

Demystifying the Equity Premium

MASSIMILIANO DE SANTIS[†]

ABSTRACT. We provide an explanation for the high equity premium, the low risk free rate, and related puzzles based on a unifying theme: identifying the risks that firms and households really face. Our main explanation for the high equity premium is that there is a small persistent component to changes in dividend growth, which makes equity prices very volatile, and hence a poor insurance instrument. Our main explanation for the low risk free rate is that individuals face not only aggregate risk, but also significant idiosyncratic risk, which increases their desire for precautionary saving. These are our only two departures from the model of Mehra and Prescott. As in Mehra and Prescott, preferences are standard; so individuals' desire for consumption smoothing remains the main explanatory factor. The model also explains the extreme volatility of stock prices: the price-dividend ratio predicted by the model based on U.S. consumption data from 1891-2001 has a correlation of 72% with the actual price-dividend ratio in the S&P 500. The model succeeds in explaining the puzzles without assuming unrealistically high risk aversion: relative risk aversion can be ten or less.

Date: November 2, 2007.

*I am grateful to Timothy Cogley, Oscar Jorda, Jason Lepore, Martine Quinzii, Kevin Salyer, Aaron Smith. I especially thank Louis Makowski, who helped me a lot from the very inception of this project, giving generously of his time.

[†]Department of economics, Dartmouth College, 6106 Rockefeller Hall, Hanover, NH 03755. E-mail: massimiliano.desantis@dartmouth.edu.

1. INTRODUCTION

A number of empirical observations about financial markets challenge our understanding of the relationship between risk and returns.

1. **The high equity premium:** the fact that stock returns have exceeded the return on short term Treasury bills by 6% on average during the last 130 years.
2. **The low risk free interest rate:** the fact that the return on short term Treasury bills has averaged less than 2% during the same period.
3. **The “excess volatility” of stocks:** the high standard deviation of stock prices and stock returns, which average about 20% annually.
4. **The cyclicity and high persistence of the price-dividend ratio:** the price-dividend ratio is procyclical and has an autocorrelation of about 90% in annual data.

These four observations constitute puzzles because a simple formulation of the consumption-based capital asset pricing model (CCAPM), as in Mehra and Prescott (1985), calibrated to U.S. data generates an equity premium that is too low, a risk free rate that is too high, a standard deviation of stock returns that is too low, and a persistence in the price-dividend ratio that is too low relative to the above actual magnitudes.

In this paper we provide an explanation for these four puzzles based on a unifying theme: identifying the particular risks that firms and households really face. We begin with a brief but systematic summary of our explanation, along with historical motivation from the work of Mehra and Prescott.

1.1. Our explanation for the high equity premium. While our main focus will be on explaining (1), let us first look at (3) since, as we shall see, the high equity premium is closely related to the extreme volatility of stock prices.

If investors in period t receive real dividends D_t , expect their future dividends to grow at a constant rate g_t , and have a constant required real rate of return on equities r^E , then the price of equities is given by the standard Gordon pricing formula

$$P_t = \frac{D_t(1 + g_t)}{r^E - g_t}$$

(Myron J. Gordon, 1962). If shocks to dividend growth are i.i.d., information at time t does not cause investors to revise g_t , so price fluctuations are *only* due to changes in the level of D_t . This type of shock to dividends is not enough to generate the amount of price variation observed in

U.S. data. Since most of the stock price variation in Mehra and Prescott's model comes from this type of shock, the observed volatility of stock prices is a puzzle for their model.

But, if shocks to dividend growth are not i.i.d., investors will also adjust their forecast about future dividend *growth* when the level of D_t changes. An important feature of recessions is that the amount of uncertainty firms face increases dramatically: Will I be able to survive the current fall in demand? If so, will the demand for my product return to its pre-recession level or will my market position be permanently hurt? Such uncertainty is likely to lower firms' (and consequently investors') expectations about future dividend growth g_t at least for a while, until the uncertainty caused by a recession has sufficiently cleared.

In terms of the Gordon pricing formula, because of the nonlinear dependence of a stock's price on g_t — due to the compounding effects of growth, — even a small change in g_t can cause a large swing in stock prices. Indeed, to match the variability of actual U.S. stock prices, the needed fluctuations in g_t are so small that the dividend growth process is almost indistinguishable from an i.i.d. process, although the pricing implications are dramatically different. This is our explanation for the excess volatility puzzle: small but persistent changes in expected dividend growth result in large variations in stock prices and stock returns.¹

We are now ready to turn to (1), the equity premium puzzle. The equity premium π^E is defined as the excess return on equities relative to the return r^f on risk free short term Treasury bills:

$$\pi^E = r^E - r^f.$$

In CCAPM models with standard expected utility, the equity premium π^E is proportional to the covariance

$$- \text{Cov} \left(\frac{u'(C_{t+1}^i)}{u'(C_t^i)}, R_{t+1} \right),$$

where $u'(\cdot)$ is an individual i 's marginal utility from consumption, and R_{t+1} is the return on stocks next period. So the fraction in the above covariance is the individual's marginal rate of substitution between consumption this period and consumption in whatever state occurs next period. If this covariance is negative, the equity premium will be positive.

¹This explanation is based on Barsky and De Long (1993). We will show how to lift their partial equilibrium analysis to a general equilibrium CCAPM model, and hence how to apply it to explaining (1). By contrast, Barsky and De Long were exclusively interested in (3) and (4). We will see that lifting their analysis, and hence giving it a general equilibrium foundation, is by no means trivial; the first way one might try will not succeed in explaining either (3) or (4)!

The intuition is that, when the covariance is negative, the individual expects stocks to return relatively little precisely in those states next period when his consumption will be relatively low (i.e., during recessions), so when his marginal utility from extra consumption $u'(C_{t+1}^i)$ (scaled by $u'(C_t^i)$) will be relatively high. This implies that, relative to adding a bond, adding an equity share to the individual's portfolio of assets will tend to exacerbate the variability of his consumption (i.e., provide him with *negative* consumption insurance). Since the individual is risk averse, he will regard equities as inferior to bonds for achieving his goal of consumption smoothing, insisting on an equity premium.

Putting it together, recessions are periods in which, on average, individuals experience a drop in consumption *and* firms' returns also drop. Thus the covariance has the right sign, which makes the standard model a good candidate for explaining the equity premium. But we have seen that, if shocks to dividend growth are i.i.d., then stock prices and returns will not fluctuate a lot, and so the covariance, although of the right sign, will be low; consequently, Mehra and Prescott could not explain the high equity premium. By contrast, we have also seen that if shocks to dividend growth g_t include a small persistent component, then stock prices and returns will fluctuate dramatically through the course of a business cycle. Thus the above covariance, and hence the equity premium, will be high. This is our main explanation for (1). *With a persistent component in dividend growth consistent with U.S. data, our model generates an equity premium of about 4.2% even in a standard, consumption-based, representative agent model like that in Mehra and Prescott.*

Let us now turn to the risk free rate puzzle, (2). Since real income is growing in the U.S. economy, the consumption-smoothing motive makes individuals want to borrow today in anticipation of their higher future income. Mehra and Prescott observed that they could explain the high equity premium if individuals were very risk averse. But the more risk averse the individual, the stronger his desire for consumption smoothing, hence the larger will be his desire for short term borrowing, and consequently the larger will be the risk free rate r^f . The upshot was that Mehra and Prescott found themselves caught on the horns of a dilemma: their model could explain (1) if and only if it failed to explain (2). We have seen that we can explain (1) and (3) by understanding more fully the risks faced by firms and individuals through the course of a business cycle. Our explanation for the low risk free rate is a continuation of this overall theme.

Unlike the representative agent model of Mehra and Prescott, realistically individuals also must bear a lot of personal, idiosyncratic risk. Will

I lose my job or not? Will I be laid off or not? This sort of risk is largely uninsurable, and it increases dramatically during recessions. Thus, any one individual faces a lot more risk than is reflected in the fluctuations of aggregate per capita consumption. This greater level of consumption risk makes any one individual more cautious about consuming today, that is, it increases his desire for *precautionary savings* just in case his fortunes flounder tomorrow. Thus, individuals who must bear both aggregate *and* idiosyncratic risk will be willing to pay a higher price for transferring one unit of consumption from today to tomorrow, which makes the risk free rate r^f much smaller than one would predict using a representative agent model. The presence of significant idiosyncratic risk, leading to a strong desire for precautionary savings, is how we explain (2), the risk free rate puzzle. *When we calibrate this risk to existing panel data evidence, our model generates a risk free rate of only 1.5%.*

The presence of idiosyncratic risk not only affects the level of the risk free rate r^f , but also the size of the equity premium. Since individuals' idiosyncratic risk is counter-cyclical, the presence of this additional risk makes the covariance above and hence the equity premium even larger than it would be in a representative agent model. Once idiosyncratic risk is taken into account, equities (when compared to bonds) look even worse as a financial instrument for consumption smoothing since, during recessions when equities will have a low return, an individual's consumption will not only be hit by an adverse aggregate shock, but also is more likely to be hit by an adverse idiosyncratic shock. *With a realistic level of idiosyncratic risk, the equity premium increases to the 6% level needed to match the empirical evidence.*

Observation (4), the fact that the price-dividend ratio in U.S. data is procyclical and has a high autocorrelation, provides indirect evidence for our hypotheses that dividend growth g_t is procyclical and has a small persistent component. Notice from the Gordon pricing formula that

$$\frac{P_t}{D_t} = \frac{1 + g_t}{r^E - g_t}.$$

If shocks to dividend growth are i.i.d., then expected dividend growth g_t would be time invariant. Consequently, the price-dividend ratio would be constant, without cyclicity or persistence. By contrast, observe the implication of our story, based on firms' increased uncertainty during recessions, for the behavior of the price-dividend ratio. The same small persistence in changes to g_t that we used to explain the volatility in stock prices also explains the procyclical behavior of the price-dividend ratio, and the persistence observed in this ratio: Because expected dividend

growth falls slightly during recessions, stock prices fall *more* than dividends during recessions, which in turn implies the price-dividend ratio falls during recessions. Expressed a little differently, fact (4) provides indirect evidence for our hypotheses about g_t through the business cycle.

Summarizing, (3) explains (1), and (4) provides evidence for our explanation of (3). It is intellectually satisfying to find that the resolution of puzzles (1), (3), and (4) is intimately connected. All this is entirely consistent with a representative agent model. To also explain (2), we depart from such a model by taking into account that individuals face not only aggregate risk, but also significant idiosyncratic risks. Throughout we follow Mehra and Prescott in assuming individuals have entirely standard preferences.

1.2. Non-standard preferences. By contrast, in response to the failures of Mehra and Prescott and many others, the asset pricing literature seems to have concluded that some departure from standard preferences was needed, habit persistence and recursive preferences being the leading candidates (see Parker and Julliard, 2005, p. 186). Commenting on the use of non-standard preferences to explain the equity premium and to calculate the welfare cost of consumption fluctuations, Robert Lucas in his AEA presidential address said:

Resolving empirical difficulties by adding new parameters
always works, but often only by raising more problems.
(Lucas, 2003, p. 8)

In accord with Lucas' insight, *implausible* parameter values are essential for explaining the above stylized facts using either habit persistence or recursive preferences.²

For example, Epstein-Zin-Weil preferences can be viewed as a generalization of standard expected utility with CRRA preferences, a generalization that allows for the intertemporal elasticity of substitution (*IES*) to differ from relative risk aversion (*RA*).³ Bansal and Yaron (2004) use preferences of this type to construct a representative agent model that replicates the above empirical regularities exceptionally well. To do so, they assume a fairly high but not implausible value for relative risk aversion ($RA = 10$), but they also assume an implausibly high value for the intertemporal elasticity of substitution ($IES = 1.5$). These parameter choices imply consumers are far from standard: the representative agent is very tolerant to fluctuations in his consumption over time since

²The use of non-standard preferences is itself not objectionable, only their use in conjunction with implausible parameter values.

³Notice one also can obtain a separation using standard expected utility theory, e.g., by introducing a labor leisure tradeoff, introducing consumption of durable goods, etc.

$IES = 1.5$, but at the same time he is very averse to fluctuations across states since $RA = 10$. This “solves” the risk free rate puzzle by making the standard consumption-smoothing motive largely irrelevant for asset pricing. What would happen if, while maintaining a separation between the two parameters (in particular, keeping $RA = 10$), we made the model closer to the standard one by assuming IES is *less* than one, as most empirical studies find? With $IES < 1$, the Bansal and Yaron model implies that stock prices relative to dividends should be *higher* in recessions than in expansions, and that stock prices should *increase* with an increase in uncertainty. These are very unrealistic features for any asset pricing model.⁴

The habit persistence model of Campbell and Cochrane (1999) also successfully matches the above four empirical regularities. In addition, it matches the predictability of stock returns. Thus Campbell and Cochrane’s pioneering work set an extremely high standard for any future asset pricing model. Their success in reverse engineering taught us the nature of the dynamic relationship between dividends, consumption, and the stochastic discount factor needed to replicate the empirical evidence. But, as another illustration of Lucas’ concern, Campbell and Cochrane assume an implausibly high steady state value for relative risk aversion ($RA = 80$). Continuing with Lucas’ quote:

It would be good to have the equity premium resolved,
but I think we need to look beyond high estimates of risk
aversion to do it.

Summarizing, models with non-standard preferences can match the empirical data only by assuming implausibly high values for RA or IES . The current model can be viewed as an important next step, showing that one can successfully match the empirical data with $RA = 10$ (or even less) *and* with $IES < 1$. This is important not only because $IES < 1$ is more realistic. But, as pointed out above, $IES < 1$ has crucial (and unpleasant) implications in Bansal and Yaron’s model. Since we follow Mehra and Prescott in assuming individuals have entirely standard preferences, the desire for consumption smoothing plays the central role in this paper, which makes our explanation for the equity premium intuitively appealing. Our departures from Mehra and Prescott just involve providing a more realistic description of the risks that firms and individuals face: our main explanation for (1) is that there is a small persistent component to changes in dividend growth (it is not i.i.d.), while our main explanation

⁴See Bansal and Yaron (2004) equations (5) and (9). We discuss their model further in later sections.

for (2) is that individuals face not only aggregate risk, but also significant idiosyncratic risks.⁵

Section 2 presents the paper's basic asset pricing model, which can be solved analytically. This model matches the historical equity premium, low risk free rate, and procyclical behavior of the dividend-price ratio. In recognition of the high standard set by Campbell and Cochrane, Section 3 presents a generalization of the basic model that, in addition to these three features, generates a countercyclical Sharpe ratio and predictable excess returns. Section 4 presents simulation results based on the generalized model, Section 5 provides a more in depth review of the literature, and Section 6 concludes.

2. BASIC MODEL

Consider an exchange economy with a single non-durable consumption good and two traded assets, a risk-free discount bond and a risky equity. Bonds are issued at time $t-1$, matured at t , and each bond has a par value of one. We assume the bond is in zero net supply. The risky equity (whose net supply we normalize to be one) pays dividend D_t and has ex-dividend price P_t . Each consumer i is endowed with labor income I_t^i and consumes C_t^i at time t . Aggregate labor income is I_t , and aggregate consumption is $C_t = I_t + D_t$. It is assumed that $I_t + D_t > 0$ for all times t . There is an infinite set of distinct consumers denoted by A . The increasing sequence of information sets $\{\mathcal{F}_t : t = 0, 1, 2, \dots\}$ available to each consumer includes the equity's dividend history, the history of equity and bond prices, and the disaggregated labor income history $\{I_s^i : i \in A, 0 \leq s \leq t\}$. At time t , consumer i holds a portfolio of shares of the risky asset θ_t^i and of the bond b_t^i . The time t budget constraint is:

$$(2.1) \quad C_t^i + \theta_t^i P_t + b_t^i \leq I_t^i + \theta_{t-1}^i (P_t + D_t) + b_{t-1}^i R_t^f,$$

where R_t^f denotes the return on a bond issued at $t-1$. Consumers have homogeneous preferences represented by a time-separable von Neumann-Morgenstern utility function with constant relative risk aversion coefficient γ and a constant subjective discount factor δ . At time 0, each consumer maximizes

$$(2.2) \quad E \left[\frac{\sum_{t=0}^{\infty} \delta^t (C_t^i)^{1-\gamma}}{1-\gamma} \mid \mathcal{F}_0 \right]$$

⁵It may be worthwhile to broaden our perspective for a moment. Models with standard preferences and plausible parameter values have been successfully used to interpret observations in growth theory, business cycle theory, labor market behavior, and so on. One wonders whether non-standard models with RA = 80 or IES = 1.5 would be so robust.

subject to the sequence of budget constraints (2.1) by choosing a sequence

$$(\theta^i, b^i, C^i) = \{\theta_t^i, b_t^i, C_t^i\}_{t=0,1,2,\dots}$$

An *equilibrium* for this economy is a security price and bond return process, (P, R^f) , and strategies $\{(\theta^i, b^i, C^i) : i \in A\}$ for the consumers such that

- (i) (θ^i, b^i, C^i) maximizes (2.2) subject to (2.1)
- (ii) markets clear, i.e., $\sum_{i \in A} \theta_t^i = 1$ and $\sum_{i \in A} b_t^i = 0$ for all t .

Market clearing implies that $\sum_{i \in A} C_t^i = C_t \equiv I_t + D_t$ for all t . An equilibrium price process for the risky asset will satisfy the following condition for all individuals $i \in A$:

$$(2.3) \quad P_t = E \left[\delta \left(\frac{C_{t+1}^i}{C_t^i} \right)^{-\gamma} (P_{t+1} + D_{t+1}) | \mathcal{F}_t \right],$$

that is, the price of an equity share equals its expected discounted payoff, where the expectation is taken with respect to the information set \mathcal{F}_t .

As in Constantinides and Duffie (1996), individuals in our economy will be subject to both aggregate and idiosyncratic income shocks. They show the existence of an equilibrium when idiosyncratic income shocks are a martingale. (Appendix C briefly describes such a process.) They also show that the equilibrium price process satisfies:

$$(2.4) \quad P_t = E \left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \exp \left[\frac{\gamma(\gamma+1)}{2} y_{t+1}^2 \right] (P_{t+1} + D_{t+1}) | \phi_t \right],$$

where the variable y_{t+1}^2 represents the variance of the cross-sectional distribution of individual consumption growth relative to aggregate growth, i.e.

$$y_{t+1}^2 = \text{Var} \left(\log \left(\frac{C_{t+1}^i / C_{t+1}}{C_t^i / C_t} \right) \right).$$

The variable y_{t+1}^2 is thus the variance of the idiosyncratic shock to consumption growth. Notice from (2.4) that P_t depends only on aggregate quantities. The information set ϕ_t differs from \mathcal{F}_t in that it does not include the disaggregated labor income histories, i.e., it is a subfield of \mathcal{F}_t and is interpreted as the information set observed by the econometrician. But the expectation in (2.4) is the same if we condition on \mathcal{F}_t , because the extra information contained in \mathcal{F}_t is not relevant in calculating the expected value.⁶

We now go beyond the Constantinides and Duffie model. We will describe a simple process for aggregate consumption growth, dividend

⁶In Appendix C, we detail the specification of the income shocks, which contain a martingale component as in Constantinides and Duffie (1996). Meghir and Pistaferri

growth, and idiosyncratic risk that leads to an analytical solution of the equilibrium and can be used to determine the level of the risk free rate, the equity premium, and the procyclical variation in the price-dividend ratio.

Let $g_{t+1}^c = \Delta c_{t+1}$ and $g_{t+1} = \Delta d_{t+1}$, where lower case letters denote logs of the uppercase counterparts. So g_t^c is log per-capita consumption growth, g_t is log dividend growth, and y_t^2 is the variance of the idiosyncratic income shock. The process for g_t^c, g_t, y_t^2 is as follows:

$$(2.5) \quad \begin{aligned} g_{t+1}^c &= \mu^c + \sigma \eta_{t+1} \\ g_{t+1} &= \mu^d + \phi x_t + \sigma \varphi_d u_{t+1} \\ x_{t+1} &= \rho x_t + \sigma \varphi_e \eta_{t+1} \\ y_{t+1}^2 &= \bar{y}^2 - \varphi_y \eta_{t+1}, \end{aligned}$$

with the two shocks η_{t+1} and u_{t+1} having correlation $\rho_{\eta,u}$. These shocks are assumed to be i.i.d. with normal distribution $N(0, 1)$. Hence σ is the standard deviation (volatility) of consumption growth. The volatility of dividend growth is $\sigma \varphi_d$, where $\varphi_d > 1$, reflecting the fact that dividends are more volatile than consumption in the data. The mean of dividend growth includes a small (relative to μ^d) time varying and persistent component x_t . Hence φ_e will be very small, to ensure that x_t is small.

Agents do not know the variance of the idiosyncratic shocks nature will pick, which is reflected in the fact that y_{t+1}^2 is a random variable with a positive standard deviation φ_y . A greater value for φ_y implies greater uncertainty. Because y_{t+1}^2 is a random variable with known distribution, individual consumption growth $\frac{C_{t+1}^i}{C_t^i}$ is a continuous mixture distribution. Agents know this distribution, and because the parameters in g_{t+1}^c and y_{t+1}^2 are constant, this distribution is time invariant, that is, individual risk is constant in this model.⁷

We now relate the model's specifications to our explanation of the four puzzles outlined in the introduction. Recall the importance of aggregate shocks — recessions and expansions — for the risks faced by both firms and households.

First consider firms. We explained in the introduction that the substantial increase in uncertainty during recessions is likely to lower firms' (and consequently investors') expectations about future dividend growth g_t at least for a while, until uncertainty has sufficiently cleared. When

(2004) provide empirical evidence that supports the presence of a martingale component in household earnings, while Deaton and Paxson (1994) provide evidence of a martingale component in consumption.

⁷See Appendix C for details; equation C.1 in the Appendix can be used to calculate the moments of the mixture distribution.

the model is hit by a recession — a large negative value of the aggregate shock η_t , — expected *future* dividend growth g_{t+1} decreases by the change in x_t , which in turn depends on η_t . Notice that with $\rho > 0$, this shock to *expected* dividend growth will persist for a while. As we saw in the introduction using the Gordon growth formula, small fluctuations in g_{t+1} can generate large fluctuations in stock prices.

Next consider households. When the economy is hit by a negative aggregate shock η_t , the variance of the distribution of the idiosyncratic shock y_t^2 is also greater than average. Thus, a negative aggregate shock also leads to a greater probability of a large, uninsurable, idiosyncratic income and consumption shock at the individual level: Will I lose my job or not? Will I be laid off or not? Notice the perfect correlation between y_t^2 and η_t . It is conceivable that idiosyncratic risk y_t^2 should also include a shock not related to η_t . But because this unrelated shock does not matter for the equity premium (it does not affect the covariance with aggregate risk), we only consider the component perfectly correlated to η_t here, and will take this into account in our calibration (see the discussion of Table 2 in Section 3.3).

Our specification of the dividend growth process basically follows Barsky and De Long (1993); so let us call the hypothesis that expected dividend growth is slightly procyclical the *Barsky-De Long hypothesis*. They showed, using a partial equilibrium model, that a small and highly persistent process (they chose $\rho = 1$) plus a volatile i.i.d. process, as in our specification of g_{t+1} , cannot be distinguished from a pure i.i.d. process. Further, Shephard and Harvey (1990) showed that standard identification techniques would favor the i.i.d. process. Thus it is empirically difficult to distinguish these two types of processes in a finite sample. But their economic implications are very different, as we explained in the introduction.

In order to lift the Barsky and De Long analysis to the general equilibrium level, a subtlety must be taken into account. Because their analysis was partial equilibrium, they could view the discount rate r^E in the Gordon growth formula as fixed.⁸ But, in general equilibrium, r^E is endogenously determined by individuals' marginal rates of substitution. In terms of the Gordon formula, we must worry that perhaps r^E will fall when g_t falls, undercutting the Barsky-De Long explanation for puzzles (3) and (4). Indeed, to lift the Barsky-De Long explanation of (3) and (4) to a general equilibrium CCAPM model (hence to give their explanation a

⁸Barsky and De Long were only interested in explaining the excess volatility of stocks, not the high equity premium, so they could restrict themselves to partial equilibrium.

general equilibrium foundation), we must strengthen the Barsky-De Long hypothesis, and hypothesize that:

aggregate shocks to income have a more persistent effect on dividend growth than on aggregate consumption growth,

where notice the growth in aggregate consumption responds not just to the growth in dividends, but also to the growth in all other forms of (real) income, notably wage income. This explains why we specify aggregate consumption growth as being i.i.d., not containing the persistent component x_t .

To see the intuition why the stronger hypothesis is needed, suppose that in a recession (a negative aggregate shock) an individual experiences a drop in the both the levels of his dividend and wage incomes. If the individual expects the drop in his current wages to also signal a small persistent drop in his future wage *growth*, his optimal response would be to consume relatively less today and hence to save relatively more today (for consumption smoothing) than if he expects wage growth to be i.i.d.⁹ Since equities are a savings tool, albeit an imperfect one, such persistence in individuals' beliefs about their wage growth will, *ceteris paribus*, increase their demand for equities and therefore stock prices. This positive effect on stock prices will tend to dampen the negative effect on prices brought about by the fall in dividend growth g_t . In terms of the Gordon pricing formula, any persistent component in aggregate wage (and aggregate consumption) growth would make r^E *fall* during recessions, which would dampen and may even completely offset the fall in g_t . The fact that the price-dividend ratio is procyclical and persistent (puzzle (4)) can be viewed as providing indirect evidence for our stronger version of the Barsky-De Long hypothesis: Without this strengthening, (4) would again become puzzling — when viewed from a general equilibrium perspective.

This brings us to a crucial difference between our model and those of Mehra and Prescott and of Bansal and Yaron. Mehra and Prescott made the simplifying assumption that dividends equal consumption:

$$D_t = C_t.$$

Thus, even if they had included the Barsky-De Long hypothesis, their model could not explain the volatility of stocks or the high equity premium because consumption growth would have the *same* persistence as dividend growth. Unlike Mehra and Prescott, Bansal and Yaron followed

⁹If expected wage growth is i.i.d., the individual expects any drop in the level of his wages to be permanent, but he expects his future wages will continue to grow at the same rate as before, although growing now from a lower starting point.

Barsky and De Long in specifying that dividend growth has a persistent component x_t ; and they distinguished between the dividend and consumption growth processes. But they assumed these two processes have the *same* persistent component (no asymmetry). Hence, under standard preferences, a recession in their model would cause both g_t and r^E to fall. In particular, since they assumed a relatively high risk aversion ($RA = 10$), in their model r^E could fall more than g_t , making prices *increase* in a recession. They thus concluded that Epstein-Zin preferences *and an IES > 1* were needed for the Barsky-De Long analysis to carry over to a general equilibrium model.¹⁰

By enriching our model to include a storage technology, our stronger version of the Barsky-De Long hypothesis may be demonstrable as a conclusion — even if *all* forms of income have the same persistence as dividends. To illustrate suppose households have access to a linear storage technology, so they need not consume all their income. Under log-normality assumptions about the stochastic income process, the log of consumption would follow a random walk with drift. Even without log-normality assumptions, consumption would still be very close to a random walk with drift, as shown in Robert Hall’s 1978 paper (see his Corollary 5). It also can be shown that with CRRA preferences and a linear storage technology, up to a second order approximation of the intertemporal MRS’s, the log of consumption follows a random walk with drift, which is our assumption.¹¹

Even in the absence of any storage technology, there are other ways to justify our strengthening of the Barsky-De Long hypothesis. When all income must be consumed, our strengthening says that the growth in firms’ profits is *more* persistent than the growth in other forms of income. Why should this be? One clue comes from the literature on irreversible investment under uncertainty. When firms’ uncertainty goes up during recessions, many potential investment projects will begin to look less profitable and, consequently, many irreversible investment projects will be prudently delayed. This will lower a firm’s future growth prospects, at least for a while, until the uncertainty caused by a recession has sufficiently cleared (see Hart, 1942; Jones and Ostroy, 1984; Bernanke, 1983; Dixit and Pindyck, 1994). Secondly and relatedly, the “consumption” we are calibrating in the model is the consumption of nondurables and services.

¹⁰This assumption also helped Bansal and Yaron to explain (2), the risk free rate puzzle. Weil (1989) showed that, even with Epstein-Zin preferences, *IES* needs to be greater than unity to match the low risk free rate.

¹¹To calibrate such an enriched model one would have to assume that the rate of return from the linear storage technology is low, say 1.5%, to match the low risk free rate.

Since aggregate shocks are felt disproportionately in the durable goods sectors, we can think the x_t shock affects the consumption of durables in the same persistent way it affects the dividend process; only the consumption of nondurables and services is i.i.d.

2.1. Solution. We can now solve for the price-dividend ratio, the risk free rate, and the equity premium of this economy, and therefore verify that our explanation of facts (1)-(4) is consistent with plausible parameter values. We begin with (3) and (4). We can solve for the price-dividend ratio by using the log-linear approximation of returns in Campbell and Shiller (1988):

$$(2.6) \quad \ln(R_{t+1}) \equiv r_{t+1} = \kappa_0 + \kappa_1 v_{t+1} - v_t + g_{t+1},$$

where v_t is the log of the price-dividend ratio, i.e. $v_t = \ln(P_t/D_t)$, while κ_0 and κ_1 are constants of approximation. In Appendix A we show that, by using the approximation in the Euler equation (2.4), the log of the price-dividend ratio v_t is an affine function of the state x_t , i.e. in this economy:

$$v_t = A_0 + A_1 x_t, \quad \text{with}$$

$$A_1 = \frac{\phi}{1 - \kappa_1 \rho} > 0.$$

Notice the pure i.i.d. shocks to dividends u_t do not affect v_t , only the persistent component x_t does. If there is a recessionary shock ($\eta_t < 0$), time varying expected growth x_t falls and so does the price of the equity relative to dividends. The approximating constant κ_1 is close to one, and so is ρ , which together imply that even a small change to the conditional mean of dividend growth can have important implications for asset prices. This ‘‘Barsky and De Long effect’’ can explain both the procyclical behavior of the price-dividend ratio and the high volatility of aggregate stock prices. Further, the price-dividend ratio will inherit the persistence of x_t . Thus our strengthening of the Barsky-De Long hypothesis allows us to explain puzzles (3) and (4) using plausible parameter values. In our benchmark simulation, we will assume $RA = 10$ and hence $IES = 1/10$.¹²

Moving to the risk free rate and the equity premium puzzles, we can use the Euler equation (2.4) to find the risk free rate and the expected return on the risky asset by using our normality assumptions. First re-write (2.4)

¹²By contrast, if $IES < 1$ in the Bansal and Yaron’s model, the coefficient in their model corresponding to our A_1 becomes *negative*. Thus they concluded that Epstein-Zin preferences *and* $IES > 1$ are needed to replicate the empirical evidence in a general equilibrium model.

as

$$(2.7) \quad E_t \left[\exp \left\{ \ln \delta - \gamma g_{t+1}^c + \frac{\gamma(\gamma+1)}{2} y_{t+1}^2 + r_{t+1} \right\} \right] = 1.$$

Then use the fact that if X is normal, $Ee^X = e^{EX+0.5\text{Var}(X)}$, and take logs of both sides to obtain

$$(2.8) \quad \ln \delta - \gamma E_t g_{t+1}^c + \alpha E_t y_{t+1}^2 + E_t r_{t+1} + \frac{1}{2} \text{Var}_t(-\gamma g_{t+1}^c + \alpha y_{t+1}^2 + r_{t+1}) = 0,$$

where $\alpha = 0.5\gamma(\gamma+1)$.¹³ The Euler condition (2.8) is valid for all assets, so we can substitute r_{t+1}^f for r_{t+1} . Therefore the risk free rate is given by

$$(2.9) \quad r_{t+1}^f = -\ln \delta + \gamma \mu^c - \alpha \bar{y}^2 - \frac{1}{2}(\sigma\gamma + \alpha\varphi_y)^2.$$

By contrast, in a representative agent economy, $r_{t+1}^f = -\ln \delta + \gamma \mu^c - \frac{\gamma^2 \sigma^2}{2}$. Since σ is only 3.6% in Mehra and Prescott's data, σ^2 is very small. So, even with high risk aversion, the precautionary saving term $\frac{\gamma^2 \sigma^2}{2}$ is second order, implying an unrealistically large risk free rate when γ is large (the risk free rate puzzle). With uninsurable idiosyncratic shocks to income, the precautionary saving motive is not second order. For example, a relative risk aversion coefficient of 10 and a cross-sectional standard deviation \bar{y} of 5% leads to a risk free-rate of less than 2%, even if $\varphi_y = 0$. So the low risk free rate, fact (2), is not a puzzle in our model.

Turning finally to (1), we can use the Euler equation again (this time with r_{t+1} instead of r_{t+1}^f) to find the expected return on the risky asset, which will now contain the covariances implied by the variance term in (2.8):

$$\begin{aligned} \text{Var}_t(-\gamma g_{t+1}^c + \alpha y_{t+1}^2 + r_{t+1}) &= (\gamma\sigma + \alpha\varphi_y)^2 + \text{Var}_t(r_{t+1}) + \\ &\quad - 2\gamma \text{Cov}_t(g_{t+1}^c, r_{t+1}) + 2\alpha \text{Cov}_t(y_{t+1}^2, r_{t+1}). \end{aligned}$$

This implies the risk premium on equity is

$$(2.10) \quad E_t(r_{t+1} - r_{t+1}^f) = \gamma \text{Cov}_t(g_{t+1}^c, r_{t+1}) - \alpha \text{Cov}_t(y_{t+1}^2, r_{t+1}) - 0.5 \text{Var}_t(r_{t+1}).$$

The first term is the covariance found in the standard, representative agent model. The second covariance comes from the presence of idiosyncratic risk, and contributes to a positive risk premium if it is negative, i.e. returns are low in times of larger idiosyncratic shocks, as is the case in our model. The third term is a Jensen's inequality term arising

¹³Notice that the normality of X is justified here. For the risky asset, it follows from (2.6) and the solution for v_t that the log of returns r_{t+1} are approximately normally distributed. For the risk free rate, since it is known at time t , normality follows from our assumptions about g^c and \bar{y}^2 .

from the fact that we are describing expectations of log returns. In effect, this term converts the expected excess returns from a geometric average to an arithmetic average.

We can substitute the log-linear approximation for r_{t+1} in the equity premium expression (2.10) using our solution for v_t , and calculate the premium for this economy:

$$(2.11) \quad E_t(r_{t+1} - r_{t+1}^f) = \gamma\sigma^2\varphi_d\rho_{\eta,u} + \gamma\sigma^2\kappa_1 A_1\varphi_e \\ + \alpha\varphi_y\kappa_1 A_1\sigma\varphi_e + \alpha\varphi_y\varphi_d\rho_{\eta,u} - 0.5 \text{Var}_t(r_{t+1}).$$

The first two terms come from $\text{Cov}(g_{t+1}^c, r_{t+1})$, while the following two terms come from $\text{Cov}(y_{t+1}^2, r_{t+1})$. Since the small persistent component of shocks to dividend growth causes large swings in equity prices, the risk effect is also magnified. Notice in fact that the premium depends on A_1 , which is large for ρ sufficiently close to one. For the parameterization that we will use later, the premium in this economy is about 6%. The first term, $\gamma\sigma^2\varphi_d\rho_{\eta,u}$, is the ‘‘Mehra and Prescott’’ equity premium, and it is between 0.7 and 1.8%, depending on the value of $\rho_{\eta,u}$ used. The second term, $\gamma\sigma^2\kappa_1 A_1\varphi_e$, is the ‘‘Barsky-De Long’’ premium, and it is the most important, adding about 3.2% to the overall premium. This is the covariance between the persistent shock to dividend growth and aggregate consumption growth; persistent negative shocks occur in recessions. The term $\alpha\varphi_y\kappa_1 A_1\sigma\varphi_e$ adds about 1% to the premium as does the last term, $\alpha\varphi_y\varphi_d\rho_{\eta,u}$. These last two terms are the covariances between the shocks to dividend growth (persistent and transitory) and idiosyncratic risk.

Because $\text{Cov}(g_{t+1}^c, r_{t+1})$ is the most important source of the premium, we must check that this covariance is in accord with estimates from the data. As we will see in Table 3, our model-generated covariance is not unrealistically high, being very close to U.S. estimates.

3. GENERALIZED MODEL

3.1. The Stochastic Process. Notice that the equity premium is constant in the economy above, and so is the variance of returns on the risky asset. The predictability literature indicates countercyclical variation in the expected excess return (Campbell and Shiller 1988a, 1988b). Estimates of the conditional variance of returns also change over time, but they do not move one for one with the conditional mean of excess returns. This implies time varying Sharpe ratios (i.e., expected excess returns per unit of standard deviation). The Sharpe ratio on the risky asset satisfies the following inequality:

$$\frac{E_t(R_{t+1} - R_{t+1}^f)}{\sigma_t(R_{t+1})} \leq \frac{\sigma_t(M_{t+1})}{E_t(M_{t+1})},$$

where M_{t+1} is the stochastic discount factor. Since $E_t M_{t+1}$ is almost constant in the data, a model that explains countercyclical Sharpe ratios should have a stochastic discount factor that is heteroskedastic, with greater variance in bad times. We now specify such a model.

In our basic model, individual risk (aggregate plus idiosyncratic) is constant over time. While aggregate risk and idiosyncratic risk are correlated, the probability distribution of individual consumption growth is time invariant because the aggregate shocks η_t are i.i.d. More realistically one would expect that, as the economy goes into a recession, individual uncertainty (risk) rises, and with it the desire for income insurance. According to this view of the economy, equities should be less valuable relative to bonds in recessions than in expansions. There is more than one way to specify a process that generates the desired time variation in the Sharpe ratio. Given the low variation in $E_t(M_{t+1})$, we will specify a process with changing uncertainty such that the risk free rate in the economy remains constant; in this we follow Campbell and Cochrane (1999).

The dividend growth and consumption growth processes are specified as in (2.5), repeated here for convenience:

$$(3.1) \quad \begin{aligned} g_{t+1}^c &= \mu^c + \sigma \eta_{t+1} \\ g_{t+1} &= \mu^d + \phi x_t + \sigma \varphi_d u_{t+1} \\ x_{t+1} &= \rho x_t + \sigma \varphi_e \eta_{t+1}. \end{aligned}$$

But y_{t+1}^2 in the generalized model is given by the following MA(1) with heteroskedastic error term:

$$(3.2) \quad y_{t+1}^2 = \bar{y}^2 - \sigma \lambda_t \eta_{t+1} + \theta \sigma \lambda_{t-1} \eta_t.$$

The specification of y_{t+1}^2 in the basic model of Section 2 corresponds to the special case of Equation (3.2) in which

$$\theta = 0 \quad \text{and} \quad \sigma \lambda_t = \varphi_y, \text{ a constant.}$$

The term $\sigma \lambda_t$ is the conditional standard deviation of y_{t+1}^2 . As in the basic model, agents face uncertainty about the variance of idiosyncratic shocks, and individual consumption growth is a known mixture distribution. But now, since the standard deviation λ_t changes with t , the mixture distribution is not time invariant, that is, individual risk is time varying in this model. In particular, recessions will be periods of greater individual risk (higher λ_t).

We use the equations of the maximal Sharpe ratio and the risk free rate for this economy to specify λ_t and the sign of the constant θ . The time varying conditional variance will generate a heteroskedastic stochastic discount factor. The time varying conditional mean, the MA(1) component

$\sigma\lambda_{t-1}\eta_t$, affects variation in the risk free rate. Using the normality assumption, we can get the risk free rate from

$$E_t m_{t+1} + E_t r_{t+1}^f + 0.5 \text{Var}(m_{t+1}) = 0.$$

Recall that $m_{t+1} = \ln \delta - \gamma g_{t+1}^c + \alpha y_{t+1}^2$, hence:

$$\begin{aligned} E_t m_{t+1} &= \ln \delta - \gamma \mu^c + \alpha E_t[y_{t+1}^2], \\ \text{Var}_t(m_{t+1}) &= \sigma^2(\gamma + \alpha\lambda_t)^2. \end{aligned}$$

So that

$$(3.3) \quad r_{t+1}^f = -\ln \delta + \gamma \mu^c - \alpha E_t[y_{t+1}^2] - \frac{1}{2}\sigma^2(\gamma + \alpha\lambda_t)^2,$$

with $E_t[y_{t+1}^2] = \bar{y}^2 + \theta\sigma\lambda_{t-1}\eta_t$.

Let's consider the Sharpe ratio next. In a Gaussian economy, the maximal Sharpe-ratio (from the Hansen-Jagannathan bound) is

$$\max_{\{\text{all assets}\}} \frac{E_t(R_{t+1} - R_{t+1}^f)}{\sigma_t(R_{t+1})} = \frac{\sigma_t(M_{t+1})}{E_t(M_{t+1})} = (e^{\sigma_t^2} - 1)^{1/2} \simeq \sigma_t,$$

where $\sigma_t = \sqrt{\text{Var}_t(m_{t+1})}$. So the maximal Sharpe ratio for this economy is approximately

$$(3.4) \quad \frac{\sigma_t(M_{t+1})}{E_t(M_{t+1})} = \sigma(\gamma + \alpha\lambda_t).$$

A countercyclical Sharpe ratio implies that the last term in the risk free rate equation (3.3) is higher in recessions (in absolute value), which means the risk free rate will be lower as the precautionary savings motive rises. In the data, the risk free rate is procyclical, but the standard deviation is only about 1.5%, very low compared to the standard deviation of the S&P 500, which is about 18%. This means that the effect of greater precautionary saving is in large part offset by an intertemporal substitution effect that makes agents want to borrow against future growth during recessions. Given the low variation in interest rate data, we decided to follow Campbell and Cochrane (1999) and match a constant interest rate; in terms of our model, we will assume that the intertemporal substitution effect completely offsets the larger precautionary saving motive in recessions. This amounts to choosing θ to be positive in the MA(1) process for y_{t+1}^2 . Say at time $t - 1$ the system is at steady state. Then $E_{t-1}y_t^2$ is the steady state constant \bar{y}^2 . If there is a recession at t , λ_t is high, which tends to increase the precautionary saving motive, thus lowering interest rates. With a positive θ , $E_t y_{t+1}^2$ is lower than \bar{y}^2 . This gives the agents an incentive to borrow during recessions (the intertemporal substitution motive), which tends to increase the risk free rate. By assuming the two motives exactly offset each other, our model will generate an

equity premium without a term premium, as in Campbell and Cochrane. Many models with changing risk, like ours, imply high variation in R_t^f and a high term premium (Jermann, 1998; Boldrin, Christiano, and Fisher, 1997), which are not empirical.

These considerations lead to the following definition of λ_t :

$$(3.5) \quad \lambda_t = \frac{1}{\sigma\alpha} \sqrt{2[c - \alpha(\theta\lambda_{t-1}\sigma\eta_t)]} - \frac{\gamma}{\alpha}.$$

Notice that λ_t is greater in recessions (when η_t is negative) than in expansions (when η_t is positive). The constant c is calibrated so that λ_t is always defined by (3.5) in our simulations, and it yields a risk free rate of 1.5% in (3.3).¹⁴ Appendix A details how the model is solved numerically and simulated.

By way of comparison, consider the maximal Sharpe ratio generated by the economy of Campbell and Cochrane (their equation (7)):

$$\max_{\{\text{all assets}\}} \frac{E_t(R_{t+1} - R_{t+1}^f)}{\sigma_t(R_{t+1})} = \gamma\sigma(1 + \lambda(s_t)).$$

The variable s_t is the log of what they call the surplus consumption ratio, i.e., consumption above habit divided by consumption. The variable s_t is low in bad times. The function $\lambda(s_t)$ is the conditional standard deviation of s_t , and it is high in bad times. The process s_t in Campbell and Cochrane (1999) is a persistent AR(1). The analog of $\lambda(s_t)$ in Campbell and Cochrane is our λ_t . But of course the interpretation is completely different because ours is not a habit model.

We emphasize that the time variation in λ_t is needed only to match *variation* in the Sharpe ratio, (i.e., variation in the price of risk), not the size of the Sharpe ratio, nor the equity premium and the level of the risk free rate. The latter features can be matched even with the simpler process (2.5), that is, using only the basic model in the previous section.

3.2. Parameter Specification. Parameters are presented in Table 1. The mean and the standard deviation of consumption growth, μ^c and σ , are the same values used by Bansal and Yaron (2004). They match the BEA data on real per capita consumption of nondurables and services for the period 1929-1998. The risk free rate r^f that we decided to match is somewhat higher than the one used by Campbell and Cochrane, who use 0.94%, but it is lower than the average annualized log-rate on six

¹⁴To derive (3.5), rewrite (3.3) as

$$\underbrace{r^f + \ln \delta - \gamma\mu^c + \alpha\bar{y}^2}_{-c} = -\alpha\theta\sigma\lambda_{t-1}\eta_t - \frac{1}{2}\sigma^2(\gamma + \alpha\lambda_t)^2$$

and solve for λ_t .

month commercial paper of 2.4% in Shiller’s dataset, which covers the sample 1871-2002.¹⁵ Siegel (1999) presents evidence that a value of 1% underestimates the return on treasury bills; the real rate both during the nineteenth century and after 1982 has been substantially higher than 1%.

The coefficients ϕ , φ_e , and ϕ_d in x_t , the persistent component in the dividend process, are chosen so that the Barksy and De Long effect powerfully affects price movements, and at the same time the implied dividend process is consistent with the data. Table 2 presents evidence that our dividend process has time series features similar to the observed data. The persistence parameter $\rho = 0.94$ is less than unity, so the dividend growth process is stationary. Bansal and Yaron use the somewhat higher value of 0.98.

TABLE 1. Parameter Choices

| Parameter | | Value |
|--|-----------------|------------------|
| Mean consumption growth (%) | μ^c | 1.89 |
| Standard deviation of consumption growth (%) | σ | 2.9 |
| Log risk-free rate (%) | r^f | 1.5 |
| Mean dividend growth (%) | μ^d | 1.3 |
| Persistence shock coefficient | ϕ | 3.2 |
| Volatility of dividend growth | φ_d | 3.2 |
| Persistence Parameter in x_t | ρ | .94 |
| Volatility of persistence shock (%) | φ_e | 9 |
| Mean idiosyncratic shock (%) | \bar{y}^2 | 6.1 ² |
| Correlation between g_t^c and g | $\rho_{\eta,u}$ | .3 |
| Subjective discount factor | δ | .91 |
| Relative risk aversion | γ | 10 |
| MA(1) coefficient in y_t^2 | θ | 0.45 |
| Constant in λ_t | c | 0.268 |

Annual values. $\alpha = \frac{1}{2}\gamma(\gamma + 1) = 55$.

The average standard deviation of the idiosyncratic shock is 6.1%. This is well below the cross-sectional variation in permanent consumption shocks reported in Carrol (1992) using PSID data and in Cogley (2002) using CEX data. We present results with higher cross-sectional heterogeneity in Section 4.4. We choose 0.3 as the correlation between consumption growth and dividend growth. There are different estimates of this correlation. Campbell and Cochrane use 0.2; Bansal and Yaron use

¹⁵See Robert Shiller’s website. As a technical aside, the use of a higher rate in our simulations helps ensure that draws from the conditional distribution of y^2 (which is a normal variate) are not negative.

0.55. We found that values in this range do not alter the results significantly. The subjective discount factor is chosen to match the risk free rate of 1.5%. The risk aversion parameter we use, $\gamma = 10$, is admittedly high, although some estimates do exceed 10 (see Parker and Julliard, 2005, for example). A high value of the risk aversion parameter helps to get a high equity premium and, in this model, to get a low risk free rate. We chose 10 because it is the highest value in the range considered plausible by Mehra and Prescott (1985), and at the same time it is a lower bound among the models that succeed in matching similar dimensions of the data. *Section 4.4 will show there is an interesting tradeoff: by increasing the cross-sectional heterogeneity, we can match the data even with a risk aversion parameter γ below 10.*

For the MA parameter θ , since it affects the stationary distribution of y_t^2 , we pick it so that we can control the tails of the stationary distribution and prevent y_t^2 from being negative, while allowing for variation in the Sharpe ratio. To give time variation to the Sharpe ratio, we need the conditional variance of y_t^2 to be time varying. We keep the variance of y_t^2 small to avoid negative values and implausibly high positive ones. The unconditional distribution of y_{t+1}^2 is summarized on the right hand panel of Table 2. We argue below that the unconditional distribution of y_{t+1}^2 is very plausible and lies on the conservative side of the evidence found in panel data.

3.3. Implied Processes. Prior to commenting on the ability of the model to replicate the empirical evidence, it is important to evaluate empirically the stochastic process for dividend growth and individual consumption, which we take as exogenous.

Consider dividends first. Statistics for the simulated dividend process are reported on the left panel of Table 2. The table shows that the time series properties of our generated process are similar to the ones found in the data. The autocorrelation functions in the data are more complex and show some negative signs. But these are not significant, and could be matched if we gave a weak lag structure to u_t in the dividend process rather than a simple i.i.d. Complex time series dynamics are just not detected in the data for the dividend process. The autoregression coefficient $\rho_{AR(1)}$ of the simulated series is even lower than in the data, hence our parameterization for dividends does not impose an unrealistic persistence (at least as measured by a standard AR(1) model). Also, based on the data, we would not be able to reject the AR(1) parameters implied by the simulated series.¹⁶

¹⁶The AR(1) is a common specification for the law of motion of dividend growth.

TABLE 2. Implied g and y^2 Processes

| Statistic | Model | Data | | y^2 process | |
|----------------|-------|----------|--------|-------------------|-------------|
| | | Estimate | 2×S.E. | | |
| μ^d | 1.3% | 1.0% | | \bar{y}^2 | $(6.2\%)^2$ |
| $\sigma(g)$ | 9.5% | 11.6% | | $\sigma(y^2)$ | .00186 |
| AC(1) | .08 | .13 | .19 | CI_y^U | 8.7% |
| AC(5) | .07 | -.02 | .19 | CI_y^L | 1.1% |
| AC(10) | .04 | .15 | .2 | $c_{AR(1)}$ | .005 |
| $c_{AR(1)}$ | .012 | .009 | .022 | $\rho_{AR(1)}$ | -.25 |
| $\rho_{AR(1)}$ | .085 | .123 | .188 | $\rho(y^2, g^c)$ | -.69 |
| $\rho(g^c, g)$ | .30 | .28 | | $\beta(y^2, g^c)$ | -0.044 |

Notes: μ^d and $\sigma(g)$ are the mean and standard deviation of dividend growth. AC(j) is the j -th autocorrelation. $c_{AR(1)}$ and $\rho_{AR(1)}$ are the constant and autoregression coefficients in the AR(1) model. $\rho(g^c, g)$ is the correlation between per-capita consumption growth and dividend growth. \bar{y}^2 is the average cross-sectional variance in the model, $\sigma(y^2)$ is the unconditional standard deviation of y_t^2 . CI_y^U and CI_y^L are the upper and lower bounds, respectively, of a 95% symmetric confidence interval for y (not y^2), and $\rho(y^2, g^c)$ is the correlation between the cross-sectional variance of consumption growth y_t^2 and per-capita growth g_t^c . The data used for the dividend process is from the annual dataset 1890-2001 on consumption and the S&P 500 dividend downloadable from Robert Shiller's website.

Moving to idiosyncratic risk, the right panel of Table 2 shows some summary statistics for the stationary distribution of y_t^2 . The mean of the generated process is 6.2%. We report a 95% confidence interval for y_t (not y_t^2), which represents the standard deviation of the idiosyncratic shock, to get an idea of the implications of the generated distribution in terms of cross-sectional inequality. With 95% probability, the cross-sectional standard deviation of individual shocks is between 1.1% and 8.7%. These are definitely not extreme values considering panel data evidence on cross-sectional variation in income and consumption (see for example Cogley, 2002, or Carrol, 1992). Carrol uses a value of 10% for *permanent* income shocks, after adjusting for measurement error, in his study of precautionary savings. Using CEX data, Cogley finds values for the cross-sectional variation on the order of 35-50%! He also warns us of the high measurement error in the data. A value of 6.2% corresponds to assuming that only 2.5% of the overall variation found in the data is true cross-sectional variation in consumption growth, while 97.5% is due to measurement error in the CEX data.

One important feature of the generalized model is the time varying individual risk given by the mixture distribution for idiosyncratic shocks, which now varies with aggregate shocks through λ_t . We can estimate the variation of idiosyncratic risk over the “business cycle” by regressing y_t^2 on per capita growth g_t^c . The projection coefficient β reported in the Table is -0.044 . This value implies that idiosyncratic risk varies from 4.9% if $g_t^c = \mu^c + \sigma$ (an expansion) to 7.1% if $g_t^c = \mu^c - \sigma$.¹⁷ Storesletten, Telmer and Yaron (2004) show evidence that cross-sectional risk varies between 12% (expansion) and 21% (recession). Our parameters thus lie on the conservative side of the empirical evidence. Section 4.4 will show that greater variation — still within the plausible range — permits the model to explain the stylized facts about asset pricing with lower values of risk aversion than $\gamma = 10$.

It is worth re-emphasizing that we used the observation that the Sharpe ratio is countercyclical and that the risk free rate is very smooth in U.S. data to specify λ_t , but no data on stock prices or returns has been used to calibrate the process. Thus, evidence presented below in Sections 4.1, 4.2, and particularly in Section 4.3 and Figure 1 present additional, although indirect, evidence in support of our stochastic process.

4. SIMULATION RESULTS

We simulate a history of 100,000 draws from our economy and calculate sample statistics that we then compare to the post war sample and long sample moments from Robert Shiller’s data on the S&P 500 (1871-2001) and consumption (1889-2001).¹⁸ Table 3 summarizes the comparison. We then turn to the long horizon predictability of returns in Table 5.

4.1. Matching Moments. As discussed above, our parameter choice implies a risk free rate of 1.5%. Given the parameters, the model generates an equity premium of 6%, which is the level that Mehra and Prescott tried to explain. Notice that there is some in-sample variation in excess returns (5.7% and 7.2% in the two samples); our value of 6% indicates that the model can explain most of the equity premium. The model also generates a standard deviation that is consistent with the data, 18%, and therefore a Sharpe ratio consistent with the data.

The level of the price-dividend ratio is 20.9, close to the long sample value, but lower than the postwar average. The standard deviation of the price-dividend ratio is also lower than the level observed, 24% versus

¹⁷Alternatively, using NBER indicators, per capita growth is 2.9% in expansions and 0.8% in recessions, which implies, given $\beta = -0.044$, that idiosyncratic risk varies between 4.9% and 6.4%.

¹⁸The data can be downloaded at <http://www.econ.yale.edu/~shiller/data.htm>.

TABLE 3. Moments of Simulated and Historical Data.

| Statistic | Model | Long Sample* | Postwar Sample* |
|----------------------|-------|--------------|-----------------|
| $\exp(E[p - d])$ | 20.92 | 22.7 | 28.0 |
| $\sigma(p - d)$ | .24 | .35 | .40 |
| $\sigma_a(p - d)$ | .24 | .26 | .29 |
| $\sigma_{93}(p - d)$ | .24 | .26 | .26 |
| $E_t[R - R^f]$ | 6.0% | 5.7% | 7.2% |
| $\sigma(R - R^f)$ | 18% | 18% | 15.0% |
| Sharpe Ratio | .33 | .32 | .46 |
| $AR(1)(p - d)$ | .91 | .84 | .86 |
| $\rho(g^c, r)$ | .57 | .37 | .34 |
| $\rho(g^c, p - d)$ | .34 | .16 | .20 |

Notes: The Long Sample is 1871-2001, and the Postwar Sample is 1950-2001. $\exp(E[p - d])$ is the average price dividend ratio. $\sigma(p - d)$ is the standard deviation of the log price-dividend ratio. $\sigma_a(p - d)$ is the standard deviation of the log dividend-price ratio adjusted for repurchases as in Appendix B and $\sigma_{93}(p - d)$ is the standard deviation using data up to 1993. $R - R^f$ is the excess return on the stock market relative to the risk free interest rate. $\rho(g^c, r)$ is the correlation between consumption growth and log returns, and $\rho(g^c, p - d)$ is the correlation between consumption growth and the log price-dividend ratio. $AR(1)(p - d)$ is the first-autocorrelation coefficient of $p - d$.

35% and 40% in the two samples. Competing explanations, like Campbell and Cochrane (1999) and Bansal and Yaron (2004), generate a volatility of 26% and 21% respectively. It is of interest to notice that the higher observed variation is entirely due to the period 1994-2001, even in the century long sample. The value of the standard deviation for both long and post-war samples up to 1993 is only 26%.

There is empirical evidence that the price-dividend ratio shifted to a new regime over the nineties (see Lettau, Ludvigson and Watcher, 2003; Boudoukh et al., 2003), which could explain a higher mean and a higher variance in the 1990's. In particular, Boudoukh et al. study the role of stock repurchases on the price-dividend ratio. They show that a constructed payout ratio (i.e., price divided by dividend plus repurchases) has the same time series properties as the price-dividend ratio prior to the 1990's, but much different during that decade. The values under $\sigma_a(p - d)$ in the table refer to the standard deviation of the price-dividend ratio adjusted to take stock repurchases into account. They are 26% and

29% respectively for the two samples. We discuss repurchases and the adjustment made to the price-dividend ratio in section 4.3 and Appendix B. This seems to explain almost entirely both the higher volatility and the higher mean in the late nineties.

The simulated log price-dividend ratio has an autocorrelation of 0.91, close to the 0.84 and 0.86 of the long and post war samples, respectively.

The model has a consumption growth-stock return correlation of 57%. This is far lower than the almost perfect correlation that Mehra and Prescott's economy or most consumption based asset pricing models imply, albeit higher than the 37% found in the long sample. Time aggregation alone can explain such discrepancy between model and data. Ermini (1991) derives a formula for the reduction in covariance between the consumption growth and returns as a function of the ratio between the interval of observation, and the decision interval, in discrete time. If the decision interval is monthly, and the interval of observation is annual (as in our calculation of observed statistics), the covariance between annual (time averaged) consumption growth and annual returns is approximately two-thirds of the covariance of the original (monthly) processes. That is, if we simulated our model at a monthly frequency (with appropriate monthly parameterization), and then calculated annual (time averaged) correlation between consumption growth and returns, the correlation would decrease approximately by 1/3, thus making the model's correlation almost equal to the correlation in the data ($2/3 \times 57\% = 38\%$).

Further, the correlation coefficient is hard to estimate and estimates vary widely across studies.¹⁹ Campbell, Lo, and MacKinlay (1997) report a correlation of 50%, very close to our generated value. Given that time aggregation substantially reduces correlation, and given the uncertainty in estimating the parameter, we feel that the generated 57% is consistent with the data.

Table 4 compares simulations of the generalized model, the basic model in Section 2, and a model in which the Barsky and De Long component to dividend growth is shut down, i.e., $\phi = 0$. As the table shows, the Barsky and De Long component is the one that generates the premium and the volatility of the price-dividend ratio. The basic model matches the moments almost as well as the generalized model, while removing the Barsky and De Long component destroys the model's ability to match the data. Individual risk alone (aggregate plus idiosyncratic) generates a premium of only 2.8%.

¹⁹This is in accord with Barsky and De Long's insight that the small persistent component in dividend growth is hard to detect.

TABLE 4. Comparison of Models

| Statistic | Generalized Model | Basic Model | $\phi = 0$ |
|--------------------|-------------------|-------------|------------|
| $\exp(E[p - d])$ | 20.92 | 25.67 | 39.16 |
| $\sigma(p - d)$ | .24 | .24 | .04 |
| $E_t[R - R^f]$ | 6.0% | 5.1% | 2.8% |
| $\sigma(R - R^f)$ | 18% | 18.16% | 9.0% |
| Sharpe Ratio | .33 | .28 | .29 |
| $\rho(g^c, r)$ | .57 | .57 | .21 |
| $\rho(g^c, p - d)$ | .34 | .34 | -.07 |

Notes: The Generalized Model is the model of Section 3, while the Basic Model is the model of Section 2, without time-varying uncertainty. The last column, under the heading $\phi = 0$ shows results from a model in which the “Barsky-De Long” component is shut down.

As motivated in Section 3.1 above, time-varying risk (equation 3.2) is important for replicating the observed predictability of excess returns. We now turn to predictability.

4.2. Autocorrelations and Return Predictability. Table 5 shows results from long horizon regressions, i.e. we regress cumulative log excess returns, $r_{t,t+1}^e + r_{t,t+2}^e \cdots + r_{t,t+j}^e$, on $p_t - d_t$, the log of the price-dividend ratio at time t . We do this for $j = 1, \dots, 5$.

TABLE 5. Long-Horizon Regressions

| Horizon | Model | | Long Sample | | Post War | |
|---------|-------------------|-------|-------------------|-------|-------------------|-------|
| | $10 \times \beta$ | R^2 | $10 \times \beta$ | R^2 | $10 \times \beta$ | R^2 |
| 1 | -1.21 | .12 | -2.6 | .08 | -3.4 | .16 |
| 2 | -1.83 | .20 | -3.4 | .13 | -3.4 | .26 |
| 3 | -1.95 | .26 | -4.9 | .20 | -4.8 | .38 |
| 4 | -2.02 | .32 | -7.3 | .36 | -10.2 | .46 |
| 5 | -2.13 | .37 | -9.3 | .31 | -14.2 | .50 |

Notes: The dependent variable is $r_{t,t+1}^e + r_{t,t+2}^e \cdots + r_{t,t+j}^e$, where j is the horizon. The independent variable is $p_t - d_t$.

We observe the pattern documented in Campbell and Shiller (1988) and Fama and French (1988): an increase in the price-dividend ratio forecasts lower future excess returns, and the R^2 increases with the horizon. Absolute values of the regression coefficients are lower than in the data, but these are imprecisely estimated, and only the post war 5-year coefficient is statistically different from the model’s.

Recall that the basic model (Section 2) implies constant expected excess returns, hence it cannot generate this kind of predictability. Both coefficients and R^2 are zero in the basic model.

4.3. Model Implications from Historical Consumption Data. We take the consumption series from Shiller's data (1890-2001) and use the parameterization in (3.1) to calculate the series of η_t . We then feed the η_t shocks to the generalized model, which produces a series of price-dividend ratios (so no stock market data is used in this simulation).²⁰ We adjust model predictions for the period 1972-2000 to take into account that repurchases have become an important source of payout for U.S. corporations using the data of Table I in Grullon and Michaely (2002). This makes the model's predictions comparable with the observed price-dividend ratio. Details of the adjustment are in Appendix B. Figure 1 presents a comparison between data on the S&P 500 (from Shiller's dataset), and the price generated by the model using consumption data. Figure 2 shows the behavior of the implied price-dividend ratio versus the real S&P 500 price series.

Clearly, given the simple nature of the model, we do not expect the implied price to match the data point by point, but the impression from the figures is that the model does a very satisfactory job in capturing some of the main events in the stock market. The correlation coefficient between the simulated price-dividend series and the observed price-dividend ratio is 72%.²¹

Successive declines in consumption drive down future expected growth. As a consequence, prices fall. The model thus accounts for the decline in prices relative to dividends in the sharp recession of 1908 and in the post WWI period. The model also captures the rise and then decline of the stock market in the early 20th century, 1901-1918, the roaring 20's and the Great Depression. It predicts an even larger drop in 1932 than actually happened, so extreme was the drop in consumption growth, which lowered expected dividend growth. Then the model tracks the recovery during WWII, and the consumption and stock market boom of the sixties, though with a lag of about 5 years. It also tracks the poor performance of the 70's and the recovery of 1982-87.

²⁰The shocks u_t are not needed as they are i.i.d.; all that is needed to calculate prices is their variance $\sigma\phi_d$, as shown in the Appendix.

²¹The unadjusted series and the adjusted one have similar correlation coefficients with the S&P data up to 1990, 47% and 44% respectively. The coefficient of the adjusted series increases during the run-up in prices of the 1990's, when repurchases were highest.

FIGURE 1. Historical S&P 500 Real Price and Model Predictions

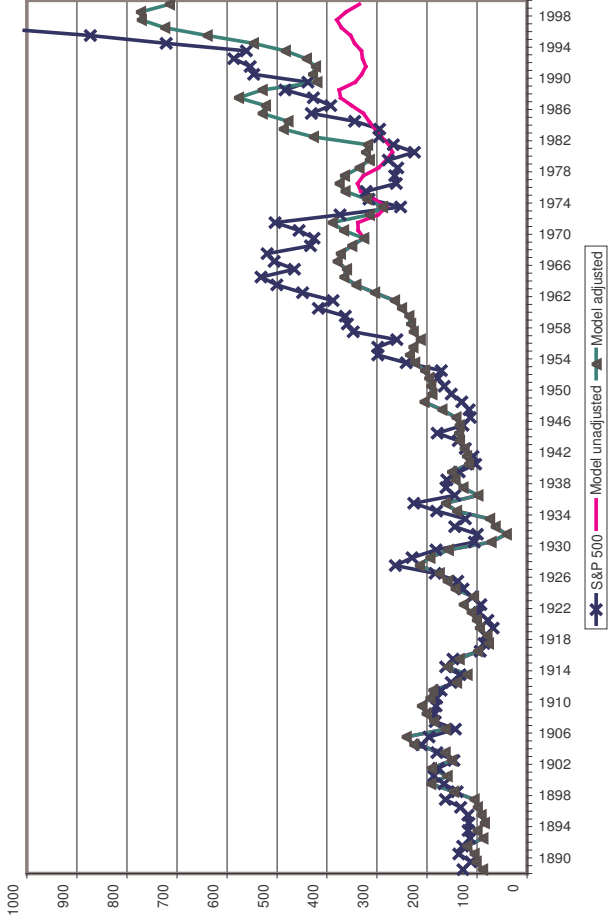
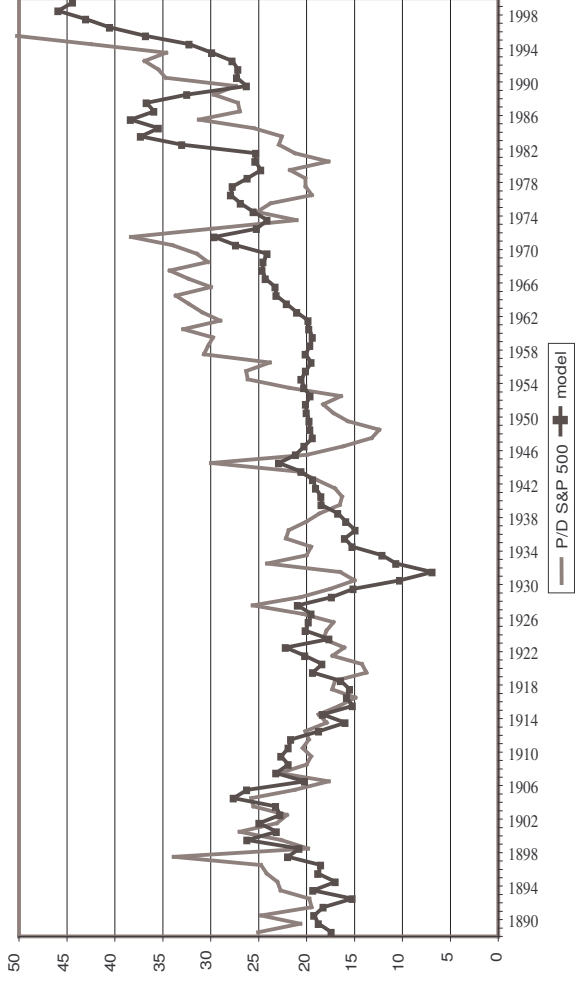
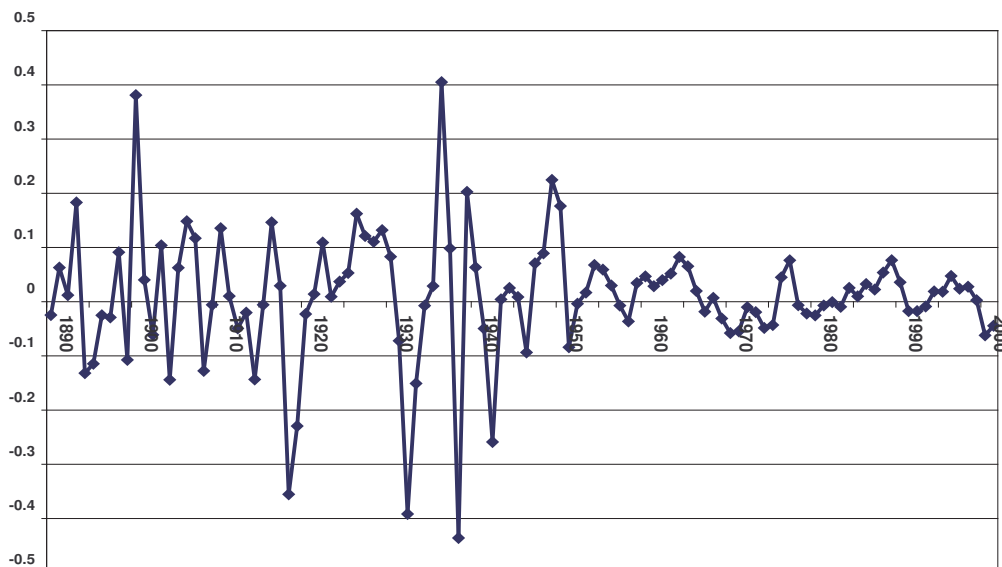


FIGURE 2. Historical Price-Dividend Ratio and Model Predictions



The worst performance of the unadjusted series (i.e., without adjusting for repurchases) is from the second half of the eighties and during the nineties when stock prices hit unprecedented levels at high speed. The unadjusted series moves with the adjusted series, but it does not vary enough. This is a common feature of models that try to explain stock market behavior using consumption data (see Campbell and Cochrane, 1999). Consumption growth was very smooth over the nineties, there was not a series of large positive shocks that is needed to move prices upward relative to dividends in the model. More generally, the smoothness of consumption applies to the entire second half of the figure, say after 1950. Consequently, model predictions after 1950 are much smoother relative to the S&P movements than before 1950. Figure 3 plots the consumption growth series used in our simulation, downloaded from Robert Shiller's web site. The standard deviation of the series is about 3%, but notice the huge difference in swings before and after 1950.

FIGURE 3. Consumption Growth



We find that we can reconcile part of the smoothness of consumption with greater swings in the price-dividend ratio if stock repurchases are taken into account. Our theoretical model's predicted price-dividend ratio should really be thought of as the predicted total payout rate, that is price divided by dividends plus repurchases. During most of our long sample, which runs from 1871 to 2001, repurchases were not significant.

Hence the model's predicted payout ratio and predicted price-dividend ratio would approximately coincide. But in the 1980's and 1990's repurchases became a significant fraction of total payout ratio to shareholders as documented in Grullon and Michaely (2002). While between 1972 and 1983 repurchases amounted to an average of 10.9% of dividend payments, between 1984 and 2000 repurchases were 57.7% of dividends, reaching a maximum of 113.11% in 2000. Most importantly, firms financed their share repurchases with funds that would otherwise have been used to increase dividends. This means that the price-dividend ratio in the data is much larger than the true payout ratio towards the end of the sample, i.e., there was a regime shift. To take account of this shift, we adjust the model's predicted price-dividend ratio series as suggested by Boudoukh et al. using Grullon and Michaely's Table 1. Details are in Appendix B. The adjustment can account for most the movements in the eighties and nineties while it does not help at the end of sample, say 1997-2001.

As yet, it is not clear how and whether we can rationalize the level of prices in the late 1990's. Possible explanations include compositional shifts of consumption towards durable goods, or an increase in the consumption of stock market investors that is not captured in the data, due perhaps to demographic effects of the baby boom generation entering peak saving years. Eugene White (2004) argues that there is fair evidence of a bubble in the late 1990s. As Campbell (1999) points out:

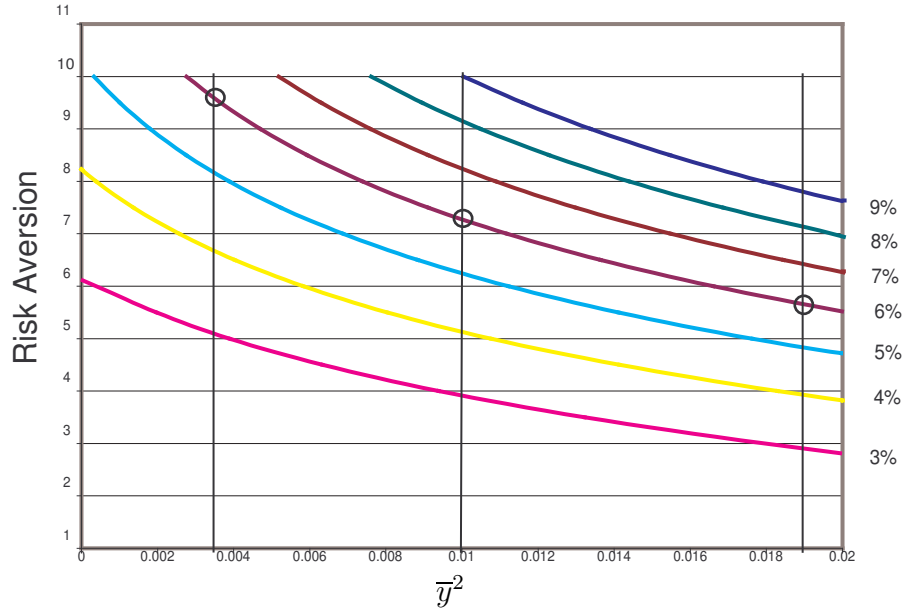
The recent run-up in stock prices is so extreme relative to fundamental determinants such as corporate earnings, stock-market participation, and macroeconomic performance that it will be very hard to explain using a model fit to earlier historical data. (p. 261)

While we leave open the possibility of a bubble in the late 1990s, we believe the model (and therefore fundamentals and rational expectations) can explain major movements in prices over the century-long sample.

4.4. Risk Aversion. The risk aversion coefficient used is 10, which is the upper bound of plausible parameters according to Mehra and Prescott (1985). The general view among economists is that 10 is too high (see Kocherlakota, 1996).

There is a sense in which a value of 10 can be justified for our economy, which comes from comparing our pricing equation with the one from a representative agent model. We can think of our economy as isomorphic to a representative agent economy in which utility and MRS's depend on y_t^2 . The representative agent will then exhibit high risk aversion, as he does in Campbell and Cochrane where risk aversion is as high as 80 at steady state. Compared to 80, our economy shows that, by adding a simple form

FIGURE 4. Tradeoff Between Risk Aversion and Idiosyncratic Risk



The circles correspond to the three alternative pairings of risk aversion and idiosyncratic risk that give a 6% equity premium, as discussed in the body of the paper. The circle on the upper left corresponds to the benchmark parameterization.

of heterogeneity to income shocks, one can go a long way toward reducing the level of risk aversion for individual agents. It is plausible to think that allowing for more general forms of heterogeneity could further reduce the level of risk aversion. It is unlikely though that such a model would yield a pricing function that depends only on aggregate variables.

We turn now to an interesting tradeoff. Recall that in our benchmark simulations, based on the parameter choices in Table 1, we used a very conservative specification for y_t^2 . Equation (2.11) shows that, for achieving a given level of the equity premium, there is a tradeoff between the level of risk aversion and the level of uncertainty about idiosyncratic shocks (φ_y in (2.11) and λ_t in the model of section 3). Figure 4 illustrates the contour lines in $\gamma - \bar{y}^2$ space; each contour line shows the combinations of risk aversion and \bar{y}^2 needed to achieve any given level of the equity premium. These are calculated from our benchmark distribution for y_t^2 by increasing the mean \bar{y}^2 and the standard deviation φ_y in (2.11), keeping the coefficient of variation constant.

For example, consider the vertical line through $\bar{y}^2 = 0.01$. This value for \bar{y}^2 corresponds to a mean value for y_t of 10%, and φ_y is calculated so that with 99% probability y_t is between 0 and 20%. When $\bar{y}^2 = 0.01$ and risk aversion $\gamma = 10$, the equity premium will equal about 9%, since the 9%-contour line crosses the vertical line at about $\gamma = 10$. Even if risk aversion falls to $\gamma = 5$, we would get a premium only slightly lower than 5% when $\bar{y}^2 = 0.01$. The vertical line through $\bar{y}^2 = 0.0038$ corresponds to our benchmark simulation, hence when $\gamma = 10$ the premium equals approximately 6%.²² Finally, the vertical line through $\bar{y}^2 = 0.019$ corresponds to assuming that with 95% probability, the cross-sectional variance is less than 15% (close to Cogley's 2002 value), which implies a standard deviation of 38%. In other words, we make the 38% standard deviation a very unlikely event, a 2- σ event. In this case, when risk aversion γ lies between 5 and 6, the risk premium still lies between 5 and 6%.

Table 6 presents two simulations suggested by Figure 4. The first is with $\gamma = 7$ and a value of $\bar{y}^2 = .01$, and the second is with $\gamma = 5$ and $\bar{y}^2 = .019$. Figure 4 shows that both these combinations should give us an equity premium close to 6%. This is confirmed in the table. The table shows that the model can reproduce a high premium with a risk aversion as low as 5, while still matching other relevant moments.²³

TABLE 6. Lower Risk Aversion and Greater Heterogeneity

| Statistic | Benchmark $\gamma = 10$ | $\gamma = 7$ | $\gamma = 5$ |
|--------------------|-------------------------|--------------|--------------|
| $\exp(E[p - d])$ | 20.92 | 22.5 | 24.8 |
| $\sigma(p - d)$ | .24 | .24 | .28 |
| $E_t[R - R^f]$ | 6.0% | 5.6% | 5.4% |
| $\sigma(R - R^f)$ | 18% | 16.4% | 17.8% |
| Sharpe Ratio | .33 | .34 | .30 |
| $\rho(g^c, r)$ | .57 | .79 | .82 |
| $\rho(g^c, p - d)$ | .34 | .37 | .36 |

Notes: We pair $\gamma = 10, 7, 5$ with $\bar{y}^2 = 0.00372, 0.01, 0.0184$ respectively. That is, higher levels of risk aversion are paired with lower levels of idiosyncratic risk. These levels of variance correspond to standard deviations of 6.1%, 10%, and 13.6% respectively.

²²The contour lines are computed using the approximate solution (2.11), so the vertical line does not exactly cross at 6% as one would expect from the benchmark simulation.

²³Parameters used are the same as in Table 1, with the exception of $\phi = 3.5$, $\varphi_d = 2.8$, $\rho = .925$, $\varphi_e = 10\%$, $\delta = .8$. These values slightly improve the fit of the simulation and do not significantly alter the implied stochastic process for dividend growth.

There are other possible modifications of the model that would go in the same direction as higher cross-sectional variance, and could allow a lower risk aversion for a given level of the premium. Consider the effect of a probability mass of consumption close to zero, as argued in Carroll (1992), or a possible multi-dimensionality of individual shocks whose relative importance varies over the business cycle, as in Krebs (2003).

5. RELATIONSHIP TO EXISTING LITERATURE

Our theory touches on several strands of the vast literature that seeks to document and resolve asset pricing puzzles. Surveys are found in Kocherlakota (1996), Campbell (2003), and Cochrane (2001). The objective here is to relate our model to that part of the literature that directly motivated our maintained assumptions, and also to relate our model to predecessors that — by using non-standard preferences, — were able to generate asset pricing features similar to ours.

Mehra and Prescott’s seminal paper is based on three assumptions: individuals maximize the expected discounted value of a stream of utilities generated by a *power utility function*; markets are *complete*; asset trading is *costless*. Taking their economy as a point of departure, our model postulates a different process for dividend growth and relaxes the assumption of complete markets.²⁴

As detailed in Section 2, we take the idea of a persistent component in shocks to dividend growth from Barsky and De Long (1993). We show how to incorporate their idea into a general equilibrium asset pricing model, and hence how to apply it to explaining the equity premium and related puzzles in a model with standard preferences. Barsky and De Long were only interested in the excess volatility puzzle (3). With our strengthening of their hypothesis, we show that their explanation of (3) survives in a general equilibrium framework. Further, the strengthening allows us to explain most of the equity premium puzzle (1), and the persistence of the price-dividend ratio (4).

Bansal and Yaron (2004) also use the Barsky and De Long idea. But they do so in a representative agent model in which individuals have Epstein-Zin-Weil preferences and an intertemporal elasticity of substitution (*IES*) greater than one. We discussed how in their model the high risk premium, the low risk free rate, and cyclicity of the price-dividend ratio depend crucially on the *IES* being greater than one. While the authors rightly point out the difficulties with estimating this elasticity, most

²⁴In the generalized model, it also adds fluctuating uncertainty about individual uninsurable shocks.

empirical evidence indicates a value less than unity (Yogo, 2004; Campbell, 2003; Vissing-Jorgensen, 2002; Hall, 1988). Indeed, the evidence in Hall, Campbell, and Yogo even casts doubts on a non-zero intertemporal elasticity of substitution. By contrast, we assume a low intertemporal elasticity of substitution, 0.1 to 0.2, consistent with the findings in Yogo. Intuitively, accepting a parameterization of preferences with relatively high risk aversion and $IES > 1$ means accepting that agents are quite averse to consumption fluctuations across states but quite flexible about consumption fluctuations across time. The main justification for Epstein-Zin-Weil preferences comes from Kreps and Porteus (1978, 1979), and runs in terms of a preference for flexibility. But the Kreps-Porteus story is based on *income* uncertainty, not consumption uncertainty as in Bansal and Yaron. So we do not know of any good economic story that would justify their choice of preference structure.

In our treatment of incomplete markets and idiosyncratic shocks, we follow Constantinides and Duffie (1996). Their main no-trade theorem shows that, under some assumptions, there exists a stochastic process for individual income that potentially can rationalize the joint behavior of aggregate consumption and asset prices, which is equation (2.4). In their model, only the cross-sectional variation in income matters for asset prices; they show that this cross-sectional variation has to co-move with returns and aggregate consumption in an appropriate way to generate an equity premium, in particular, idiosyncratic risk must be negatively correlated with aggregate shocks (like in our basic model of Section 2).

Recent econometric work has studied the relevance of cross-sectional variation in consumption (idiosyncratic risk) for explaining the equity premium. The evidence is mixed. Using CEX data, Cogley (2002) finds that cross-sectional variation generates premia of 2% or less for preferences with a low degree of risk aversion. Brav, Constantinides, and Geczy (2002) find that a pricing equation which takes individual risk into account is not rejected in the CEX data. We are pleased that our model of idiosyncratic risk is based on conservative specifications, in the sense of being consistent with the findings of Cogley. As the mixed empirical findings suggest, the Constantinides and Duffie model is far from enough to take us all the way.

In this paper we go beyond the basic framework of Constantinides and Duffie. We supplement their abstract model with a stochastic process for aggregate consumption, dividend growth, and idiosyncratic risk. In accord with the high standards set by Campbell and Cochrane, our specification does not just aim at matching the equity premium, but also at successfully matching the other major empirical features of asset prices, including the low risk free rate and excess return predictability. As discussed in Sections 1 and 2, our strengthening of the Barsky-De Long hypothesis

explains most of puzzle (1), the high equity premium. *Idiosyncratic risk is mainly important for explaining puzzle (2), the low risk free rate.*

Any model that wants to generate time variation in the price of risk and predictable excess returns needs to have a stochastic discount factor that is heteroskedastic. In Campbell and Cochrane (henceforth CC) this is achieved by postulating that the sensitivity of the marginal rates of substitution to consumption shocks changes over the business cycle, and so heteroskedasticity is part of the definition of the habit process. In Bansal and Yaron (2004), heteroskedasticity is embedded in the per-capita consumption growth process, and it is interpreted as changing risk; time variation in risk is assumed independent of the business cycle.

In our model, it is fluctuating uncertainty about the idiosyncratic shock that gives rise to a heteroskedastic stochastic discount factor. Thus, as in Bansal and Yaron, risk premia are time-varying because of changing risk, albeit of a different type. By contrast, in CC the factor that drives risk premia is changing risk aversion. This is because, as consumption moves closer to “habit,” the curvature of their utility function (the second derivative with respect to consumption) increases. This makes marginal rates of substitution more sensitive to consumption variation, an important feature to explain the equity premium. In both CC and our model, a strong precautionary saving motive in the model generates a low risk free rate. But the source of the need for insurance is different in the two models. In ours it is individual risk, and individuals’ uncertainty about the distribution of this risk, that generate a strong precautionary saving motive. In CC, the source is per capita risk, and the need for keeping consumption above habit.

Even successful models like CC have difficulty matching the price-dividend ratio in the 1990’s. In confronting the model’s predictions (when fed actual consumption shocks) with the historical behavior of the price-dividend ratio in U.S. data, it helps to take stock repurchases into account. Grullon and Michaely (2002) report evidence for the period 1972–2000 that repurchases have become an important source of payout for U.S. corporations, and firms finance their share repurchases with funds that would otherwise be used to increase dividends. Boudoukh, Michaely, Richardson, and Roberts (2003) construct a payout yield that adjusts the dividend yield for repurchases. They find that the adjustment can explain the lack of predictive power of the dividend yield over the 1990’s, and they suggest that asset pricing models that relate cash flow to asset pricing (like ours) should take this into account. We do so and find that model predictions are much improved over the period, confirming their intuition.

6. CONCLUSION

In this paper we provide an explanation for the equity premium and related puzzles based on a unifying theme: identifying the risks that firms and households really face. Our main explanation for the high equity premium is that there is a small persistent component to changes in dividend growth, which makes equity prices very volatile, while our main explanation for the low risk free rate is that individuals face not only aggregate risk, but also significant idiosyncratic risk. These are the only two departures from the model of Mehra and Prescott. Preferences are standard, and so our explanation of the major stylized facts does not rely on psychologies that may be hard to understand. Rather, in our explanation, individuals' desire for consumption smoothing is the main driving factor. To our knowledge, this is the first paper in the literature that matches relevant features of the data using standard preferences; further, we do so with plausible parameter values: risk aversion can be between 5 and 10, hence the intertemporal elasticity of substitution can be between .1 and .2.

Besides providing an explanation for the high equity premium and the low risk free rate, our model explains the extreme volatility of stock prices, the persistence in the price-dividend process, return predictability, and the low correlation between consumption growth and excess returns. These are the major, known features of aggregate stock market behavior.

Finally we show how, even with plausible parameter values, Barsky and De Long's insight — that the dividend process is much more important for price movements than simple autoregressive specifications would suggest — can be lifted to the general equilibrium level, and hence used to explain the equity premium. While we do not have production and investment in our economy, one possible interpretation of the persistent component to dividends is in terms of irreversible choice under uncertainty, which may also explain why the effects of recessions are felt more by equities than by aggregate consumption of nondurables, as we suggested in Section 2.

APPENDIX A. SOLVING FOR THE EQUILIBRIUM PRICE-DIVIDEND
RATIO

A.1. Basic Model. Notice that the stochastic process in (2.5) is linear with constant variances and covariances. This implies that the variance term in the Euler equation (2.8)

$$\ln \delta - \gamma E_t g_{t+1}^c + \alpha E_t y_{t+1}^2 + E_t r_{t+1} + \frac{1}{2} \text{Var}_t(-\gamma g_{t+1}^c + \alpha y_{t+1}^2 + r_{t+1}) = 0$$

is constant. We can substitute for r_{t+1} using the log linear approximation (2.6). Rearranging terms, this gives

$$\ln \delta - \gamma \mu^c + \alpha \bar{y}^2 + \frac{c}{2} = -E_t[\kappa_0 + \kappa_1 v_{t+1} - v_t + \mu^d + \phi x_t],$$

where c is the constant variance. Solving for v_t

$$(A.1) \quad v_t = B + \kappa_1 E_t v_{t+1} + \phi x_t,$$

with $B = \ln \delta - \gamma \mu^c + \alpha \bar{y}^2 + c/2 + \mu^d + \kappa_0$. The equation (A.1) can be solved by recursive substitution using the fact $E_t x_{t+j} = \rho^j x_t$ to yield

$$v_t = \underbrace{\frac{B}{1 - \kappa_1}}_{A_0} + \underbrace{\frac{\phi}{1 - \kappa_1 \rho}}_{A_1} x_t.$$

The variance

$$\text{Var}_t(-\gamma g_{t+1}^c + \alpha y_{t+1}^2 + r_{t+1}) = \text{Var}_t(-\gamma g_{t+1}^c + \alpha \bar{y}^2) + \text{Var}_t(r_{t+1}) + \text{Cov}_t(r_{t+1}, -\gamma g_{t+1}^c + \alpha \bar{y}^2)$$

The first variance equals $(\gamma \sigma + \alpha \varphi_y)^2$, while the second is

$$\begin{aligned} \text{Var}_t(r_{t+1}) &= \text{Var}(r_{t+1} - E_t r_{t+1}) = \text{Var}_t(\kappa_1(v_{t+1} - E_t v_{t+1}) + g_{t+1} - E_t g_{t+1}) \\ &= \text{Var}_t(\kappa_1 A_1 \varphi_e \sigma \eta_{t+1} + \sigma \varphi_d u_{t+1}) \\ &= (\kappa_1 A_1 \varphi_e \sigma)^2 + \sigma^2 \varphi_d^2 + 2\kappa_1 A_1 \varphi_e \sigma^2 \varphi_d \rho_{\eta,u}. \end{aligned}$$

So both variances are constant. Lastly, the covariance term is equal to the covariance between the innovations to the stochastic discount factor and returns, i.e.

$$\begin{aligned} \text{Cov}_t(-\gamma g_{t+1}^c + \bar{y}_{t+1}^2, r_{t+1}) &= \text{Cov}_t(-(\gamma \sigma + \alpha \varphi_y) \eta_{t+1}, \kappa_1 A_1 \varphi_e \sigma \eta_{t+1} + \sigma \varphi_d u_{t+1}) \\ &= -(\gamma \sigma + \alpha \varphi_y)(\kappa_1 A_1 \varphi_e \sigma + \sigma \varphi_d \rho_{\eta,u}), \end{aligned}$$

which is the equation of the constant premium in this economy.

A.2. Generalized Model. First we get rid of the shock u_t , since it does not determine state variables and it can be integrated out. Then we use quadrature-based rules to approximate the functional equation implied by the first order conditions. Notice now that the process (3.1)-(3.5) is a function of two shocks (u_{t+1}, η_{t+1}), and two state variables ($x_t, \lambda_{t-1}\eta_t$), or equivalently ($x_t, E_t y_{t+1}^2$). In fact, it is enough to know x_t and $E_t y_{t+1}^2$ to know the distribution of the process (3.1)-(3.2). We have:

$$(A.2) \quad \begin{aligned} y_{t+1}^2 &= E_t y_{t+1}^2 - \sigma \lambda (E_t y_{t+1}^2) \eta_{t+1}, \\ E_{t+1} y_{t+2}^2 &= \bar{y}^2 + \theta \sigma \lambda (E_t y_{t+1}^2) \eta_{t+1}, \\ \lambda (E_t y_{t+1}^2) &= \frac{1}{\sigma \alpha} \sqrt{2[c - \alpha E_t y_{t+1}^2]}, \end{aligned}$$

so we can define the state vector $s_t = (s_{1t}, s_{2t}) \equiv (x_t, E_t y_{t+1}^2)$.

Consider the Euler equation for the price-dividend ratio from (2.4) and denote by s_t the vector of state variables. Then

$$(A.3) \quad \begin{aligned} \frac{P_t}{D_t}(s_t) &= E \left[\underbrace{\delta \exp(-\gamma g_{t+1}^c + \alpha_t y_{t+1}^2)}_{h(s_{t+1})} \left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \exp(g_{t+1}) \middle| s_t \right] \\ &= E [E [h(s_{t+1}) \exp(g_{t+1}) | \eta_{t+1}, s_t] | s_t] \end{aligned}$$

by the law of iterated expectations. Given η_{t+1} and s_t , s_{t+1} is measurable, hence we have:

$$\frac{P_t}{D_t}(s_t) = E [h(s_{t+1}) E [\exp(g_{t+1}) | \eta_{t+1}, s_t] | s_t].$$

We can solve the integral $E [\exp(g_{t+1}) | \eta_{t+1}, s_t]$ using the normality assumption:

$$E [\exp(g_{t+1}) | \eta_{t+1}, s_t] = \exp\{E[g_{t+1} | \eta_{t+1}, s_t] + \frac{1}{2} \text{Var}(g_{t+1} | \eta_{t+1}, s_{t+1})\}$$

with

$$\begin{aligned} E[g_{t+1} | \eta_{t+1}, s_t] &= \mu^d + \phi x_t + \sigma \varphi_d \rho_{\eta,u} \eta_{t+1} \\ \text{Var}(g_{t+1} | \eta_{t+1}, s_{t+1}) &= \sigma^2 \varphi_d^2 (1 - \rho_{\eta,u}^2) \end{aligned}$$

given that u_t and η_t are jointly normal. This yields

$$\frac{P_t}{D_t}(s_t) = \underbrace{\exp\{\mu^d + \phi x_t + \frac{1}{2} \sigma^2 \varphi_d^2 (1 - \rho_{\eta,u}^2)\}}_{c(s_t)} E [h(s_{t+1}) \exp(\sigma \varphi_d \rho_{\eta,u} \eta_{t+1}) | s_t],$$

so we can write

$$(A.4) \quad \frac{P}{D}(s_t) = c(s_t)E \left[\delta \exp\{-\gamma g_{t+1}^c + \alpha y_{t+1}^2 + \sigma \varphi_{d\rho\eta,u}\eta_{t+1}\} \left(1 + \frac{P}{D}(s_{t+1})\right) \right] (s_t),$$

which means that we need to integrate only over one dimension, η_{t+1} .

Denote $P/D(s_t)$ by $v(s)$, dropping the time subscript for convenience. We can write:

$$(A.5) \quad v(s) = c(s) \int K(s, s') (1 + v(s')) f(s'|s) ds'$$

$$(A.6) \quad = \int \psi(s, s') (1 + v(s')) f(s'|s) ds'$$

making the appropriate substitutions for $K(\cdot)$ and $\psi(\cdot)$. Define $\lambda(s, s') \equiv \psi(s, s')(1 + v(s'))$ and

$$I[\lambda](s) = \int \lambda(s, s') f(s'|s) ds' = \int \lambda(s, s') \frac{f(s'|s)}{\omega(s')} \omega(s') ds'$$

where $\omega(s')$ is a strictly positive weighting function. The integral can be approximated by the quadrature rule for $\omega(\cdot)$. Let s'_k and w_k , $k = 1, 2, \dots, N$, denote the abscissa and weights for an N point quadrature rule for the density $\omega(s')$. The approximation based on this rule to $I[\lambda](s)$ is

$$(A.7) \quad I_N[\lambda](s) = \sum_{k=1}^N \lambda(s'_k, s) \pi_k(s),$$

where

$$(A.8) \quad \pi_k(s) = \frac{f(s'_k|s)}{N(s)\omega(s'_k)} w_k$$

and

$$N(s) = \sum_{i=1}^N \frac{f(s'_i|s)}{\omega(s'_i)} w_i$$

so that the weights π_k sum up to unity. Estimating (A.6) at points $s_k = s'_k$ for $k = 1, \dots, N$ using $I_N[\lambda](s)$ gives

$$v_j = \sum_{k=1}^N (1 + v_k) \psi_{j,k} \pi_{j,k}, \quad j = 1, \dots, N$$

where $\psi_{j,k} = \psi(s'_j, s'_k)$, $\pi_{j,k} = \pi_{j,k}(s'_j)$ and $v_j = v(s'_j)$.

The $\{v_{j=1}^N\}$ are the solutions to the asset pricing equations if one views the law of motion of the state vector as a discrete Markov chain with range $\{s'_k\}$ and transition probabilities $\pi_{j,k} = P(s' = s'_k | s = s'_j)$. We choose the weight function $\omega(\cdot)$ to be the distribution of $(x_{t+1}, E_{t+1}y_{t+2}^2)$

conditional on the steady state values of the variables, $(x_0, y_0^2) = (0, \bar{y}^2)$, so $\omega(s') = f(s'|s_0)$.

APPENDIX B. ADJUSTING FOR REPURCHASES

We solve for the price-dividend ratio as detailed in Appendix A. Denote by P_t/D_t the observed S&P 500 price-dividend ratio, and by v_t^* the model's payout ratio. We want to compute a price-dividend ratio using model predictions so that it is comparable to P_t/D_t . Denote this model-generated price-dividend ratio as v_t , so that we can write

$$v_t = a_t v_t^*,$$

for some value of a_t . If there were no repurchases, a_t would be one all the time. But with repurchases, $a_t > 1$, i.e., the price-dividend ratio should be greater than the payout ratio. The figures discussed in section 2 assume that $a_t = 1$ for the sample 1891-1971, while for the sample 1972-2001 we calculate

$$a_t = \frac{R_t^p}{D_t} + 1,$$

where R_t^p is expenditure on repurchase of common stock. Notice that the total payout ratio is $\frac{R_t^p + D_t}{P_t}$. If R_t^p is zero, $a_t = 1$. We compute R_t^p/D_t for the period 1973-2001 using data from the column denoted $\sum_i REPO / \sum_i DIV$ in Table I of Grullon and Michaely. The assumption is that their sample is representative enough so that the same a_t is applicable to the S&P 500. The figures in the text report v so calculated.

APPENDIX C. IDIOSYNCRATIC SHOCKS AND EQUILIBRIUM

Individual consumption depends both on labor income and the return from a portfolio of assets. Labor income (I_t^i in (2.1)) is defined by

$$I_t^i = \delta_t^i C_t - D_t,$$

where δ_t^i is the individual shock to labor income. Since aggregate consumption satisfies $C_t = I_t + D_t$, an infinite number of agents is needed so that a law of large numbers can be applied to yield $\sum_i \delta_t^i = 1$ at each point in time. For idiosyncratic shocks to be relevant, they have to be non-stationary. A common feature of earlier models with uninsurable income, like Lucas (1994), Telmer (1993) is that the time series of the ratio of each consumer's labor income to aggregate labor income I_t^i/I_t is stationary. With low persistence, consumers are able to come close to the complete-market rule of complete risk sharing. Meghir and Pistaferri (2004) provide evidence of a significant martingale component in households' earning processes using data from the PSID. Further, they show

that the variance of the idiosyncratic shock is related to the business cycle. In Constantinides and Duffie, the process δ_t^i is the following martingale:

$$\delta_t^i = \exp \left\{ \sum_{s=1}^t \left(\eta_s^i y_s - \frac{y_s^2}{2} \right) \right\}$$

y_t is the cross-sectional standard deviation of consumption growth, and it depends on aggregates at t . The aggregates are determined first, then the shocks η_t^i are handed out. η_t^i is assumed to be standard normal $N(0, 1)$. Recall that for η normal $E[\exp(\eta k - (k^2/2))] = 1$, which implies that δ^i is a geometric martingale. Further, we have

$$(C.1) \quad \frac{\delta_{t+1}^i}{\delta_t^i} = \exp \left\{ \eta_{t+1}^i y_{t+1} - \frac{1}{2} y_{t+1}^2 \right\}.$$

Individual consumption growth $\frac{C_{t+1}^i}{C_t^i} = \frac{\delta_{t+1}^i}{\delta_t^i} \frac{C_{t+1}}{C_t}$, so it is a mixture distribution (y_{t+1} is a random variable with known distribution), and (C.1) can be used to find moments of individual consumption.

Constantinides and Duffie (1996) prove that there exist a unique equilibrium with no trade in this economy and the pricing Euler equation takes equals (2.4) in the body of the paper.²⁵

²⁵See the appendix in their paper for a formal proof.

REFERENCES

- [1] Ravi Bansal and Amir Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59:1491–1509, 2004.
- [2] Ben S. Bernanke. Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics*, 98:85–106, February 1983.
- [3] Michele Boldrin, Lawrence Christiano, and Jonas Fisher. Habit persistence and asset returns in an exchange economy. *Macroeconomic Dynamics*, 1(2):312–32, 1997.
- [4] J. Boudoukh, Roni Michaely, Matthew Richardson, and Michael Roberts. On the importance of measuring payout yield: Implications for empirical asset pricing. manuscript, 26 pages, 2003.
- [5] Alon Brav, George Constantinides, and Christopher Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, 110:793–824, 2002.
- [6] J. H. Campbell and R.J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1:195–227, 1988.
- [7] J. H. Campbell and R.J. Shiller. Stock prices, earnings, and expected dividends. *Journal of Finance*, 43:661–676, 1988.
- [8] J. Y. Campbell, A.W. Lo, and C. A. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, 1997.
- [9] John Campbell. Comment. In Ben Bernanke and Julio Rotemberg, editors, *NBER Macroeconomics Annual*, pages 253–262. Cambridge: MIT Press, 1999.
- [10] John Y. Campbell. Consumption-based asset pricing. In George Constantinides, Milton Harris, and Rene Stulz, editors, *Handbook of the Economics of Finance*. Norton, 2003.
- [11] John Cochrane. *Asset Pricing*. Princeton University Press, Princeton, New Jersey, 2001.
- [12] Timothy Cogley. Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey. *Journal of Monetary Economics*, 49:309–334, 2002.
- [13] Angus Deaton and Christina Paxson. Intertemporal choice and inequality. *The Journal of Political Economy*, 102(3):437–467, 1994.
- [14] Avinash K. Dixit and Robert S. Pindyck. *Investment Under Uncertainty*. Princeton University Press, Princeton, New Jersey, 1994.
- [15] Luigi Ermini. Reinterpreting a temporally aggregated consumption cap model. *Journal of Business & Economic Statistics*, 9:325–328, 1991.
- [16] Eugene F. Fama and Kenneth R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22:3–25, 1988.
- [17] Myron J. Gordon. The savings investment and valuation of a corporation. *The Review of Economics and Statistics*, 44(1):37–51, 1962.
- [18] G. Grullon and R. Michaely. Dividends, share repurchases, and the substitution hypothesis. *Journal of Finance*, 57:1649–1684, 2002.
- [19] Robert Hall. Stochastic implications of the life cycle permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, 86:971–987, Dec 1978.
- [20] Robert Hall. Intertemporal substitution in consumption. *Journal of Political Economy*, 96:339–357, 1988.

- [21] Albert G. Hart. Risk, uncertainty and the unprofitability of compounding probabilities. In O. Lange, F. McIntyre, and T.O. Yntema, editors, *Studies in Mathematical Economics and Econometrics*, pages 110–118. 1942.
- [22] Urban Jermann. Asset pricing in production economies. *Journal of Monetary Economics*, 42:167–247, 1998.
- [23] Robert Jones and Joseph Ostroy. Flexibility and uncertainty. *Review of Economic Studies*, 51(1):13–32, 1984.
- [24] Narayana R. Kocherlakota. The equity premium: It’s still a puzzle. *Journal of Economic Literature*, 34:42–71, 1996.
- [25] Tom Krebs. Welfare cost of business cycles when market are incomplete. Brown University, manuscript, 28 pages, 2004.
- [26] D. M. Kreps and E. L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46:185–200, 1978.
- [27] D. M. Kreps and E. L. Porteus. Temporal von neumann-morgenstern and induced preferences. *Journal of Economic Theory*, 20:81–109, 1979.
- [28] Deborah J. Lucas. Asset pricing with undiversifiable risk and short sales constraints: Deepening the equity premium puzzle. *Journal of Monetary Economics*, 34:325–341, 1994.
- [29] Robert E. Jr. Lucas. Macroeconomic priorities. *American Economic Review*, 93(1):1–14, March 2003. AEA Presidential Address.
- [30] Costas Meghir and Luigi Pistaferri. Income variance dynamics and heterogeneity. *Econometrica*, 72(1):1–32, 2004.
- [31] R. Mehra and E. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15:145–162, 1985.
- [32] Jonathan Parker and Christian Julliard. Consumption risk and the cross section of stock returns. *Journal of Political Economy*, 113(1):185–222, February 2005.
- [33] N. G. Shepard and A. Harvey. On the probability of estimating a deterministic component in the local level model. *Journal of Time Series Analysis*, 11:339–347, 1990.
- [34] Jeremy J. Siegel. The shrinking equity premium. *Journal of Portfolio Management*, 26:10–17, 1999.
- [35] K. Storesletten, C.I. Telmer, and A. Yaron. Cyclical dynamics in idiosyncratic labor market risk. *Journal of Political Economy*, 112:695–717, 2004.
- [36] Chris Telmer. Asset-pricing puzzles and incomplete markets. *Journal of Finance*, 48:1803–32, December 1993.
- [37] Annette Vissing-Jørgensen. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853, August 2002.
- [38] Eugene N. White. Bubbles and busts: the 1990s in the mirror of the 1920s. Duke and the University of North Carolina Conference: Understanding the 1990s, the Long-Run Perspective, March 2004.
- [39] Motohiro Yogo. Estimating the elasticity of intertemporal substitution when instruments are weak. *The Review of Economics and Statistics*, 86(3):797–810, August 2004.