents model, asset prices may end up being determined by (the intertemporal marginal rates of substitution of) a very small group of agents. Because only the equilibrium interactions can tell who will be in this group, it is hard a priori to guess what the difference is between two-agent and many-agent versions of the same setup. These points are illustrated in our discussion below of how the market price of risk is determined in the present model.

Constantinides and Duffie, at another end of the spectrum of possible models, do consider a richer population structure, but they reach their results only by restricting the income processes of agents to a very narrow class. It turns out, as we show in this paper, that this restriction leads to a distribution of asset holdings in the population that is highly unrealistic: By construction, *every agent* in the economy has to have the same level of asset holdings. Given that real-world asset holdings are quite dispersed among consumers, in fact significantly more so than are individual earnings, this feature of the Constantinides and Duffie model seems highly problematic.[2] Relatedly, in terms of asset price determination, a feature of the Constantinides and Duffie model is that all agents have interior portfolio decisions, i.e., the group of agents determining asset prices is the whole population.

The economic framework that we use is based on our earlier work in Krusell and Smith (1996b) where we introduce aggregate productivity shocks into the continuum-of-agents, precautionary-savings version of the neoclassical growth model studied by Aiyagari (1994). Our earlier paper shows that equilibria in that framework could be characterized without using all moments of the income and wealth distributions (a result we refer to as "approximate aggregation"), and thus that it is computationally feasible to explore this class of models. There, however, we consider only one asset—aggregate (risky) capital. The introduction of a second asset—a riskless bond—into this environment is an important robustness check but poses nontrivial computational problems. The bond market needs to clear at each date and state, and to see whether this is possible with a bond price that can only depend on a few moments of the wealth distribution is the main challenge in this respect. We find, fortunately, that an extension to our computational algorithm for finding equilibria works quite well in the two-asset model, and that equilibrium quantities are very similar in the one- and two-asset versions of the model.

Like Mehra and Prescott (1985) and Ríos-Rull (1994), who compute equity premia in contexts without idiosyncratic risk, and like the studies of two-agent models with idiosyncratic risk and incomplete markets, we find that the equity premium in our model is nowhere near that observed in postwar U.S. data.[3] However, our model is able to produce values for the market price of risk which are closer to the data. This result is not without interest, because one can argue that the NYSE-based measure of the market price of risk is much closer to its model counterpart than is the NYSE return premium. The model's risky asset is the aggregate capital stock, only a small part of which consists of NYSE assets. Thus, one should not expect the model equity premium to coincide with the NYSE return premium (in particular, the return on aggregate capital fluctuates much less than do NYSE stock returns). However, one can extract implications from NYSE data for the market price of risk (which is a per-unit measure), even though these data do not contain direct measures of the return to total capital. If the model's borrowing constraints are very tight, the implied values for the market price of risk are very large, indeed, as large as those seen in the data. With very lax constraints, the market price of risk falls to a substantially lower level (it falls by about a factor of 100), at which point it is only 15% higher than in the corresponding representative-agent model.

Calibrated versions of the present model and the two-agent models thus have in common that they imply unrealistic asset prices (unless borrowing is severely restricted). However, the economics behind these results are different in some important respects. Whereas individual consumption variability is quite limited in the two-agent frameworks—agents are able to use the assets to trade to almost perfectly insured positions—we find substantial individual consumption variability. For example, the unconditional standard deviation of individual consumption is about four times that of aggregate consumption, and at any moment in time the variation in consumption growth rates across consumers is very large. On the face of it, this finding may seem like great news for asset prices, at least relative to the two-agent economies. However, increasing idiosyncratic consumption variability by an order of magnitude (holding aggregate consumption variability constant) is not sufficient for significantly raising either the equity premium or the market price of risk.

To see why, consider the portfolio first-order condition for any agent who holds both assets. This condition, which resembles the pricing-kernel condition from a standard representative-agent model, can be used to express the ratio of the expected excess return on equity to its standard deviation—a ratio that we define to be the market price of risk for our economy—as a function of the joint distribution of the agent's future marginal utility of consumption and the equity return. Although the variability of individual consumption enters the determination of the market price of risk through this formula, the effects of simply increasing idiosyncratic volatility on the market price of risk are not clearcut. In particular, to understand how the market price of risk changes, it is crucial to know how the added volatility is related to the equity's return. This point is made by Mankiw (1986), who shows that the equity premium can actually fall with increased idiosyncratic consumption variability if the added variability is concentrated in states in which the equity return is high. Mankiw's argument suggests that an increased market price of risk requires the idiosyncratic consumption volatility to be higher in bad times than in good times.

We use the above arguments to derive a bound for the market price of risk that applies in incomplete-markets economies. This bound makes clear the role of idiosyncratic consumption volatility, and it shows how Mankiw's argument applies in our context. In our economy, because employment is more likely in good aggregate states, the variability of individual consumption is negatively correlated with the asset return. Therefore, at least on a qualitative level, the idiosyncratic risk in our model does contribute to an increase in the market price of risk and to an increase in the equity premium.

From a quantitative perspective, however, very large fluctuations in individual consumption are needed to produce enough suffering from risk to warrant a large increase in the market price of risk. This observation goes back to the work by Lucas (1987), Cochrane (1989), Atkeson and Phelan (1994), Krusell and Smith (1996a), Tallarini (1996), and others: Even if idiosyncratic consumption variability is raised by what seems to be a large amount, insurance in this class of models is still excellent *in terms of utility*.

The key equilibrium determinant of the market price of risk in an incomplete-markets economy is thus the variability in marginal utility of the agents who hold both risky and riskless assets. For our parameterizations, this ends up being a small subset of the agents.[4] Such an agent's view about risk depends on his asset position and on his current income and preference shocks. In economies with very tight restrictions on borrowing, these agents turn out to be quite poor and to expect very large (i.e., well above average) fluctuations in consumption and marginal utility; hence the high market price of risk. In contrast, when the borrowing constraint is more lax, the agents with interior portfolio solutions are in much more of an average position, and do not suffer much from fluctuations.

Our results on the market price of risk are not directly comparable to those in existing models, because the calibrations differ and the assumptions on the set of assets differ (many of these studies also do not report the market price of risk). Telmer (1993) also obtains high values for the market price of risk when borrowing is severely constrained, but these values are accompanied by large negative values for the risk-free rate.[5] Lucas (1995) finds high values as well, but with a different set of assets and a different calibration. Finally, Constantinides and Duffie (1996) can obtain any market price of risk, but as we pointed out, this virtue is accompanied by sharply counterfactual implications for wealth heterogeneity.

We organize the paper as follows. In Section 2 we describe our general framework of analysis. We present the results in steps. In the first step, we discuss the role of persistence of individual shocks. To make this discussion as transparent as possible, we restrict attention to a setup without aggregate uncertainty. Specifically, we show how the distribution of asset holdings in the Aiyagari (1994) model contracts as the labor income process becomes more and more persistent. This discussion, which is limited to the determination of the riskless rate and which also explores the connection with Constantinides and Duffie (1996), is contained in Section 3.1. Thereafter, in Section 3.2, we report the results from the model with aggregate uncertainty and equity-premium and portfolio-choice implications. In addition to income shocks, there are preference shocks in this model: ex ante identical agents have different discount factors ex post, which introduces differences in savings behavior among agents that are sufficient to make the distribution of wealth much more skewed and thus conform with that observed in the data. This section also contains our bounds calculations for the market price of risk. Section 4 concludes.

2. THE MODEL

2.1. Primitives

We describe the full model in this section. Whereas the agents are ex ante identical, there is ex post heterogeneity deriving from two sources: idiosyncratic, only partially insurable shocks to labor productivity and to the discount factor. That is, at each point in time, two agents may differ in their current productivity and degree of patience and, as we see later, in accumulated wealth.

Our framework is a version of the stochastic growth model. There is a large (measure 1) population of infinitely lived consumers. There is only one consumption good per period and preferences satisfy

$$E_0 \sum_{t=0}^{\infty} \beta_t U(c_t).$$

The accumulative discount factor β_t follows the stochastic process

$$\beta_t = \tilde{\beta}\beta_{t-1},$$

and we assume that $\tilde{\beta}$ is a finite-state Markov chain. We defer further description of this process to Section 3.2; in Section 3.1, $\tilde{\beta}$ is assumed to be constant. For the period utility function, we use $U(c) = \log(c)$.

The aggregate technology is standard:

$$y = zk^{\alpha}l^{1-\alpha},$$

where $\alpha \in [0, 1]$, y is output, k is the aggregate capital input (i.e., the sum of individual agents' holdings of capital), l is the aggregate labor input (i.e., the sum of individual agents' labor supplies, measured in efficiency units), and z is a shock to aggregate productivity. Capital accumulation is also standard:

$$k' = (1 - \delta)k + i,$$

with $\delta \in [0, 1]$, where k' denotes capital next period and i investment. Finally, the total resources y can either be consumed or invested.

The shock to aggregate productivity takes on one of two values: $z \in \{z_b, z_g\}$, with $z_g > z_b$. The matrix

$$\begin{bmatrix} \pi_{gg} & \pi_{bg} \\ \pi_{gb} & \pi_{bb} \end{bmatrix}$$

describes the evolution of the aggregate shock: $\pi_{ss'}$ is the probability that the aggregate shock next period is $z_{s'}$ given that it is z_s this period.

In each period, each agent is endowed with one unit of time that is devoted to production because agents do not value leisure. The productivity, or efficiency, of an agent's endowment of time is denoted by ϵ. The labor productivity shock ϵ is independent of the preference shock and evolves in an idiosyncratic fashion according to one of two stochastic processes. In Section 3.1, ϵ follows an AR(1) in logs:

$$\log \epsilon' = \rho \log \epsilon + \varepsilon',$$

where ε is i.i.d. and normally distributed. This income process is used to examine the behavior of the model as the autoregressive parameter ρ is increased toward 1. In this context, we assume a law of large numbers that states that the (cross-sectional) average of labor productivity is constant across time. Total labor supply (measured in efficiency units) therefore can be normalized to 1 without loss of generality. Finally, in this setup, an agent's productivity shock and the aggregate productivity shock are assumed to be independent of each other.

In Section 3.2, we do not consider permanent shocks and therefore use a simpler process for labor productivity. In particular, we assume that $\epsilon \in \{0, 1\}$: The agent is employed when $\epsilon = 1$ and unemployed when $\epsilon = 0$. We also assume that unemployed agents receive an exogenous amount g of goods, which can be interpreted as the value of home production (it is thus a form of unemployment insurance).[6] We assume that g is independent of market conditions. In this setup, unlike in Section 3.1, the individual and aggregate shocks follow a joint first-order Markov structure, where we allow aggregate conditions to affect idiosyncratic ones in a manner suggested by Mankiw (1986). In particular, the condition that adverse aggregate shocks have the effect of increasing idiosyncratic income risk, and thus the cross-sectional income variance, takes a very natural form here: The probability of becoming unemployed is higher in bad times. We use u_g and u_b to denote the unemployment rates in good and in bad times, respectively.[7] The marginal distribution of the aggregate productivity shock is the same as in Section 3.1, i.e., it is determined by the transition matrix given above. The joint transition process for (z, ϵ) is as follows:

$$\begin{bmatrix} \pi_{gg00} & \pi_{bg00} & \pi_{gg10} & \pi_{bg10} \\ \pi_{gb00} & \pi_{bb00} & \pi_{gb10} & \pi_{bb10} \\ \pi_{gg01} & \pi_{bg01} & \pi_{gg11} & \pi_{bg11} \\ \pi_{gb01} & \pi_{bb01} & \pi_{gb11} & \pi_{bb11} \end{bmatrix},$$

where $\pi_{ss'\epsilon\epsilon'}$ is the probability of transition from state (z_s, ϵ) today to $(z_{s'}, \epsilon')$ tomorrow. For consistency with the above marginals and unemployment rates, the transition probabilities need to satisfy

$$\pi_{ss'00} + \pi_{ss'01} = \pi_{ss'10} + \pi_{ss'11} = \pi_{ss'}$$

and

$$u_s \frac{\pi_{ss'00}}{\pi_{ss'}} + (1 - u_s) \frac{\pi_{ss'10}}{\pi_{ss'}} = u_{s'}$$

for all (s, s').

2.2. Market Arrangement

There are two assets available for net saving and for protection against risk: There is a riskless bond, and there is aggregate capital. It is assumed that all agents face the same returns on these assets. It should be clear how the assets can be used to protect against risk: The two assets' returns are linearly independent, and so, they can be used to span a larger space than if there were only one asset. Because the idiosyncratic and the aggregate shocks are correlated, it is clear, moreover, how the risky capital asset can be used, particularly to protect against the idiosyncratic risk. Agents who are afraid of becoming unemployed identify capital as an asset that pays off in precisely the opposite way in which they wish to insure: It pays poorly when times are poor and any given agent is more likely to be unemployed and thus in need of income. Because employment status is positively autocorrelated, we therefore anticipate that unemployed agents particularly will desire a portfolio with relatively little capital.

We use k and b to denote the individual's beginning-of-period holdings of capital and bonds, respectively. We restrict end-of-period asset holdings as follows: $k' \geq \underline{k} > -\infty$, $b' \geq \underline{b} > -\infty$. These restrictions reflect the need to rule out Ponzi schemes but also can be used as a way of limiting the effective ability of agents to smooth consumption. The restrictions on k' and b' imply that next period's wealth cannot fall below a lower bound.

Letting the total amount of capital in the economy be denoted $\bar{k}$ and the total amount of labor supplied $\bar{l}$, our constant-returns-to-scale production function implies that the price of one unit of labor services is $w(\bar{k}, \bar{l}, z) = (1 - \alpha) z (\bar{k}/\bar{l})^{\alpha}$ and that the return to capital services is $r(\bar{k}, \bar{l}, z) = \alpha z (\bar{k}/\bar{l})^{\alpha-1} - \delta$. It is an important feature of this environment that all agents who save in capital receive the same return.

Because we employ a recursive definition of equilibrium, we need to specify the relevant set of aggregate state variables. By this, we mean, loosely speaking, those current variables that have independent impact on current or future equilibrium prices.[8] Our set of state variables, then, contains the aggregate productivity shock and the distribution of agents over their individual wealth, preference, and employment status. The individual's vector of wealth, ω (which we define as the total value of labor income and gross asset income), current $\tilde{\beta}$, and employment ϵ is his relevant individual-specific state variable: Agents that differ with respect to this vector will, in general, behave differently with respect to savings propensities, portfolio choices, and so on. The distribution variable over this vector, Γ, thus keeps track of how many agents are, say, in each wealth interval for each value of the preference/employment status. Note that whether an agent carried his wealth in capital or bonds is not relevant as a separate state variable: The portfolio choice is a control variable in this setup.

The equilibrium law of motion for the aggregate state variable is jointly given by the transition matrix for z, which is exogenous, and by a function H, which is endogenous: $\Gamma' = H(\Gamma, z, z')$. The role of the aggregate state variable and its law of motion from the point of view of the individual is, of course, to predict future prices. The equilibrium pricing functions are given by the two expressions for w and r above, where $\bar{k}$ can be computed by integrating the wealth distribution and $\bar{l}$ is given by $1 - u$; and the bond pricing function $q(\Gamma, z)$.[9]

The individual's optimization problem thus can be cast in terms of the following dynamic programming problem:

$$v_i(\omega, \epsilon; \Gamma, z) = \max_{c,k',b'} \{U(c) + \tilde{\beta}_i E[v_j(\omega', \epsilon'; \Gamma', z') \mid z, \epsilon]\}$$

subject to

$$c + k' + q(\Gamma, z)b' = \omega, \tag{1}$$

$$\omega' = (r(\bar{k}', \bar{l}', z') + 1)k' + b' + \epsilon' w(\bar{k}', \bar{l}', z') + (1 - \epsilon')g, \tag{2}$$

$$\Gamma' = H(\Gamma, z, z'), \tag{3}$$

$$(k', b') \geq (\underline{k}, \underline{b}), \tag{4}$$

where the subscript on the value function indicates the value of the discount factor; the laws of motion for z, $\tilde{\beta}$, and ϵ are implicit in the expectations operator. The decision rules coming out of this problem are denoted by the functions f_i^k and f_i^b: $k' = f_i^k(\omega, \epsilon; \Gamma, z)$ and $b' = f_i^b(\omega, \epsilon; \Gamma, z)$. We suppress the subindex i when we refer to the vector of decision rules.

A recursive competitive equilibrium then is summarized by a law of motion H, the individual's functions v, f^k, and f^b, and pricing functions (r, w) and q such that (v, f^k, f^b) solves the consumer's problem; r and w are competitive (i.e., given by marginal productivities as expressed above); H is generated by f^k, i.e., the appropriate summing up of agents' optimal choices of capital given their current status in terms of wealth and employment; and f^b generates bond market clearing, i.e., $\int f^b(\omega, \epsilon; \Gamma, z)\, d\Gamma = 0$.

3. FINDINGS

We first discuss the effects of persistence in individual income shocks (Section 3.1) in the context of no aggregate shocks, and then move on to the model with aggregate uncertainty in Section 3.2.

3.1. Effects of Persistence in Individual Income

We now analyze a special case of the model in Section 2: We set $z_g = z_b$ and $\beta_t = \tilde{\beta}^t$, where $\tilde{\beta}$ is a constant. This special case does not allow us to study risk premia, but it does allow nontrivial determination of the risk-free rate of interest and of the implications for equilibrium wealth dispersion. Our model here is very closely related to the neoclassical growth-model setup with partially uninsurable income risk in Aiyagari (1994) and also is quite similar to the endowment economy that Huggett (1993) studies for the purpose of analyzing the risk-free rate puzzle.[10]

3.1.1. Calibration. We calibrate the model in this section to a period of a year, and select the discount factor and the depreciation rate in approximate accordance with Aiyagari (1994). To be able to characterize equilibria also for the case with a unit root in log labor income, it is necessary to prevent the distribution of labor productivities from spreading out indefinitely. This can be done in several ways, and we follow Constantinides and Duffie (1996) who assume a (constant) positive risk of dying in an alteration of their baseline model.[11] We assume that this risk, π_d, is 0.02, leading to an effective discount rate of about 0.96, and we assume that newly born agents receive an initial level of capital equal to the mean capital stock.[12] We use a capital share of 0.36 and a borrowing (or short-sales) constraint for capital of zero. In sum, our parameter vector is $(\tilde{\beta}, \alpha, \delta, \pi_d, \underline{k}) = (0.98, 0.36, 0.10, 0.02, 0)$.

For the parameters of the labor income process, we let the autocorrelation ρ vary across experiments (we use $\rho \in \{0.5, 0.98, 0.99, 1\}$) and maintain a constant value of the conditional standard deviation of log labor income.[13] We set the latter so that, at $\rho = 1$, the unconditional standard deviation of log labor income is similar to that used in Aiyagari's study. More precisely, we set the standard deviation of ε_t equal to 0.04 and we select its mean so that, when integrated over the population, the total labor input in efficiency units equals 1.

We solve for the prices in the stationary equilibrium of this economy using the method described by Aiyagari (1994). To solve for the decision rules of a typical agent, given prices, we iterate on the value function in Bellman's equation using cubic splines to interpolate between grid points (see the Appendix for a more complete description of the algorithm and for a discussion of its accuracy). When ρ, the degree of idiosyncratic labor income persistence, is equal to 1, the resulting unit root in log labor productivity prevents us from using a discrete-state Markov chain to approximate the process governing the evolution of log labor productivity. Instead, for all values of ρ, we perform numerical integration using eight-point Gaussian quadrature.

3.1.2. Equilibrium: Wealth dispersion and the risk-free rate. The results from varying ρ—the degree of idiosyncratic income persistence—are contained in Table 1.

TABLE 1. Effect of increased idiosyncratic income persistence

	ρ			
Variable	0.50	0.98	0.99	1.00
Equilibrium interest rate, %	4.12	4.07	4.06	4.07
Aggregate capital	11.60	11.67	11.68	11.66
SD of capital	1.47	6.42	5.94	0.36
Skewness of capital	−0.03	2.58	3.60	4.98
Gini coefficient of capital	0.067	0.255	0.217	0.018

The table shows that the risk-free rate and several aspects of the distribution of capital are nonmonotone in ρ. At low values for ρ, aggregate capital and the standard deviation, skewness, and Gini coefficient of the distribution of capital across agents are all low. As persistence is increased, these variables all increase. However, as ρ gets close to 1, some striking facts appear: The distribution of capital (almost) collapses onto one point! At $\rho = 0.99$, the standard deviation equals 5.94, and at $\rho = 1$, it equals 0.36. Similarly, the Gini coefficient for the asset distribution goes from 0.217 to 0.018. Finally, the skewness (as well as the kurtosis) of the capital distribution is monotone increasing in ρ.

Thus, at high enough levels of persistence, as shocks in some sense become worse and worse, agents are more and more averse to holding low levels of capital: With a low level of capital and high income persistence, the chance of a string of bad shocks that takes the agent down to very low consumption levels is larger. On the other hand, the rate of return on savings is significantly below the discount rate, and so, agents are also more unwilling to hold high levels of capital: When an agent is well insured, what mainly matters for his savings decision is the riskless rate relative to the discount rate. Recall that the average holdings of capital in this economy are endogenous, so that, for example, more risky processes lead to higher average capital holdings. In addition, the return to capital is determined by the marginal productivity of capital, so that the capital stock cannot increase too much without significantly lowering the return to capital. In other words, the higher the average capital holdings, the lower is the return, and hence the more will well-insured agents dissave, and hence the limit to the increase in the economywide stock of capital as ρ goes to 1.

3.1.3. Wealth dispersion in the Constantinides and Duffie model. At this point it is useful to recall the results of Constantinides and Duffie (1996). They use a slightly different individual income process and show that an equilibrium with no trade obtains under certain conditions. In this equilibrium, which has aggregate uncertainty as well, all agents hold exactly the same portfolios by construction. To clarify the connection between this result and the one we obtain here, let us consider the constraints faced by an agent in the Constantinides and Duffie model.

When there is no aggregate uncertainty, the agent's budget reads

$$c_{it} + ps_{i,t+1} = I_{it} + (p+d)s_{i,t}$$

at time t, where i refers to the individual, s_i is the individual's share of the aggregate asset, p is the price of the asset, d its dividend, and I_i the individual's income. The process for income is as follows: $I_{it} = e^{z_{it}} - d$, with z_{it} following a random walk whose innovation is normally distributed with mean $-y^2/2$ and variance y^2, i.e., z_{it} has negative drift and an innovation variance that increases with the absolute value of the drift. This particular combination of innovation drift and variance ensures that the economywide (average) income I has a constant mean over time if z_{i0} is normally distributed, because normality implies that $E(e^{z_{i,t+1}}) = e^{E(z_{i,t+1})+\mathrm{Var}(z_{i,t+1})/2} = e^{E(z_{it})-y^2/2+[\mathrm{Var}(z_{it})+y^2]/2} = e^{E(z_{it})+\mathrm{Var}(z_{it})/2} = E(e^{z_{it}})$, where the E's and Var's are the cross-sectional expectation and variance operators, respectively.[14] Also note that one peculiarity in this framework is that labor income can be negative.

To see why the Constantinides and Duffie formulation yields a no-trade equilibrium, note that the first-order condition to the agent's problem reads as follows, assuming an interior solution[15]:

$$qc_{it}^{-\sigma} = \tilde{\beta} E_t \left[c_{i,t+1}^{-\sigma}\right],$$

where $q \equiv 1/[(p+d)/p]$ denotes the equivalent of the price of a one-period riskless bond with a face value of 1.[16] This equation can be rewritten as follows:

$$q = \tilde{\beta} E_t \left[\left(\frac{c_{i,t+1}}{c_{it}}\right)^{-\sigma}\right].$$

In a no-trade situation, the asset holding is constant over time: $s_{it} = s_i$. Using this in the equation, we obtain

$$q = \tilde{\beta} E_t \left[\left(\frac{e^{z_{i,t+1}} - d + s_i d}{e^{z_{it}} - d + s_i d}\right)^{-\sigma}\right].$$

This equation is satisfied for all agents (that is, for all z_{it}'s and s_i's) *only* if $s_i = 1$ for all agents. In other words, no trade implies equal asset holdings. In this case, the last equation becomes

$$q = \tilde{\beta} E_t \left[e^{-\sigma(z_{i,t+1}-z_{it})}\right] = \tilde{\beta} e^{\sigma(1+\sigma)y^2/2}.$$

This equation expresses the equilibrium risk-free rate (i.e., q^{-1}) as a function of fundamentals. We see that $q^{-1} < \tilde{\beta}^{-1}$, and that the gap is increasing in y, the parameter governing the variance and (negative) drift of the idiosyncratic shocks to income growth, and increasing in the degree of risk aversion. In other words, the risk-free rate can be made arbitrarily small by increasing the variance of individual shocks alone.

Note as a curiosity of the solution that it implies that the real value of assets carried forward equals p, i.e., all agents in the economy are right at the borrowing constraint (see note 15 for further details).

3.1.4. Comparisons. There is a straightforward interpretation of the Constantinides and Duffie equations in terms of a model with capital and production: Let $d = (r - \delta)\bar{k}$, $p = \bar{k}$, $s_{it} = k_{it}/\bar{k}$, and $e^{z_{it}} = w\epsilon_{it}$, in which case the budget can be rewritten as

$$c_{it} + k_{i,t+1} = w\epsilon_{it} - (r - \delta)\bar{k} + (1 + r - \delta)k_{it},$$

which is close to the budget constraint in the neoclassical model that we examine in this paper. There is only one difference: the addition of the negative term $-(r-\delta)\bar{k}$ on the resource side of the constraint. For our calibration, this term is small relative to the other resource terms in the budget. The fact that it is small in relative terms explains why the $\rho = 1$ case of the Aiyagari setup is similar, but not identical, to the Constantinides and Duffie case. Thus, the wealth distribution collapses, but it does not quite collapse onto one point (the standard deviation of the distribution is a little larger than zero).

3.1.5. Remarks. The Constantinides and Duffie model with permanent shocks leads to a striking and counterfactual result—the no-trade steady state has complete equality in asset holdings—and the model we look at has the same essential features when shocks are permanent. A few remarks are worthwhile at this point. First, the omission of aggregate uncertainty in our discussion is not critical to our central message in this section; it is straightforward to amend the setup to include aggregate shocks and argue that the asset distribution will collapse.

Second, the no-trade result in Constantinides and Duffie's paper is often "explained" with reference to the fact that shocks are permanent. The logic is as follows. For an agent to want to borrow when hit by a negative shock, it seems necessary that the agent also expect better times in the future relative to other agents: Otherwise, the borrowing does not help in smoothing consumption over time. However, when shocks are permanent, the current shock is the best predictor of future shocks, and so, in relation to others at least, better (or worse) times are never expected. The results from the Aiyagari-style model makes clear that this line of reasoning is incomplete: This model has permanent shocks, but there is trade nevertheless. It is clear, for example, that the form of preferences and the precise way in which shocks are permanent matters for whether there will be trade or not.

Third, can we conclude that permanent labor income shocks always lead to an equilibrium with a(n essentially) collapsed wealth distribution? We cannot. For example, with different preferences, assets do not need to be equal across agents: With a negative exponential utility function, any asset distribution is a steady state with no trade if the income shocks follow a random walk in levels (this is easy to check). However, a random walk in levels is problematic because it implies that income and consumption will be negative with probability 1 for all agents. Another way out is as follows. It is possible to amend the Constantinides and Duffie model to allow for heterogeneity in asset holdings by letting agents have different labor income processes. If, namely, agent i has $I_{it} = e^{z_{it}} - s_i d$, then a no-trade equilibrium in which agent i holds s_i units of the asset is possible.[17] However, this makes labor and asset income negatively correlated in the population. Finally, a life-cycle model with a unit root in (log) labor income may not lead to a counterfactual wealth distribution. This possibility is under investigation by Storesletten et al. (1996).

Fourth and finally, we would like to point out that the collapse of the asset distribution could to a large extent also have been anticipated from the work of Deaton (1991). Deaton analyzes the decision problem of an agent facing the kind of shocks assumed in Constantinides and Duffie's model, and he shows that, when $q > \tilde{\beta} e^{-\sigma(1+\sigma)y^2/2}$ (using our notation from above), there is an absorbing state: zero asset holdings (which equals the borrowing constraint in his setting). That is, agents draw down on their asset holdings and when zero is reached they stay there forever and let consumption equal labor income. Deaton, however, does not show that the limit case $q = \tilde{\beta} e^{-\sigma(1+\sigma)y^2/2}$ leads to the same outcome, and does not remark that a stationary equilibrium can be constructed using this equation to pin down the equilibrium price.

3.2. Aggregate Uncertainty

The preceding section established what we consider a central, and counterfactual, implication from assuming permanent idiosyncratic shocks to (log) labor income: It leads to a collapse of the asset distribution, which in reality is much more dispersed than the distribution of labor income. However, there are *a priori* reasons to rule out permanent income shocks as well. Heaton and Lucas (1996) and others argue, with reference to studies of individuals, that individual shocks are not permanent. Actually, their argument for using a model with only temporary shocks can be strengthened. In a model such as those we discuss here, agents can only be interpreted as dynasties (because they live forever). Thus, the appropriate counterpart of the model's labor income process in the data is really not the process for given *individuals*: It is the income process of *families*. And here there is less disagreement: There is significant regression to the mean if one compares children's income to that of their parents [for a discussion, see Stokey (1996)]. In other words, even if individuals' income processes had a unit root component in the data, one should not calibrate the present kind of model to have a unit root.

In what follows, we present results from our full model with aggregate uncertainty and less-than-permanent shocks to individual income. We first present results from a model that has no preference shocks and illustrate that the wealth distribution so derived also has much less skewness than in the data. This necessitates a departure from that framework, and we go on to show that our particular approach—to use (a small amount of) randomness and ex post heterogeneity in discount factors—is at least reasonably successful at replicating the distribution of asset wealth. Thus, it is the asset pricing implications from this last framework that we argue should be taken the most seriously.

3.2.1. Calibration. We stay in the spirit of Heaton and Lucas (1996) and Aiyagari (1994) and calibrate the income processes to have less than full persistence. We do this in the simplest possible way by using a two-state Markov process as representing employment and unemployment. Our calibrated model also has aggregate risk, which we calibrate by letting the value of the technology shock be 1.01 in good times and 0.99 in bad times (the model in this section is calibrated to quarterly data). The unemployment rate is chosen to equal 0.04 in good times and 0.1 in bad times. These values for the technology shock and the unemployment rate lead to fluctuations in the macroeconomic aggregates, which have roughly the same magnitude as the fluctuations in observed postwar U.S. time series. The amount of home-produced goods (the unemployment insurance) is equal to around 9% of the average employed wage. This parameter plays the role of making bad shocks less bad, and hence agents are less averse to holding low levels of assets, which is a significant help in matching the left tail of the asset distribution. We select the values of the stochastic process for (z, ϵ) so that the average duration of both good and bad times is eight quarters and so that the average duration of an unemployment spell is 1.5 quarters in good times and 2.5 quarters in bad times. We also impose $\pi_{gb00}\pi_{gb}^{-1} = 1.25\pi_{bb00}\pi_{bb}^{-1}$ and $\pi_{bg00}\pi_{bg}^{-1} = 0.75\pi_{gg00}\pi_{gg}^{-1}$.[18]

3.2.2. Numerical implementation. As in Krusell and Smith (1996b), the computational methods here focus on finding stationary stochastic equilibria only. A stationary stochastic equilibrium is a recursive equilibrium described by an ergodic set $\mathcal{D}$ of distributions (i.e., a set such that once there, the economy never leaves the set) and an invariant probability measure $\mathcal{P}$ over this set. This means that we will be looking for functions H and q that approximate their true counterparts over $\mathcal{D}$.

Our earlier work explored the similarities of marginal propensities to consume across agents in this economy. When marginal propensities are the same, equilibrium prices depend only on total wealth, and not on its distribution. We outlined an algorithm with the following key elements:

1. Agents perceive that prices depend only on a subset of the moments of Γ and that the laws of motions for these moments depend only on this same set of moments.
2. The resulting behavior for individual savings is simulated over a long period of time, and it is thus possible to compare the behavior of the moments in question.
3. If the simulated time series for the moments are very close to those perceived by the agents, the obtained behavior is argued to reflect very small deviations from perfectly rational expectations, and we thus would have a candidate approximate equilibrium.

It turned out in the set of models we studied that it was sufficient with one moment to obtain remarkably small prediction errors.

In the present setup, we follow our original algorithm as closely as possible, because it is, *a priori*, likely that only the mean will matter for prices in this environment as well. In particular, one interpretation of our original results is that the idiosyncratic uninsurable risk that gives rise to the deviations from aggregation leads to only very little dispersion in the sensitivity to risk among poor and rich agents. Here, the idiosyncratic risk is likely to be even better protected against, because there is a larger set of assets available for insurance.

To simplify presentation, we thus describe our computational algorithm with Γ represented by one moment only; the inclusion of more moments is conceptually straightforward. In addition, we simplify here by assuming that $\tilde{\beta}$ is deterministic and the same for all agents. Thus, let H and q be represented through

$$\log \bar{k}' = \begin{cases} a_0 + a_1 \log \bar{k} & \text{if } z = z_g \\ b_0 + b_1 \log \bar{k} & \text{if } z = z_b, \end{cases} \tag{5}$$

$$q = \begin{cases} c_0 + c_1 \log \bar{k} & \text{if } z = z_g \\ d_0 + d_1 \log \bar{k} & \text{if } z = z_b. \end{cases} \tag{6}$$

Further, define the following problem for an individual agent:

$$v(\omega, \epsilon; \bar{k}, z) = \max_{k', b'} \{U(\omega - k' - qb') + \tilde{\beta} E[v(\omega', \epsilon'; \bar{k}', z') \mid \epsilon, z]\} \tag{7}$$

subject to (2), (4), (5), and (6). Now, the most straightforward extension of our earlier methods would be to iterate on H and q by solving (7) and generating time series for $\bar{k}$ and total bondholdings. The hope would be that there is a vector (a_0, a_1, b_0, b_1, c_0, c_1, d_0, d_1) such that there is a close fit for $\bar{k}$ and such that total bondholdings almost always are close to zero. Unfortunately, this procedure does not work in our economy. The problem is that total bondholdings turn out to follow something close to a autoregressive process with a coefficient just below 1; this means that even though total bondholdings can be set to sum to zero on average over time by choice of the mentioned vector, they will display large, long swings, thus not enabling anything close to market clearing at all dates and states.[19]

To achieve market clearing in bonds, we use an alternative procedure. This procedure is of independent interest, because it can be applied to any kind of additional market-clearing condition in this type of model.[20] Assume that agents solve the following problem at each point in time:

$$\tilde{v}(\omega, \epsilon; \bar{k}, z, q) = \max_{k', b'} \{U(\omega - k' - qb') + \beta E[v(\omega', \epsilon'; \bar{k}', z') \mid \epsilon, z]\} \tag{8}$$

subject to (2), (4), (5), and (6). Thus, agents make portfolio choices based on an arbitrary current value q for the bond price, and perceive their indirect utility of wealth to be given by the v function defined in problem (7). That is, agents take the current price to equal q and perceive future prices to be given by the function (6). This problem will generate portfolio decision rules $\tilde{f}^k$ and $\tilde{f}^b$ which depend explicitly on the value for q. For any given distribution Γ, it then typically will be possible to find at least one value q that clears bond markets exactly. We thus employ the following algorithm:

(1) Solve problem (7) given a vector $(a_0, a_1, b_0, b_1, c_0, c_1, d_0, d_1)$ representing the law of motion for total capital and the bond-pricing function.
(2) Simulate using problem (8):
  (a) Fix an initial wealth/employment distribution for a large number of agents and an initial value for z.
  (b) Solve problem (8) and iterate on q until the bond market clears exactly.
  (c) Find next period's wealth/employment distribution by using the decision rules just obtained and by drawing new values for the shocks.
  (d) Repeat a large number of times, obtaining a long times series, of which the first part is discarded.
(3) Use the obtained time series to regress $\log \bar{k}_{t+1}$ and q_t on constants and $\log \bar{k}_t$ for each value of z.
(4) Use some metric to compare the coefficient estimates to those taken as given by the agents. If they are the same, proceed to the next step in the algorithm. If not, update the coefficient vector and go to step 1.
(5) Upon convergence of the parameter vector, goodness-of-fit statistics can be used to evaluate how far the individual's expectations are from being perfectly rational. If the fit is not satisfactory, use different functional forms for H and q and/or add more moments.

Before we proceed to the results, let us define what we mean by the equity premium, conditional on the current state variables $\bar{k}$ and z, in this setup. In any given period, it is given by

$$P(z' = z_g \mid z)(r(\bar{k}', \bar{l}', z_g) + 1) + P(z' = z_b \mid z)(r(\bar{k}', \bar{l}', z_b) + 1) - q^{-1},$$

where $\bar{k}'$ is determined by (5) and q is the market-clearing bond price in the given period.[21] This definition of the equity premium reflects the fact that the value of the firm is equal to the value of the aggregate capital stock, whose price per unit, expressed in units of the current consumption good, is 1.

3.2.3. Solution and simulation parameters. We solve the consumer's problem (7) by computing an approximation to the value function on a grid of points in the state space. We use cubic spline and polynomial interpolation to compute the value function at points not on the grid. This numerical algorithm is similar to one used by Krusell and Smith (1996b). The Appendix describes the algorithm in detail and discusses its accuracy. The Appendix also describes the approach that we use to solve problem (8) and to find the market-clearing bond price. In our simulations we include 10,000 agents and 3,500 time periods, where we discard the first 1,000 periods. The initial wealth distribution in each of the simulations is a typical wealth distribution from an economy in which the only asset is productive capital [such economies are examined in detail by Krusell and Smith (1996b)].

3.2.4. Results for the model without preference shocks.

3.2.4.1. Predictions for quantities and prices. We first present the model without preference shocks. We consider two different borrowing constraints for bonds: The extreme case of no allowed borrowing in bonds—$\underline{b} = 0$—and the more generous $\underline{b} = -2.4$, which amounts to around half of an agent's average annual income. Short sales are not allowed for capital: $\underline{k} = 0$. The equilibrium laws of motion for aggregate capital are as follows (we report the case of a loose bond constraint only; the equilibrium laws of motion, and their fits, are quite similar across the models here and those considered later):

$$\log \bar{k}' = 0.092 + 0.963 \log \bar{k},$$
$$R^2 = 0.999999$$

in good times and

$$\log \bar{k}' = 0.082 + 0.965 \log \bar{k},$$
$$R^2 = 0.999999$$

in bad times. Using our simulated sample (consisting of 2,500 time periods) we plot tomorrow's aggregate capital against today's aggregate capital. This graph (see Figure 1) is a clear illustration of the fit.

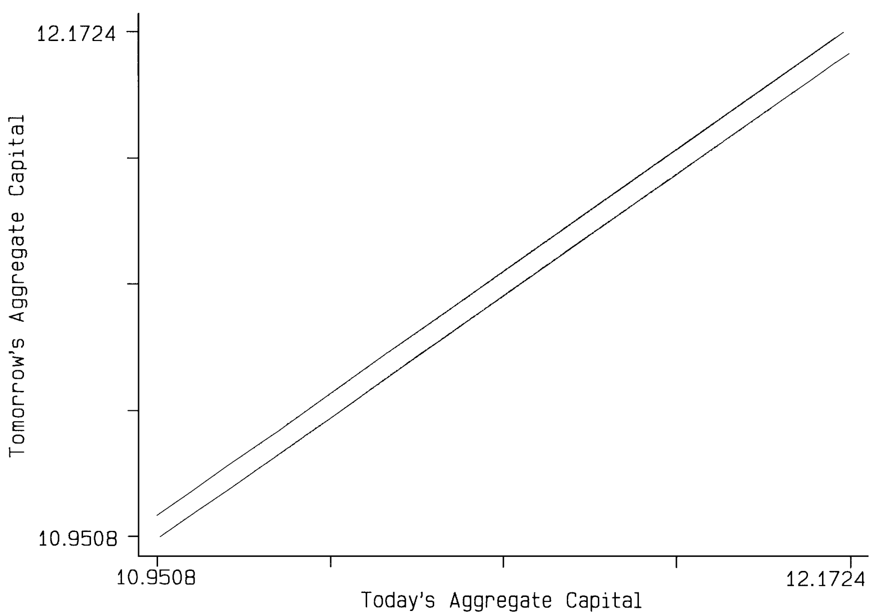

FIGURE 1. Tomorrow's vs. today's aggregate capital: top line = good aggregate state, bottom line = bad aggregate state.

The bond pricing function also can be approximated very closely using z and $\bar{k}$ alone:

$$q = 0.899 + 0.0519 \log \bar{k} - 0.006(\log \bar{k})^2,$$
$$R^2 = 0.99999999$$

in good times and

$$q = 0.904 + 0.0502 \log \bar{k} - 0.006(\log \bar{k})^2,$$
$$R^2 = 0.99999999$$

in bad times. Using our simulated sample, we plot today's market-clearing bond price against today's aggregate capital. Again, the fit is excellent. See Figure 2.

Our earlier paper [Krusell and Smith (1996b)] contains a careful examination of the robustness of the finding that there is "approximate aggregation" in the sense that the law of motion for aggregate capital depends only slightly on the distribution of the wealth. This paper also contains a detailed discussion of the reasons underlying this result. Suffice it to say here that the main arguments carry over to this model, i.e., the effective insurance achieved with only one asset in this class of models is excellent in utility terms (although individuals' consumptions vary significantly, indeed much more than does aggregate consumption). Another asset, of course, only strengthens the argument.

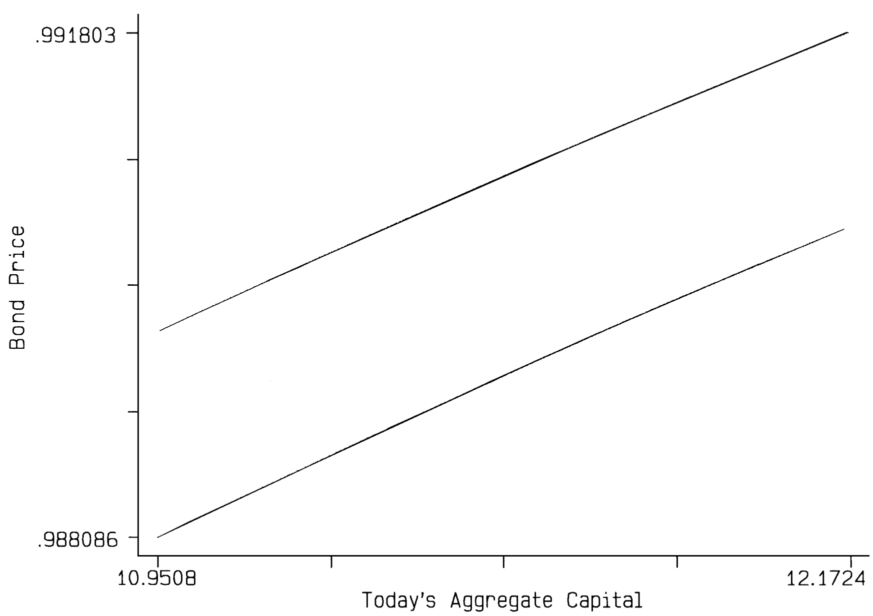

FIGURE 2. Bond price vs. aggregate capital: top line = bond return in good times, bottom line = bond return in bad times.

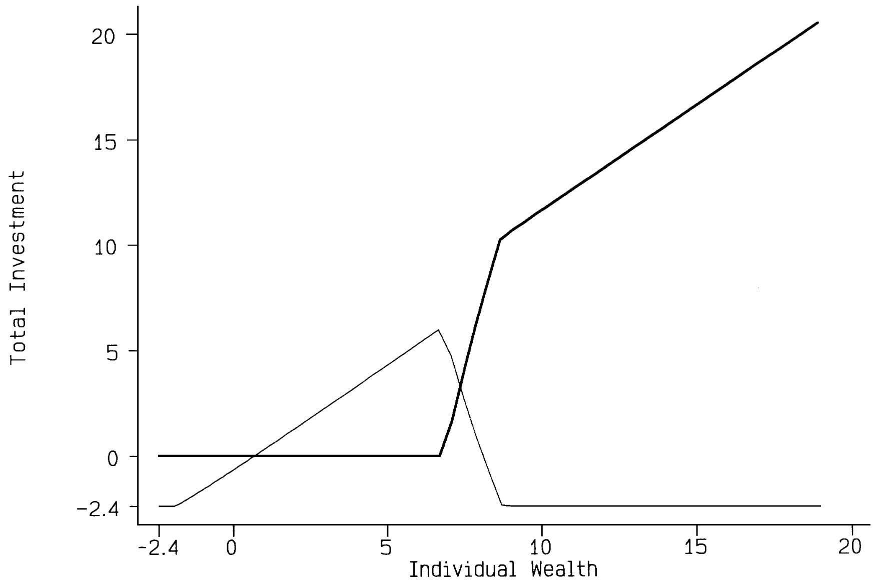

FIGURE 3. Portfolio decision rules of an employed agent: thick line = capital, thin line = bonds.

In the economy with tight constraints, all agents are at the zero-bond constraint, but few agents are constrained in capital. In contrast, when the bond constraint is relaxed to −2.4, we see many agents piling up at each of the constraints for capital and bonds (about a quarter of the population at the former and more than half at the latter). This finding comes from the extreme nature of portfolio choices of agents in this model: Only about 10% of the population are at interior portfolio solutions (on average, 25% of the population is against the short-sales constraint on capital, whereas 65% of the population is against the borrowing constraint on bonds). Poor/unfortunate agents hold bonds and go as short in capital as they can (to zero in this case), whereas rich/fortunate agents borrow as much as they can in bonds and put all their savings into capital. Figure 3 illustrates the portfolio choice of a typical employed individual and Figure 4 illustrates the portfolio choice of a typical unemployed individual (the figures are drawn for a given set of values for the aggregate state variables). A key feature of these figures is that only employed agents within a narrow wealth range have interior portfolio decisions.

Some summary equilibrium statistics are reported in Table 2.[22] The reported statistics are sample means computed using our simulated sample consisting of 2,500 time periods.

Table 2 shows that the equity premia generated by our framework are not much closer to those we see in the data than are those in other studies (recall that the average postwar equity premium is a little below 2% on a quarterly basis).[23] The variation in the equity premium across good and bad aggregate states is not

TABLE 2. Aiyagari-style model without preference shocks

Borrowing constraints on bonds and capital	Risk-free rate, %		Equity premium, %		Mean capital average
	Good times	Bad times	Good times	Bad times	
$\underline{b} = 0$, $\underline{k} = 0$	1.04	0.90	0.0181	0.0116	11.66
$\underline{b} = -2.4$, $\underline{k} = 0$	1.06	0.93	0.00015	0.00015	11.65

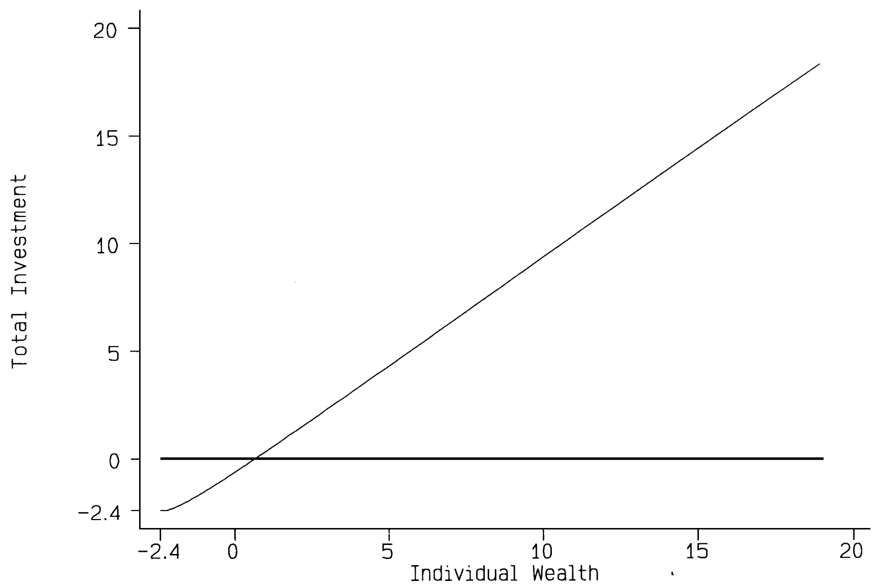

FIGURE 4. Portfolio decision rules of an unemployed agent: thick line = capital, thin line = bonds.

substantial. A tightening of the borrowing constraint, however, does lead to a large increase in the equity premium in relative terms: in particular, the equity premium increases by two orders of magnitude when no borrowing is allowed.

3.2.4.2. Equity premium and market price of risk: a theoretical digression. The analytical tools developed by Hansen and Jagannathan (1991) can be used to gain a better understanding of the determination of the equity premium in our model.[24] Consider an agent with an interior portfolio solution.[25] The Euler equations for such an agent imply that

$$E_{z',\epsilon' \mid z,\epsilon,I}[m'(R'_e - R_f)] = 0, \qquad \textbf{(9)}$$

where $E_{x \mid y}$ represents integration with respect to the distribution of x for a given value of y, where x and y can be vectors. In (9), $R_f = q^{-1}$ is the risk-free rate in the current period, R_e' is the realized gross return on capital (net of depreciation) in the next period, $m' \equiv \beta U'(c')/U'(c)$ is the agent's intertemporal marginal rate of substitution, and I contains the individual and aggregate state variables relating to wealth ((ω, Γ), or $(\omega, \bar{k})$ in our computational procedure). Now,

$$\begin{aligned} E_{z',\epsilon' \mid z,\epsilon,I}[m'(R_e' - R_f)] &= E_{z' \mid z,\epsilon,I}\{E_{\epsilon' \mid z',z,\epsilon,I}[m'(R_e' - R_f)]\} \\ &= E_{z' \mid z,\epsilon,I}\{(R_e' - R_f)[E_{\epsilon' \mid z',z,\epsilon,I}(m')]\}, \end{aligned}$$

where we are using, in turn, the definition of conditional expectation and the fact that the excess return on equity does not depend on the idiosyncratic shock ϵ'.[26] This means that equation (9) can be rewritten as

$$E_{z' \mid z,\epsilon,I}\{[E_{\epsilon' \mid z',z,\epsilon,I}(m')](R_e' - R_f)\} = 0. \tag{10}$$

This equation is at the core of asset pricing in incomplete-markets economies. As in the representative-agent model, it involves the expectation of a product of the excess return and a factor reflecting intertemporal marginal rates of substitution. However, in the incomplete-markets model, the entire expectation is individual-specific, and the intertemporal marginal rate of substitution factor is an *average of future states* for the individual under consideration.[27]

Following Hansen and Jagannathan (1991), the definition of covariance can be used to rewrite equation (10) as

$$\begin{aligned} &E_{z' \mid z,I}(R_e' - R_f) = \sigma_{z' \mid z,I}(R_e' - R_f) \\ &\quad \times \left\{ -\mathrm{corr}_{z' \mid z,\epsilon,I}[R_e' - R_f, E_{\epsilon' \mid z',z,\epsilon,I}(m')] \frac{\sigma_{z' \mid z,\epsilon,I}[E_{\epsilon' \mid z',z,\epsilon,I}(m')]}{E_{z' \mid z,\epsilon,I}[E_{\epsilon' \mid z',z,\epsilon,I}(m')]} \right\}, \end{aligned} \tag{11}$$

where $\mathrm{corr}_{x \mid y}$ denotes conditional correlation and $\sigma_{x \mid y}$ denotes conditional standard deviation (both with respect to the distribution of x, given y). The conditional expected equity premium therefore can be viewed as the product of two terms, the first being the conditional standard deviation of the equity return and the second being what we define to be the market price of risk. In our setup, by virtue of the two-state process for z and the fact that the marginal utility of consumption (averaged across the employment outcome) is low when the aggregate productivity shock (and hence the return on equity) is high, the correlation that appears in equation (11) is exactly -1.[28] In the particular case of two aggregate states, thus, we have that

$$\frac{E_{z' \mid z,I}(R_e' - R_f)}{\sigma_{z' \mid z,I}(R_e' - R_f)} = \frac{\sigma_{z' \mid z,\epsilon,I}[E_{\epsilon' \mid z',z,\epsilon,I}(m')]}{E_{z' \mid z,\epsilon,I}[E_{\epsilon' \mid z',z,\epsilon,I}(m')]}, \tag{12}$$

i.e., the market price of risk simply equals the ratio of any unconstrained agent's (conditional) standard deviation of the *average value* of m' divided by his (conditional) expectation of this same object. By construction, this ratio is the same for all unconstrained agents.

Equation (12) is very useful for analyzing the determination of the equity premium. To increase the equity premium in our model economies, one must either increase the volatility of the equity return or increase the market price of risk (or both). We find that our incomplete-markets model (with or without borrowing) displays slightly less volatility in aggregate capital (and hence less volatility in the equity return) than does its representative-agent (complete-markets) counterpart. Thus, for the incomplete-markets model to give an increase in the equity premium, it must have a higher market price of risk.

3.2.4.2.1. What determines the market price of risk? Equation (11) focuses the attention on the time-series properties of $E_{\epsilon' \mid z', z, \epsilon, I}[m']$, i.e., on the average value of an individual's marginal rate of substitution across realizations of the idiosyncratic risk, conditional on the aggregate state. It therefore can be misleading to focus on the amount of idiosyncratic risk as the sole determinant of the market price of risk. To illustrate this point, note that because $\sigma_{z' \mid z, \epsilon, I}[E_{\epsilon' \mid z', z, \epsilon, I}(m')] < \sigma_{z', \epsilon' \mid z, \epsilon, I}(m')$, it follows that

$$\frac{\sigma_{z' \mid z, \epsilon, I}[E_{\epsilon' \mid z', z, \epsilon, I}(m')]}{E_{z' \mid z, \epsilon, I}[E_{\epsilon' \mid z', z, \epsilon, I}(m')]} < \frac{\sigma_{z', \epsilon' \mid z, \epsilon, I}(m')}{E_{z', \epsilon' \mid z, \epsilon, I}(m')}.$$

The ratio on the right-hand side of this equation often has been used in the literature [see, e.g., Weil (1992) or Telmer (1993)] to obtain an upper bound on the market price of risk implied by a given economic model. By using conditioning information, we obtain the tighter bound indicated in equation (12): In our economy with two aggregate states, the bound is an equality; when there are more aggregate states, it is an upper bound [see Cochrane and Hansen (1992) for a similar use of conditioning information]. In contrast, the bound $\sigma_{z', \epsilon' \mid z, \epsilon, I}(m')/E_{z', \epsilon' \mid z, \epsilon, I}(m')$ is not a good indicator of the potential magnitude of the market price of risk in a model with idiosyncratic risk.[29] Whereas $\sigma_{z', \epsilon' \mid z, \epsilon, I}(m')$ clearly increases with the amount of idiosyncratic risk, thus suggesting a potentially higher market price of risk, so does the difference between the looser and the tighter bound, because $\mathrm{Var}_{z', \epsilon' \mid z, \epsilon, I}(m') - \mathrm{Var}_{z' \mid z, \epsilon, I}[E_{\epsilon' \mid z', z, \epsilon, I}(m')] = E_{z' \mid z, \epsilon, I}[\mathrm{Var}_{\epsilon' \mid z', z, \epsilon, I}(m')]$! In our model economies, for example, the looser bound is as much as eight times larger than the tighter bound. These results help explain how it is possible that the market price of risk remains small in our model economies despite the fact that, unlike in many previous models in the literature, individual consumption is much more volatile than aggregate consumption (in the model without preference heterogeneity, for example, the standard deviation of individual consumption is about four times that of aggregate consumption).

Given these arguments, what would constitute a sufficient condition for an increase in the market price of risk? At this point, it is helpful to recall Mankiw's (1986) analysis. In our model, as in Mankiw's, two features of the environment are crucial for understanding the qualitative effects of idiosyncratic risk on the equity premium: the convexity of marginal utility and the relative standard deviations of the idiosyncratic shock in the good and bad states. To see why, suppose for the moment that there is no idiosyncratic risk, so that there is no cross-sectional variation in consumption levels. Let c be the consumption of a typical agent in the current period and let c' be the consumption of a typical agent in the next period; assume that c' takes on the value c_g if the good state obtains tomorrow and c_b if the bad state obtains tomorrow. The marginal utility of consumption tomorrow therefore takes on the two values $U'_g \equiv U'(c_g)$ and $U'_b \equiv U'(c_b)$. Letting π be the conditional probability of the good state occurring tomorrow, it is easy to see that $E_{z' \mid z,\bar{k}}[U'(c')] = \pi U'_g + (1-\pi)U'_b$ and $\sigma_{z' \mid z,\bar{k}}[U'(c')] = \pi^{1/2}(1-\pi)^{1/2}(U'_b - U'_g)$, where we have assumed that $c_g > c_b$, in which case $U'_b - U'_g > 0$. The market price of risk therefore can be written

$$\frac{\sigma_{z' \mid z,\bar{k}}\left[\dfrac{\beta U'(c')}{U'(c)}\right]}{E_{z' \mid z,\bar{k}}\left[\dfrac{\beta U'(c')}{U'(c)}\right]} = \frac{\sigma_{z' \mid z,\bar{k}}[U'(c')]}{E_{z' \mid z,\bar{k}}[U'(c')]} = \frac{\pi^{1/2}(1-\pi)^{1/2}(U'_b - U'_g)}{\pi U'_g + (1-\pi)U'_b}. \qquad \textbf{(13)}$$

As long as $U'_b - U'_g > 0$, this expression is increasing in the value of U'_b. Now imagine introducing a mean-preserving spread in consumption levels *in the bad state only*.[30] As one can see from equation (11), to compute the market price of risk in this setup, we need to replace U'_b in equation (13) with the *average* marginal utility of consumption in the bad state, and this average is greater than U'_b if and only if U' is convex. Here, because we are using constant-relative-risk-aversion preferences, the marginal utility is indeed globally convex. Introducing mean-preserving idiosyncratic variation in consumption in the bad state therefore has the effect of increasing the average marginal utility of consumption in the bad state, thereby increasing the market price of risk as expressed in equation (11). A similar argument holds if one introduces mean-preserving spreads in consumption in both the good and bad states in such a way that the average marginal utility of consumption in the bad state increases by more than the average marginal utility of consumption in the good state.[31]

In our economy the cross-sectional standard deviation of labor income is (ignoring variations in the wage rate) roughly 50% larger in the bad state than in the good state. Although individuals do smooth consumption in response to idiosyncratic shocks, the higher cross-sectional variation in labor income in the bad state causes the cross-sectional dispersion of intertemporal marginal rates of substitution to be larger in the bad than in the good state.[32]

3.2.4.2.2. Market price of risk: quantitative findings. Having established the role and determinants of the market price of risk in our economy, what are its quantitative properties in the computed equilibrium? In the incomplete-markets model with a loose borrowing constraint, the market price of risk is around 0.0025 (averaged across a typical realization for the aggregate state) whereas in the representative-agent version of our model it is 0.0022. In the model without borrowing, the market price of risk increases dramatically to around 0.211. The increase in the market price of risk therefore explains why the equity premium in the model without borrowing is so much larger than the one in the model with borrowing. Similarly, the equity premium is only slightly higher in the incomplete markets with borrowing than in the representative-agent model.

As shown by Hansen and Jagannathan (1991), observed asset returns can be used to provide a lower bound on the actual market price of risk. For example, as reported by Lettau and Uhlig (1997), quarterly returns data for a market portfolio such as the S&P 500 yield a lower bound on the market price of risk of about 0.27. Our model without borrowing therefore generates a market price of risk that is reasonably close to the one in the data.[33]

It is also possible to understand why the borrowing constraint is so important for the market price of risk in our model. As explained in note 21, asset prices in the no-borrowing economy are determined by the agent with the highest subjective evaluation of the bond payoff. This is an unemployed agent who is just indifferent between holding no capital and holding positive amounts of capital. Such an agent has very low wealth today, and so, the agent's consumption in the next period is very sensitive to the outcome of the employment shock. Moreover, because the employment shock is persistent, this agent is much more likely to be unemployed in the next period if the bad aggregate state rather than the good aggregate state obtains. Therefore, conditional on the realization of the aggregate state, the variance of this agent's consumption (looking across realizations of the employment outcome) will be much higher in the bad aggregate state than in the good aggregate state. From equation (13), it is clear that in this case the market price of risk will be very high.

On the other hand, when significant borrowing is permitted, it turns out that the agents who determine asset prices (i.e., the agents with interior portfolio decisions) are employed and have much higher wealth than in the economy without borrowing. This makes the future marginal utilities fluctuate much less: First, the persistence of the employment shock implies that, for these agents, the probability of unemployment in the next period is only moderately higher in the bad state than in the good state. Second, the higher wealth of these agents provides them with insurance against bad realizations of the employment shock, which is quite effective in utility terms. This last point goes back to the analyses of Lucas (1987), Cochrane (1989), Atkeson and Phelan (1994), Krusell and Smith (1996a,b), and Tallarini (1996): Within this class of models, agents can achieve insurance that is near perfect in utility terms without making it near perfect in consumption terms. It thus would take an order-of-magnitude higher consumption variability to increase the market price of risk in this economy.

3.2.4.3. Predictions for the distribution of wealth. Table 3 has the implications of the model for the (average) shape of the wealth distribution (wealth in the table refers to asset wealth).

TABLE 3. Distribution of wealth

Model	% of wealth held by top 1%	5%	10%	20%	30%	Fraction with wealth <0, %	Gini coefficient
$\underline{b}=0$	3	11	20	35	47	0	0.26
$\underline{b}=-2.4$	3	13	23	39	52	0.5	0.33
Data	30	51	64	79	88	11	0.79

There is a large gap between the models and the data.[34] The Gini coefficients are much too low in the model, which reflects a failure to match the data at both ends of the distribution: The rich do not save enough, and the poor save too much. Note, finally, from this table that looser constraints allow the wealth distribution to spread out somewhat. In addition, the positive rate-of-return difference in portfolios between rich and poor contributes to the distribution of wealth spreading out, but this effect is not a major one because the equity premium is small.

Explanation of the extreme relative skewness of the wealth distribution is an open question. In what follows, we adopt a relatively simple structure that has the ability to generate a more dispersed wealth distribution for a given distribution of income shocks. For a more detailed discussion of different ways in which the wealth distribution can be matched, see our earlier paper and Quadrini and Ríos-Rull (1996).

3.2.5. Results for the model with preference shocks. We now turn to the model with preference shocks. As will be apparent, this model is capable of roughly matching the wealth distribution, and it is therefore the model whose asset pricing implications we argue should be taken the most seriously.

The preference parameters are chosen in line with our earlier paper: We subject the experiment to the requirements that the differences in discount factors are not large, and that their distribution is symmetric around its mean. More precisely, we assume that $\tilde{\beta}$ can take on three values—0.9858, 0.9894, and 0.9930—and that the transition probabilities are such that (1) the invariant distribution for β's has 80% of the population at the middle β and 10% at each of the other $\tilde{\beta}$'s; (2) immediate transitions between the extreme values of $\tilde{\beta}$ occur with probability zero; and (3) the average duration of the highest and lowest $\tilde{\beta}$'s is 50 years. We choose the latter number to match roughly the length of a generation, because we view the model as capturing some elements of an explicit overlapping-generations structure with altruism (i.e., parents care about the utility of their children) and less than perfect correlation in genes between parents and children (i.e., there is "regression to the mean" in the rate at which the current generation discounts the utility of future generations).

We omit the laws of motion for aggregate capital and the bond functions; they are similar to our previous case, and the fit is still excellent. In particular, although

TABLE 4. Distribution of wealth

	% of wealth held by top					Fraction with	Gini
Model	1%	5%	10%	20%	30%	wealth <0, %	coefficient
$\underline{b} = 0$	20	46	61	74	79	0	0.66
$\underline{b} = -2.4$	23	55	73	87	92	11	0.82
Data	30	51	64	79	88	11	0.79

marginal propensities to save are now more dissimilar across agents, almost all wealth is now held by richer agents among whom marginal propensities do not differ much, and this group in effect determines total wealth accumulation.[35] As before, the equilibrium portfolio choices are largely characterized by corner solutions in the $\underline{b} = -2.4$ case, and about 4% of the population have interior solutions (on average, about 58% of the population is against the short-sales constraint for capital and about 38% of the populations is against the borrowing constraint for bonds). For each type of agent, as indexed by current $\tilde{\beta}$ and employment status, there is an intermediate range of wealth levels for which an interior solution obtains; for lower values, the agent holds no capital, and for higher values, the agent is at the lower bound for bonds. For unemployed agents, the wealth ranges with interior solutions occur at very high levels of wealth, whereas for employed agents they occur at fairly low levels of wealth.

Table 4 summarizes the wealth distributions in our models with preference heterogeneity (as in the model with a constant $\tilde{\beta}$, we consider one version with $\underline{b} = 0$ and one with $\underline{b} = -2.4$).

The differences between these and our previous results are striking. The Gini coefficients go up significantly to the range observed in U.S. data; there is a large concentration of agents at low (including negative) levels of asset holdings; and although the fraction of wealth held by the 1% richest is still not quite as high as in the data, it is close. Note also that the effect of the borrowing constraint is quite visible: When negative values for bonds are allowed, it leads to an increase in the steady-state Gini coefficient from 0.66 to 0.82.

As discussed by Krusell and Smith (1996b), the behavior of the model without preference heterogeneity is very similar to the behavior of its complete-markets (representative-agent) counterpart. Thus it is perhaps not too surprising that the model without preference heterogeneity does a poor job of explaining observed asset prices. The model with preference heterogeneity, however, behaves very differently in some respects than a complete-markets (representative-agent) model. In particular, the model with preference heterogeneity displays significant departures from permanent-income behavior (the consumption-output correlation, for example, is much higher in the model with preference heterogeneity than in the model without preference heterogeneity). This model therefore holds some promise of doing a better job of matching asset prices. The model's implications for asset prices are contained in Table 5.

TABLE 5. Aiyagari-style model with preference shocks

Borrowing constraints on bonds and capital	Risk-free rate, %		Equity premium, %		Mean capital average
	Good times	Bad times	Good times	Bad times	
$\underline{b} = 0$, $\underline{k} = 0$	1.00	0.85	0.0183	0.0117	11.91
$\underline{b} = -2.4$, $\underline{k} = 0$	1.02	0.89	0.00017	0.00016	11.84

Although the wealth distribution in this version of the Aiyagari model is quite realistic, asset prices are not: Equity premia are minute regardless of the tightness of borrowing constraints. As in the model without preference heterogeneity, the large difference in the equity premia between the models with and without borrowing can be explained by differences in the market price of risk, which is (on average) about 0.214 in the model without borrowing and about 0.0026 in the model with borrowing.[36] The model with preference heterogeneity (and a realistic wealth distribution) therefore leads to small increases in both equity premia and market prices of risk relative to the model without preference heterogeneity (and an unrealistic distribution of wealth). Although these are steps in the right direction, the model with preference heterogeneity and borrowing falls well short of the data in its asset-pricing behavior.

4. CONCLUSIONS

In this paper we study a class of asset-pricing models based on partially uninsurable income risk and wealth heterogeneity. Compared to the existing studies of similar setups, we emphasize the importance of focusing on the models' implications for heterogeneity: Unless the heterogeneity that the model generates is quantitatively reasonable, we do not think the asset-pricing implications can be taken very seriously. When this principle is applied to the models with unit roots in (log) labor income that we look at here, the associated asset prices indeed need to be taken with a big grain of salt, because the asset distributions in these models tend to be collapsed: All agents hold the same (or nearly the same) amounts of assets. In the data, by contrast, assets are distributed across agents in a highly skewed fashion. We thus argue the need to move on to models with more realistic asset distributions, and we consider one such model in this paper.

The implications for asset prices in the models we are willing to take seriously are quite consistent with the findings in the earlier papers by Telmer (1993), Den Haan (1996a), and Heaton and Lucas (1996): We are far from being able to explain asset prices. In this sense, our news is not good, or, to put it in less negative terms, our news is really no major news. The little bit of news that we provide concerns the market price of risk: In our model, tight borrowing constraints can lead to a large market price of risk.

Does our paper close the door to all explanations for the large observed average equity premium that are based on heterogenous agents and incomplete insurance?

Given our findings regarding the importance of borrowing constraints, one direction for future research is to introduce other kinds of frictions, such as costs of adjusting aggregate capital, as in Cochrane (1991) and Christiano and Fisher (1995), or transactions costs, as in Aiyagari and Gertler (1991) and Heaton and Lucas (1995, 1996). Adjustment costs, in particular, make capital less useful relative to bonds as an asset for buffering both aggregate and idiosyncratic shocks, so that the equity premium has to increase to induce consumers to hold capital.[37]

The model that we construct to match the wealth distribution—a model with randomness in agents' preferences—is not the only possible model that can achieve this goal. The exploration of alternative frameworks for studying asset prices and wealth distribution has really only begun. Here as well, exploring frictions and incorporating more institutional detail may prove fruitful.

NOTES

1. Den Haan (1996b) also looks at models with a continuum of agents.

2. This result could perhaps have been anticipated from the earlier work by Deaton (1991), who considered the decision problem faced by a consumer living in a Constantinides and Duffie world (for the case of a constant interest rate). Deaton showed that under certain conditions such a consumer has a tendency to deplete all asset holdings and set consumption equal to current income forever after. Constantinides and Duffie close the partial-equilibrium Deaton model by setting the interest rate so that consumers behave in exactly this way, in the case in which there is no outside asset, or so that agents draw down on their holdings of the aggregate asset exactly to the point where they all hold the per-capita stock of this asset.

3. Ríos-Rull (1994) studies an overlapping-generations model with production and with consumers who live for 55 periods. He considers only aggregate risk and computes asset prices for various assumptions on to what extent this risk can be shared among agents of different ages.

4. The portfolio-choice implications from our baseline model are that most agents specialize completely, i.e., they hold either only bonds or only capital; and poor (particularly unemployed) agents hold bonds, and less poor agents hold capital. At each point in time, there is only a very narrow band of wealth levels (which depends on the agent's employment status) such that the agent does not choose a corner solution for his portfolio problem.

5. See Table II of Telmer (1993).

6. Note that in this case total resources are equal to $y + ug$, where u is the (state-dependent) unemployment rate.

7. Because the unemployment rate takes on only two values, we have effectively assumed a law of large numbers for idiosyncratic shocks. The total supply of labor is therefore $1 - u_g$ in good times and $1 - u_b$ in bad times.

8. This definition is circular to some extent, because it may happen that if a variable is included, it will have independent impact on prices, whereas if it is not, there still exists an equilibrium. When there are multiple equilibria in this sense, it is less obvious how to characterize the whole set of equilibria.

9. Because aggregate bondholdings must equal zero,

$$\bar{\omega} \equiv \int \omega \, d\Gamma = (r(\bar{k}, z) + 1)\bar{k} + (1 - u(z))w(\bar{k}, z) + u(z)g.$$

Given $\bar{\omega}$ and z, this equation has a unique solution for $\bar{k}$.

10. Other papers that study dynamic general equilibrium economies with incomplete markets include Aiyagari (1994), Castañeda et al. (1994), Coleman and Liu (1994), and Huggett (1995, 1996).

11. An alternative would be to have negative drift in the process for log labor income.

12. This particular assumption is also an effort to stay within a framework that is as close to that of Constantinides and Duffie (1996) as possible.

13. In his experiments, Aiyagari lets ρ vary from 0 to 0.9 and maintains a constant unconditional standard deviation of log income. We keep conditional variance constant to isolate the effects of changing the degree of persistence in individual income.

14. With individuals dying at a constant rate, the evolution of total income also depends on the starting income of newborns and on the death rate.

15. In this section, as in Constantinides and Duffie (1996), we assume that the period utility function takes the form $U(c) = (1-\sigma)^{-1}c^{1-\sigma}$, where $\sigma > 0$ is the coefficient of relative risk aversion.

16. There is, at least implicitly, a borrowing constraint in the Constantinides and Duffie model; otherwise, agents could run up a debt of any size to provide for complete insurance or a Ponzi scheme. The loosest possible constraint allowing the agent to pay off any debt almost surely equals $-\underline{I}q/(1-q)$, where $\underline{I}$ is defined as the lowest labor income realization [see Aiyagari (1994) for details]. That is, in the Constantinides and Duffie formulation, the loosest possible constraint on the total value of asset holdings is that it never go below $dq/(1-q) = d/(d/p) = p$. In our formulation, this lower bound is simply a nonnegativity requirement, because we have $\underline{I} = 0$.

17. One could also allow for heterogeneity in asset holdings in the Constantinides and Duffie model by letting $I_{it} = s_i(e^{z_{it}} - d)$ (we thank John Heaton for suggesting this possibility). This formulation, however, is very hard to reconcile with some of the key features of the wealth and income data: About 10% of the U.S. population has negative net worth, and the earnings distribution is much less dispersed than is the wealth distribution.

18. Because of the high computational costs (especially in the two-asset version of the model), we do not report results for values of risk aversion greater than 1. In Krusell and Smith (1996b), we find that higher values of risk aversion tend to strengthen the approximate aggregation result. Higher values of risk aversion would clearly make the asset prices in our model economies look better relative to the data. However, in light of the findings reported in this paper, it seems clear that unless risk aversion is very high, the asset-pricing implications of our model will continue to be very far from the data (provided borrowing constraints are not severe).

19. This form for total bondholdings derives from the fact that the individual's log savings are close to linear in log wealth with a coefficient just below 1 in this kind of growth model; see Krusell and Smith (1996b).

20. For example, although we did not use this procedure in Krusell and Smith (1996b) to solve the model in the case where labor supply is endogenous, it could have been applied there.

21. Note that this is the conditional expected equity premium. One could also define the ex post equity premium as $r(\bar{k}', \bar{l}', z) + 1 - q^{-1}$. By the law of iterated expectations, the unconditional expectations of the conditional expected equity premium and of the ex post equity premium are equal to each other.

22. Note that the equilibrium for the economy with $\underline{b} = 0$ can be calculated by looking at an economy in which capital is the only asset. In particular, we define the price of the bond to be the highest subjective price of the payoff the bond would give in the population (as measured by the expected intertemporal marginal rate of substitution evaluated at the bond payoff). This price would lead all agents to hold negative amounts of bonds if they could, except that individual whose subjective bond price is selected as the market price; this individual is therefore indifferent at zero bondholdings. It turns out that this individual is an unemployed agent who is just willing to hold positive amounts of capital.

23. Rouwenhorst (1995), who studies a complete-markets model with production, finds equity premia of roughly the same size as the premium that we find for the model with the loose borrowing constraint.

24. We thank Harald Uhlig for helpful discussions about the results in this section.

25. In the model without borrowing, consider instead the agent who determines the bond price as discussed in note 21.

26. In more explicit terms, this derivation reads as follows:

$$\sum_{i=b}^{g} \sum_{j=0}^{1} m'(R_e' - R_f) P(z' = z_i, \epsilon' = j \mid z, \epsilon, I)$$
$$= \sum_{i=b}^{g} \sum_{j=0}^{1} m'(R_e' - R_f) P(\epsilon' = j \mid z, \epsilon, z', I) P(z' = z_i \mid z, \epsilon, I)$$
$$= \sum_{i=b}^{g} \left[\sum_{j=0}^{1} m' P(\epsilon' = j \mid z, \epsilon, z', I) \right] (R_e' - R_f) P(z' = z_i \mid z, I).$$

27. Constantinides and Duffie (1996) rewrite the agent's Euler equation in a parallel way. In their setup, the expectations are not individual-specific, because all agents' intertemporal marginal rates of substitution follow the same process in the equilibrium they construct.

28. Let x be a random variable that equals x_1 with probability p and x_2 with probability $1 - p$. Let f and g be arbitrary functions such that $f(x_1) \neq f(x_2)$ and $g(x_1) \neq g(x_2)$. Then corr$[f(x), g(x)]$ is either 1 or -1 depending on the sign of $[g(x_2) - g(x_1)]/[f(x_2) - f(x_1)]$.

29. Lettau and Uhlig (1997) make a similar point.

30. Exactly what needs to be changed in terms of economic primitives for this experiment is not clear, of course, because consumption is endogenous.

31. Note that if the average marginal utility of consumption increases by more in the good state than in the bad state, then the market price of risk falls.

32. A similar mechanism accounts for the ability of the model of Constantinides and Duffie (1996) to explain equity premia. In particular, see equation (19) of their paper and the discussion that follows.

33. Thus, the failure of the no-borrowing model to generate equity premia anywhere near those in the data therefore can be attributed to the lack of variability in aggregate capital (and the consequent lack of variability in the return on equity).

34. The wealth distribution data are taken from Wolff (1994) and Díaz-Giménez et al. (1996).

35. The fit does worsen somewhat for the laws of motion for aggregate capital, but the fit for the bond-pricing function is still remarkable.

36. It is straightforward to derive a version of equation (11) that takes into account the idiosyncratic variation in discount rates. Note that the conditional correlation that appears in this version of equation (11) is again equal to -1.

37. Aiyagari and Gertler (1991) and Heaton and Lucas (1995, 1996) show that models with transactions costs can be successful in generating low bond returns/large equity premia provided that these transaction costs are of a certain form and minimum magnitude.

REFERENCES

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

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

Atkeson, A. & C. Phelan (1994) Reconsidering the costs of business cycles with incomplete markets. In S. Fischer & J.J. Rotemberg (eds.), *NBER Macroeconomics Annual 1994*, pp. 187–207. Cambridge, MA: MIT Press.

Campbell, J.Y. (1997) Asset Prices, Consumption, and the Business Cycle. Manuscript, Harvard University.

Castañeda, A., J. Díaz-Giménez & V. Ríos-Rull (1994) Unemployment Dynamics, Aggregate Fluctuations, and Income Distribution. Manuscript, Federal Reserve Bank of Minneapolis.

Christiano, L.J. & J.D.M. Fisher (1995) Asset Prices and Returns: Implications for Business Cycle Analysis. Manuscript, Northwestern University.

Cochrane, J. (1989) The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives. *American Economic Review* 79, 319–337.

Cochrane, J.H. (1991) Production-based asset pricing and the link between stock returns and economic fluctuations. *Journal of Finance* 46, 207–234.

Cochrane, J.H. & L.P. Hansen (1992) Asset pricing lessons for macroeconomics. In O.J. Blanchard & S. Fischer (eds.), *NBER Macroeconomics Annual 1992*, pp. 115–165. Cambridge, MA: MIT Press.

Coleman, W.J., II & M. Liu (1994) Incomplete Markets and Inequality Constraints: Some Theoretical and Computational Issues. Manuscript, Fuqua School of Business, Duke University.

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

de Boor (1978) *A Practical Guide to Splines*. New York: Springer–Verlag.

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

Den Haan, W.J. (1994) Credit Conditions, Cross-Sectional Dispersion, and ARCH Effects in a Dynamic Equilibrium Model. Manuscript, University of California at San Diego.

Den Haan, W.J. (1995) Aggregate Shocks and Cross-Sectional Dispersion. Manuscript, University of California at San Diego.

Den Haan, W.J. (1996a) Heterogeneity, aggregate uncertainty, and the short-term interest rate. *Journal of Business and Economic Statistics* 14, 399–411.

Den Haan, W.J. (1996b) Understanding Equilibrium Models with a Small and a Large Number of Agents. Discussion paper 96-34, University of California at San Diego.

Díaz-Giménez, J., V. Quadrini & J.-V. Ríos-Rull (1996) Measuring inequality: Facts on the distribution of earnings, income, and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review* (forthcoming).

Hansen, L.P. & R. Jagannathan (1991) Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99, 225–262.

Heaton, J. & D.J. Lucas (1995) The importance of investor heterogeneity and financial market imperfections for the behavior of asset prices. *Carnegie-Rochester Conference Series on Public Policy* 42, 1–33.

Heaton, J. & D.J. Lucas (1996) Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy* 104, 433–487.

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

Huggett, M. (1995) The One-Sector Growth Model with Idiosyncratic Shocks. Manuscript, Centro de Investigcation Economica—ITAM.

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

Johnson, S.A., J.R. Stedinger, C.A. Shoemaker, Y. Li & J.A. Tejada-Guibert (1993) Numerical solution of continuous-state dynamic programs using linear and spline interpolation. *Operations Research* 41, 484–500.

Krusell, P. & A.A. Smith, Jr. (1996a) Rules of thumb in macroeconomic equilibrium: A quantitative analysis. *Journal of Economic Dynamics and Control* 20, 527–558.

Krusell, P. & A.A. Smith, Jr. (1996b) Income and Wealth Heterogeneity in the Macroeconomy, Working paper 399, Rochester Center for Economic Research.

Lettau, M. & H. Uhlig (1997) Preferences, Consumption Smoothing, and Risk Premia. Manuscript, CentER for Economic Research, Tilburg University.

Lucas, R.E., Jr. (1987) *Models of Business Cycles*, New York: Basil Blackwell.

Lucas, D.J. (1995) Asset pricing with undiversifiable income risk and short sales constraints: deepening the equity premium puzzle. *Journal of Monetary Economics* 34, 325–342.

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

Marcet, A. & K.J. Singleton (1990) Equilibrium Asset Prices and Savings of Heterogeneous Agents in the Presence of Portfolio Constraints. Manuscript, Universitat Pompeu Fabra.

Mehra, R. & E.C. Prescott (1985) The equity premium: A puzzle. *Journal of Monetary Economics* 15, 145–162.

Press, W.H., B.P. Flannery, S.A. Teukolsky & W.T. Vetterling (1994) *Numerical Recipes*. Cambridge, England, UK: Cambridge University Press.

Quadrini V. & Ríos-Rull, J.-V. (1996) Models of the Distribution of Income, Wealth and Earnings. Research Department staff report, Federal Reserve Bank of Minneapolis.

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. In T.F. Cooley (ed.), *Frontiers of Business Cycle Research*, pp. 294–330. Princeton, NJ: Princeton University Press.

Stokey, N.L. (1996) Shirtsleeves to Shirtsleeves: The Economics of Social Mobility. Manuscript, Nancy L. Schwartz Lecture at Northwestern University.

Storesletten, K., C. Telmer & A. Yaron (1996) Persistent Idiosyncratic Shocks and Incomplete Markets. Manuscript, Graduate School of Industrial Administration, Carnegie Mellon University.

Tallarini, T.D., Jr. (1996) Risk-Sensitive Real Business Cycles. Manuscript, Graduate School of Industrial Administration, Carnegie Mellon University.

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

Weil, P. (1992) Asset pricing lessons for macroeconomics: comment. In O.J. Blanchard & S. Fischer (eds.), *NBER Macroeconomics Annual 1992*, pp. 169–181. Cambridge, MA: MIT Press.

Wolff, E.N. (1994) Trends in household wealth in the United States, 1962–83 and 1983–89. *Review of Income and Wealth* 40, 143–174.

APPENDIX

Here we give a description of the numerical techniques used to solve the consumer's dynamic programming problem (7). The algorithm is similar to one used by Johnson et al. (1993). We also describe how we solve the associated optimization problem (8) and how we find the market-clearing bond price in step 2b of the algorithm described in Section 3.2.2. The algorithm is described in the context of the model without preference shocks; it is straightforward to modify the algorithm to keep track of the additional state variable in the model with preference shocks.

The objective of the numerical algorithm for solving problem (7) is to approximate the four functions $v(\omega, 1; \bar{k}, z_g)$, $v(\omega, 1; \bar{k}, z_b)$, $v(\omega, 0; \bar{k}, z_g)$, and $v(\omega, 0; \bar{k}, z_b)$. We accomplish this task by approximating the values of the functions on a coarse grid of points in the $(\omega, \bar{k})$ plane and then using cubic spline and polynomial interpolation to calculate the values of these functions at points not on the grid. The numerical algorithm is in many ways analogous to value function iteration.

The following steps describe the numerical procedure. First, choose a grid of points in the $(\omega, \bar{k})$ plane (we give some details below about how we choose these points).

Second, choose initial values for each of the four functions at each of the grid points. We generally use the value function for the economy with only one asset [see Krusell and Smith (1996b)] as the initial condition for each of the functions.

Third, for each of the four (z, ϵ) pairs, maximize the right-hand side of Bellman's equation at each point in the grid. In this maximization, we allow the agent to select any values for capital and bondholdings. We use various interpolation schemes to compute the value function at points not on the grid (we describe the interpolation schemes in greater detail below). To find the optimal choices for capital and bondholdings, we use an algorithm that allows for the possibility that consumers will be at a corner. In particular, at each grid point in the agent's state space, we first determine whether the agent has an interior portfolio solution. We make this determination by (1) setting the choice for bondholdings equal to the lower bound on bondholdings, solving for the implied optimal capital holding using a one-dimensional search procedure, and calculating the implied utility level; (2) proceeding to a choice for bondholdings that slightly exceeds the borrowing constraint and repeating the procedure; if the utility so calculated is lower, stop, and determine that a corner solution for bonds has been obtained; if not, then go to the corner for capital and restart a parallel procedure. If, by following steps 1 and 2, we find that the agent has an interior portfolio solution, we use a bisection procedure to search over holdings of one of the assets, solving for holdings of the other asset as in step 1. When performing the search in step 1, we use a Newton–Raphson procedure unless the agent is at a corner with respect to both assets, in which case we use a bisection search procedure. This algorithm for finding the optimal portfolio choice relies on single-peakedness of payoffs in portfolio weights but is fast and gives very accurate approximations to the wealth ranges with interior portfolio decisions. Once we have computed the optimal values for capital and bondholdings, we insert these values into the right-hand side of Bellman's equation to obtain a new value for the value function at the specified grid point.

Fourth, compare the new optimal values generated by the third step to the original values. If the new values are close to the old values, then stop; otherwise, repeat the third step until the new and old values are sufficiently close.

We now comment on the choice of a grid in the $(\omega, \bar{k})$ plane and on the interpolation schemes that we use. Because there is generally not much curvature in the value function in the $\bar{k}$ direction, we use a small number of grid points in this direction and we use polynomial interpolation to compute the value function for values of $\bar{k}$ not on the grid. If there are n_k points in the $\bar{k}$ direction, then polynomial interpolation fits a polynomial of order $n_k - 1$ to the function values at these points (so that the polynomial fits the values exactly). This polynomial is then used to compute the value function in between grid points. We compute the value of the interpolating polynomial using Neville's algorithm, as described by Press et al. (1994, Ch. 3). This algorithm avoids the numerical instabilities associated with computing the coefficients of the interpolating polynomial. We generally use four equally spaced points in the $\bar{k}$ direction.

In the ω direction, there is generally a fair amount of curvature in the value function, especially for values of ω near zero. In this direction, therefore, we use cubic spline interpolation, which fits a piecewise cubic function through the given function values, with one piece for each interval defined by the grid. This piecewise cubic function satisfies the following restrictions: (1) It matches the function values exactly at the grid points, and (2) its first and second derivatives are continuous at the grid points. Cubic splines can be computed efficiently by solving a set of tridiagonal linear equations [see, e.g., the description in Press et al. (1994, Ch. 3) or de Boor (1978, Ch. 4)]. We generally use roughly 70 grid points in the ω direction, with many grid points near zero (where there is a lot of curvature) and fewer grid points for larger values of ω (where there is less curvature). We find that our results are not sensitive to increasing the number of grid points in either the ω or $\bar{k}$ directions.

To combine these two interpolation schemes, we therefore proceed as follows, where n_ω is the number of grid points in the ω direction and n_k is the number of grid points in the $\bar{k}$ direction. First, for each of the n_ω values of ω, use polynomial interpolation to compute the value function at the desired value of $\bar{k}$. This set of interpolations yields n_ω values of the value function, one for each value of ω on the grid. Second, use cubic spline interpolation using the n_ω values to calculate the value function for values of ω that are not on the grid. Note that because the values of $\bar{k}$ at which interpolated values must be computed are known at the beginning of each of the iterations on the value function (specifically, there are $2n_k$ such values—n_k points in the $\bar{k}$ direction times two possible outcomes for $\bar{k}'$, given $\bar{k}$), the required cubic splines need be computed only once for each iteration of the algorithm; once computed, it is easy to use these splines to calculate interpolated values.

We now describe how we compute the decision rules defined by problem (8) for a given value of q. Because the values of $\bar{k}$ and z are given by their current realizations in the simulations, we need only consider how agents' decisions vary with the individual state variables ω and ϵ. For each value of ϵ, we compute optimal savings and portfolio decisions on a grid of points for ω using the algorithm described above. To determine optimal decisions at points not on the grid, we proceed in two steps. First, we use linear interpolation between grid points to compute the optimal level of total savings for wealth levels that are not on the grid. Second, to determine how savings is split up between the two assets, we use a bisection procedure to locate the endpoints of the wealth ranges over which the agent has an interior portfolio decision. For values of wealth in this range, we approximate the optimal choice for capital using cubic spline interpolation between grid points. The optimal choice for bonds is then determined by the optimal level of total savings and the definition of savings. For values of wealth outside this range, the optimal choice for holdings of one of the assets is at a corner; the optimal choice for the other asset can then be determined from the optimal level of savings and the definition of savings.

Finally, to compute the market-clearing bond price $q^\star$, we need to solve the equation $D(q) = 0$, where $D(q)$ is the aggregate demand for bonds given a bond price q. We use a combination of methods to find $q^\star$. First, we bracket $q^\star$ with the two values

$$q_{low} = (1 - 10^{-6})\hat{q} \quad \text{and} \quad q_{high} = (1 + 10^{-6})\hat{q},$$

where $\hat{q}$ is the bond price predicted by the approximate bond-pricing function, given $\bar{k}$ and z. We then use a bisection procedure until, at the current bond price, at least some agents have an interior portfolio decision. Specifically, the bisection procedure works as follows:

1. Compute the midpoint $q = (q_{low} + q_{high})/2$.
2. Compute optimal decision rules, given q.
3. If all agents are short in bonds, then the bond price is too high: Set $q_{high} = q$.
4. If all agents are long in bonds, then the bond price is too low: Set $q_{low} = q$.
5. Return to step 1 and continue iterating until at least some agents have an interior portfolio decision.

We then switch to a version of the secant method described by Press et al. (1989, Ch. 9), or, if this method fails, to Brent's method as described by Press et al. (1989, Ch. 9). For each new candidate bond price q, we solve for the decision rules of agents and then estimate $D(q)$

by computing total bondholdings given the current distribution of wealth in our simulated sample. We continue iterating using either the secant method or Brent's method until $D(q)$ is close enough to zero (i.e., less than 10^{-3}). Although we could continue iterating using the bisection method, we find that switching to a more efficient root-finding algorithm leads to large computational savings.

Throughout the numerical work, our central concern is accuracy rather than speed or computational feasibility. For this reason, we use tight convergence criteria (percentage changes between successive iterations of less than 10^{-6} in most cases) and we choose reasonably large values for the number of grid points (when solving the consumer's problem), the number of quadrature points (when computing the conditional expectation in the consumer's problem for the case of an AR(1) productivity shock), the number of agents in the simulated cross-sectional distributions, and the length of the simulations. That is, the numerical results do not change appreciably when we choose smaller values for these parameters of the numerical algorithm. We find that, given a good initial condition, iterations on the value function lead to monotonic convergence (as predicted by the contraction mapping theorem). In all cases, converged decision rules conform to the predictions of economic theory. As discussed in the main text, the R^2's for the aggregate laws of motion and the pricing functions are very high in all cases, indicating that the inclusion of additional moments would change the results only slightly. Other goodness-of-fit measures [see Krusell and Smith (1996b) for further details] also show that the aggregate laws of motion and the pricing functions fit very well.