CHAPTER 2

Risk-Based Explanations of the Equity Premium

John Donaldson
Columbia University

and

Rajnish Mehra
University of California, Santa Barbara, and NBER

We are grateful to Francisco Azeredo, Ravi Bansal, George Constantinides, Vito Gala, Alok Khare, and Edward Prescott for their comments on earlier drafts of this chapter.

Abstract

This essay reviews the family of models that seek to provide aggregate risk-based explanations for the empirically observed equity premium. Theories based on non-expected utility preference structures, limited financial market participation, model uncertainty, and the small probability of enormous losses are detailed. We impose the additional requirements that candidate models yield consistent intertemporal portfolio choice and that a representative agent can be constructed that is independent of the underlying heterogeneous economy's initial wealth distribution. While many models are able to replicate a wide variety of financial statistics including the premium, few satisfy these latter criteria as well.

JEL Classification: D50, D52, E30, E32, G00, G11, G12, D90, D91, D10, D11

Keywords: equity premium puzzle, habit persistence, borrowing constraints, limited participation, incomplete markets, idiosyncratic income shocks, life-cycle savings, non-expected utility preferences, prospect theory, rank-dependent, expected utility, disaster events, survivorship bias

INTRODUCTION

Mehra and Prescott (1985) demonstrated that the empirical equity premium (the return earned on a well-diversified "market portfolio" in excess of that earned by a risk-free T-bill) is an order of magnitude greater than can be rationalized *as a premium for bearing non-diversifiable risk* in the context of the standard neoclassical representative agent paradigms of financial economics.[1] Inevitably, the subsequent literature has principally focused on modifying the standard paradigm to generate much larger non-diversifiable premia.[2] As a companion exercise, these efforts have adopted an ancillary goal of replicating the risk-free rate, rate of return volatilities, return autocorrelations, return cross correlations, and return correlations with consumption growth. In this chapter, we review this literature, which falls into four broad categories:

1. Preference-based theories: These theories seek to modify an agent's preferences so that he is acutely sensitive in some way to poor consumption outcomes. Outcomes may be "poor" in either an absolute sense or in a relative sense vis-à-vis some external benchmark, as in the habit formation literature. Illustrations sometimes require that agent preferences be defined over objects directly reflective of the agent's portfolio experience, such as his recent gains or losses. While this latter feature is outside the standard expected utility construct, axiomatic foundations are nevertheless sometimes available. Section 1 of this essay focuses on these theories.[3]
2. Disaster scenarios: Researchers have also explored models where utility is of the standard CRRA variety, yet in which the worst outcomes are much more disastrous than the worst cases possible in the standard Mehra and Prescott (1985) context. Rietz (1988) and Barro (2006), in particular, fall into this category.[4] Literatures 1 and 2 are dual to one another: either agents in the model must effectively be very sensitive to bad outcomes, or it is the outcomes themselves that must be very bad.
3. Trading frictions: Under either of the categories above, the premium will still be counterfactually small if agents can somehow insure against their worst consumption outcomes. Since the standard Mehra and Prescott (1985) representative agent construct is implicitly one of complete markets, perfect insurance opportunities are available in that context. Furthermore, for models parameterized to the per capita aggregate consumption process of the U.S. economy, agents do not find themselves in situations of extreme consumption risk. There must then

[1] The following chapter (Mehra and Prescott (2008b)) surveys the literature on non-risk-based explanations of the equity premium.

[2] Mehra and Prescott (1985) can match the premium with a very high CRRA. Alternatively, we can view the literature as seeking to match the premium (and other financial quantities) with a low CRRA.

[3] For example, Abel (1990), Bansal and Yaron (2004), Benartzi and Thaler (1995), Boldrin, Christiano, and Fisher (2001), Campbell and Cochrane (1999), Constantinides (1990), Epstein and Zin (1991), and Ferson and Constantinides (1991).

[4] See also Mehra and Prescott (1988) and Brown, Goetzman, and Ross (1995).

be added features to the model, which deny at least some of the agents with risky consumption profiles access to a full set of state claims. Incomplete markets (whether in the form of restricted financial market participation for some agents or in the form of allowing trade in only a restricted set of assets) and collateral constraints are possibilities that have been explored.[5,6,7]

4. Model uncertainty: Nearly all the asset pricing literature presumes rational expectations; the relevant agents in the economy know the true probability distributions governing the relevant state variables: dividends, wages, etc. The model uncertainty literature presumes that this is not the case; it thereby introduces another source of uncertainty, a feature that gives rise to an additional risk premium for equity securities. Barrillas et al. (2006) and Weitzman (2007) are cases in point.

We do not address this entire literature in full detail, as many of the topics are dealt with in specific chapters of the present volume. Category 3, in particular, is thoroughly considered in the essays by Constantinides and by Heaton and Lucas, particularly as regards classical incompleteness (too few securities are traded).[8] We thus consider only the impact of borrowing constraints. Loss aversion and narrow framing are carefully considered in the Barberis and Huang chapter.[9] As it constitutes a preference-related modification, we consider it in the present chapter but in an abbreviated form. As a field of study, model uncertainty as applied to the premium is in its infancy. While potentially very fruitful, the literature is small and our discussion will be correspondingly brief. Most research has followed the preference route, and this is the literature that we will emphasize.[10] Whenever possible, we cast the discussion within the original Mehra and Prescott (1985) discrete-time representative agent setting.

[5]For incomplete market formulations, see Bewley (1982), Brav, Constantinides and Geczy (2002), Constantinides and Duffie (1996), Heaton and Lucas (1997, 2000), Krebs (2000), Lucas (1994), Lustig and van Nieuwerburgh (2005), Mankiw (1986), Mehra and Prescott (1985), Storesletten, Telmer, and Yaron (2007), and Telmer (1993).

[6]For restricted participation and collateral constraints, see Aiyagari and Gertler (1991), Alvarez and Jerman (2000), Bansal and Coleman (1996), Basak and Cuoco (1998), Constantinides, Donaldson, and Mehra (2002), Daniel and Marshall (1997), Danthine, Donaldson, and Mehra (1992), He and Modest (1995), Heaton and Lucas (1996), and Luttmer (1996), McGrattan and Prescott (2001), and Storesletten, Telmer, and Yaron (2004).

[7]Attanasio, Banks, and Tanner (2002), Brav, Constantinides, and Geczy (2002), Brav and Geczy (1995), Mankiw and Zeldes (1991), and Vissing-Jorgensen (2002) are models that also incorporate restricted participation.

[8]Constantinides (2008) and Heaton and Lucas (2008).

[9]Barberis and Huang (2008).

[10]Gabaix and Laibson (2001) and Heaton (1995). The reader is also referred to the excellent surveys by Kocherlakota (1996), Cochrane (1997), Campbell (2001, 2003), and DeLong and Magin (2007), preliminary draft. These papers review the proposed explanations for the magnitude of the premium, and the associated models. For an up-to-date international perspective on the empirical magnitude of the premium, see Dimson, Marsh, and Staunton (2008). Goetzmann and Ibbotson (2008) review the conceptual issues in its measurement over the long term.

1. ALTERNATIVE PREFERENCE STRUCTURES

1.1. Preliminaries

The focus of this section is to assess the extent to which plausible preference structures alone can rationalize the basic stylized facts as they relate to asset returns and the equity premium. The analysis in the previous chapter shows that the isoelastic CRRA preferences used in Mehra and Prescott (1985) can be made consistent with the observed equity premium only if the coefficient of relative risk aversion is implausibly large.

Before proceeding with our discussion of alternative preference structures, we examine in detail some especially attractive features of the CRRA class, which makes it the "preference function of choice" in modern finance and macroeconomics. These features, while desirable in their own right for logical or empirical reasons, also constitute common properties that any reasonable preference specification should display.

(i) Equilibrium Return Stationarity

CRRA preferences result in a stationary equilibrium return process. This property is readily demonstrated in the context of the Mehra and Prescott (1985) model, where both the equity return and the return on the risk-free asset follow stationary processes, despite the fact that the level of output in the economy is growing over time and hence is itself non-stationary.[11] This is consistent with the statistical evidence on the time series of asset returns over the past 100 years. Asset returns appear to be stationary, although the level of stock prices and the magnitude of aggregate dividends have grown enormously. Any serious preference structure should yield equilibrium return series with this feature.

(ii) Aggregation

Actual asset prices are formed via the trading behavior of large numbers of heterogeneous investors as each attempts to maintain his individual optimal portfolio composition, given his information on the future distribution of returns. Equilibria in such economies are difficult to characterize.

If financial markets are competitive and complete, and agent preferences are expected utility, there will, in general, exist, by construction, a representative (single-agent) economy with the same aggregate consumption series as the heterogeneous agent economy and the same asset price functions. These economies are comparatively easy to analyze. In addition, if the representative agent can be constructed in a manner that is independent of the underlying heterogeneous agent economy's initial wealth distribution, we say the economy displays aggregation.

Aggregation (vis-à-vis the existence of a representative agent) is the stronger, more restrictive, and more desirable property. It implies that assets may be priced in the representative agent economy without knowledge of the wealth distribution in the

[11] See Eqs. (8)–(11) in Mehra and Prescott (1985) or Appendix B of Chapter 1 of this volume.

underlying heterogeneous agent counterpart. Under aggregation, results derived in the representative agent economy are general and robust. Aggregation also permits the use of the representative agent for welfare comparisons.[12] In what follows, we provide a brief description of how the representative agent is constructed and then point out the additional structure imposed by aggregation.

Consider an exchange economy *of* $k = 1, 2, \ldots K$ agents. We make the following notational identifications: c_t^k represents agent k's period t consumption, while q_t^e and q_t^b, respectively, denote the period t prices of the equity and one period risk-free security. The period t price of a state claim to one unit of consumption in period $t + j$, if state θ occurs, is denoted by $q_{t,j}^{\theta}$. By analogy, the corresponding period t desired holdings of these securities by agent k are, respectively, $z_t^{e,k}$, $z_t^{b,k}$, and $z_{t,j}^{\theta,k}$. For notational simplicity, we have suppressed the dependence of these quantities and prices on the period t state. With this notation in mind, the period t decision problem faced by an arbitrary agent k is

$$\max E \left\{ \sum_{t=0}^{\infty} \beta^t u\left(c_t^k\right) \right\}$$

subject to

$$c_t^k + z_{t+1}^{e,k} q_t^e + z_{t+1}^{b,k} q_t^b + \sum_{j=1}^{\infty} \sum_{\theta} z_{t+1,j}^{\theta,k} q_{t,j}^{\theta} \leq z_t^{e,k}\left(q_t^e + Y_t\right) + z_t^{b,k} + z_t^{\theta,k} + \sum_{j=1}^{\infty} \sum_{\theta} z_{t,j}^{\theta,k} q_{t,j}^{\theta}.$$

In equilibrium, both goods and asset markets must clear:

$$\sum_{k=1}^{K} c_t^k = Y_t,$$

$$\sum_{k=1}^{K} z_t^{e,k} = 1,$$

$$\sum_{k=1}^{K} z_t^{b,k} = 0,^{13}$$

$$\sum_{k=1}^{K} z_{t,j}^{\theta,k} = 0 \text{ for all } j, \theta.$$

Under competitive market conditions, equilibrium is Pareto optimal. This means the equilibrium consumption allocations above are also the solution to

[12]Suppose, alternatively, that aggregation fails. Then each possible initial wealth distribution, via its associated representative agent, will display its own asset pricing characteristics. No general statements may be made.

[13]The assumption that bonds are in zero net supply is not innocuous, although it is customary in the literature. In particular, it allows representative agent models to match the risk-free rate more easily. For a recent perspective see Gomes and Michaelides (2006).

$$\max E \sum_{t=0}^{\infty} \sum_{k=1}^{K} \lambda^k u\left(c_t^k\right)$$

subject to

$$\sum_{k=1}^{K} c_t^k \leq Y_t,$$

where the weights $\left\{\lambda^1, \ldots, \lambda^K\right\}$ in general directly depend on the initial endowment distributions $\left\{z_0^{e,1}, \ldots, z_0^{e,K}\right\}$, $\left\{z_0^{b,1}, \ldots, z_0^{b,K}\right\}$ and $\left\{\left\{z_{0,j}^{\theta,k}\right\}_{\theta,j}, k = 1, \ldots, K\right\}$.

For the given $\left\{\lambda^1, \ldots, \lambda^K\right\}$ we define the utility function of the representative agent as $u(C_t) = \sum_{k=1}^{K} \lambda^k u\left(c_t^k\right)$, where $C_t = Y_t = \sum_{k=1}^{K} c_t^k$.[14] These weights correspond to the inverses of the Lagrange multipliers on the consumers' decentralized problems. Roughly speaking, consumers with greater endowments are accorded higher λs. The equilibrium stock and bond prices in the heterogeneous agent economy are then identical to those arising in a homogeneous agent economy where a "representative agent," constructed as above, maximizes[15]

$$E_0 \left\{ \sum_{t=0}^{\infty} \beta^t u\left(C_t\right) \right\}, \quad 0 < \beta < 1,$$

subject to

$$C_t + z_{t+1}^e q_t^e + z_{t+1}^b q_t^b \leq \left(q_t^e + Y_t\right) z_t^e + z_t^b,$$

where, in equilibrium, $z_t^e = 1$ and $z_t^b = 0$ $\forall t$. Under this equivalence, the representative agent is a stand-in for the weighted average of the economy's members.

As we have noted, the weights and thus the asset price series itself will, in general, depend on the initial wealth distribution. This is an enormously complicating feature as it suggests that asset prices in any period are in significant ways dependent on wealth distributions many years previously. If all agents in the decentralized multi-agent economy have preferences of the CRRA type, however, the weights $\left\{\lambda^1, \ldots, \lambda^K\right\}$ will be independent of the initial wealth distribution, a fact that makes the use of representative agent models for asset pricing exercises more plausible. More formally, aggregation requires that the period utility function must be of the form $u\left(c_t^k\right) = e^{-\alpha c_t^k}$ or $u\left(c_t^k\right) = (\gamma/\gamma - 1)\left(\alpha^k + \gamma c_t^k\right)^{1-1/\gamma}$, where the α^k are parameters unique to agent k. To emphasize this point again, even if the individual preferences assume neither of these forms (but are still expected utility), there will continue to

[14]We use capital letters to denote economy wide aggregates. For representative agent economies, individual and aggregate quantities coincide, so we use the aggregate notation in those contexts. In situations of clear agent heterogeneity, for example in OLG models, we denote individual agent quantities by lower case variables.

[15]See Prescott and Mehra (1980) for an early articulation of this and Constantinides (1982) for a formal analysis of this decentralization perspective.

exist a representative agent whose preferences are the weighted average of the individual agent preferences. However, they cannot be used to make any statements about "off-equilibrium" paths.[16]

(iii) Time-Consistent Planning

Time consistency implies that the optimal future-contingent portfolio decisions made at $t = 0$ remain the optimal decisions even as uncertainty resolves and intermediate consumption is experienced.

When considering multi-period decision problems, time consistency is a natural property to propose. In its absence, one would observe portfolio rebalancing not motivated by any event or information flow but rather simply motivated by the (unobservable) changes in the investor's preference ordering as time passed. Asset trades would be motivated by endogenous and unobservable preference characteristics and would thus be mysterious and unexplainable.

To understand the utility restrictions that time consistency imposes, let us consider a two-period context, $t = 0, 1$, where any one of $s \in S$ possible states may occur the next period. Denote by $c_1(s)$ the investor's consumption at date $t = 1$ contingent on state "s" occurring. Johnsen and Donaldson (1985) demonstrate that if the overall (both periods' consumption) utility function is to exhibit time-consistent planning, there must exist continuous and monotone increasing functions $f(\,)$ and $\{u_s\ (\,) : s \in S\}$ such that

$$u\big(c_0, c_1(s) : s \in S\big) = f\big(c_0, u_s\big(c_0, c_1(s)\big) : s \in S\big),$$

where $u_s(c_0, c_1(\mathrm{s}))$ is the state s contingent utility function. The recursive preference structure of Kreps and Porteus (1978), which makes explicit the preference for the timing of uncertainty resolution, is closely related. Note that under this formulation, preferences in any future state $\hat{s}$ may be conditioned on consumption today c_0, but not on consumption planned for states $s' \neq \hat{s}$ that, de facto, have not occurred. In effect, the utility function must be of a form such that utility representations in future states can be recursively nested as individual arguments of the overall utility function, a condition fulfilled by the expected utility representation. We are thus, assuredly, in a time-consistent planning context under the expected utility assumption.[17]

1.2. Coincidence of Risk and Time Preferences in CRRA utility

One restriction imposed by the CRRA class of preferences is that the coefficient of risk aversion is rigidly linked to the elasticity of intertemporal substitution. One is the reciprocal of the other. If an individual is averse to variation of consumption across different states *at a particular point of time,* then this feature implies he will

[16]Within the context of seeking to replicate the equity premium, the statement is intended to remind the reader that, without aggregation, the asset pricing properties may become unique to the particular underlying initial wealth distribution. As a result, the same asset pricing results may not apply across various initial wealth distributions and associated representative agents. No statements of any generality can be made.

[17]Time-consistent planning is satisfied in the Mehra and Prescott (1985) setup, because they assume an expected utility preference representation, in addition to the CRRA period utility specification.

be averse to *consumption variation over time as well*. Since, on average, consumption is growing over time, the agents in the Mehra and Prescott (1985) setup have little incentive to save. The demand for bonds is low, and as a consequence the risk-free rate is counterfactually high.

We illustrate this in the context of a deterministic two-period model where agents have preferences of the form

$$u(c,\gamma) = \frac{c^{1-\gamma}}{1-\gamma}$$

and we solve the following problem:

$$\max\ u(c_0,\ \gamma) + \beta\, u(c_1,\ \gamma)$$

subject to

$$c_0 + s_0 \leq Y_0,$$

$$c_1 \leq s_0(1+r).$$

The solution is

$$\left(\frac{1}{\beta(1+r)}\right)^{\frac{1}{\gamma}} = \left(\frac{Y_0 - s_0}{(1+r)s_0}\right) = \left(\frac{c_0}{c_1}\right).$$

As $\gamma \mapsto \infty$ (greater CRRA), $\left(c_0/c_1\right) \mapsto 1$, the agent wishes for an extremely smooth consumption profile across time, a preference that is maintained under uncertainty. In a setting without uncertainty, this effectively means that the agent dislikes growth in his intertemporal consumption profile. In an asset pricing environment in which the representative agent's consumption is growing exogenously through time (in order that it be properly calibrated), the agent will thus demand a very high rate of return in order to hold securities, the very possession of which increases his intertemporal consumption discrepancy. From the perspective of intertemporal consumption smoothing, the parameter $1/\gamma$ is referred to as the elasticity of intertemporal substitution (hereafter EIS). While indeed increasing the premium, increasing the CRRA has the counterfactual consequence of increasing mean returns to a level much in excess of what is observed, especially in the case of the risk-free security.

In order to confirm these effects, let us explore the mean returns and volatilities for a number of CRRAs in the context of the basic Mehra and Prescott (1985) model and calibration.[18]

From Table 1 there is no doubt that increasing risk aversion increases the equilibrium equity premium: as the agent increasingly dislikes consumption variation of any sort, he must be paid an increasing premium in order to be willing to hold risky securities in preference to risk-free ones. It also improves the volatilities of the respective

[18]This model is described in greater detail in Appendix B to Chapter 1 (Mehra and Prescott (2008a)).

TABLE 1[i]
The Mehra and Prescott (1985) Model: Various CRRAs (Rates of return annualized in percent)

	U.S. Data[ii]	$\gamma = 2$	$\gamma = 3$	$\gamma = 5$	$\gamma = 9$
Er^e	6.98	7.84	9.58	12.83	18.25
SDr^e	16.54	4.34	4.99	6.36	9.24
Er^f	0.80	7.56	9.10	11.85	15.93
SDr^f	5.67	1.06	1.61	2.73	4.99
Er^p	6.18	0.28	0.48	0.97	2.31
SDr^p	16.67	4.20	4.70	5.69	7.61

[i] Same calibration as Mehra and Prescott (1985), but for various γ.
[ii] Data for the period 1889–1978.

return series. But the cost in other counterfactuals is very great. In particular, even as the CRRA is not yet absurdly large ($\gamma = 9$), the mean equity and risk-free returns become much too high, at, respectively, 18.25 and 15.93 percent annualized. This is a direct consequence of the single-preference parameter construct: in order to induce the representative agent to hold either type of security—an action that reduces the extent of his intertemporal consumption smoothing—security prices must fall to very low levels. Equivalently, the returns on all types of securities must be very high.[19]

There is no a priori reason that the desire to smooth consumption across states within a period and the desire to smooth consumption intertemporally should be tightly bound together. After all, one may speak of intertemporal consumption smoothing in a no-risk environment (in a context, say, of motivating savings behavior), and one may as well speak of risk aversion in an uncertain, atemporal environment. The situation considered previously requires that the same parameter describe the extent of sensitivity to both of these variations. Empirical studies seem to suggest, in fact, that investors are more desirous of a smooth intertemporal consumption stream than they are of an atemporal risk-free environment.

1.3. Separating Risk and Time Preferences: Epstein–Zin and others

Epstein and Zin (1989, 1991, 2001) and Kreps and Porteus (1978) have emphasized a class of preferences that they term "generalized expected utility" (GEU), which

[19] Mehra and Prescott's (1985) calibration admits a small degree of negative autocorrelation in the growth rate of consumption. Azeredo (2007) argues that the Mehra and Prescott (1985) consumption autocorrelation estimates may be due to consumption data mismeasurement prior to 1929 and that mild positive autocorrelation is more accurate. If the stochastic process governing the growth rate of consumption is positively serial correlated, then as the risk aversion increases progressively, the equity premium eventually declines and turns negative. The earlier assertion that a higher CRRA leads to a larger premium is therefore consumption process-specific. The consumption process underlying most tables and model evaluations in this survey, however, is that of Mehra and Prescott (1985) in order to benchmark a basis of comparison.

allows independent specification of the coefficient of risk aversion and the elasticity of intertemporal substitution.[20] These preferences separate time and risk preferences in a way that preserves the time-consistency property of the previous section.[21]

The basic notions involved in separating time and risk preferences are roughly summarized as follows: first characterize preferences over riskless consumption sequences $(C_0, C_1, C_2, \ldots)$ with a Koopmans' (1960) time aggregation $V(\,)$ where

$$U({}_tC) = V(u(C_t), U({}_{t+1}C)).^{22,23} \tag{1}$$

In the above expression, $U(\,)$ is the overall utility representation, C_t is the period t consumption, $u(\,)$ is the period utility function, and $({}_tC)$ denotes a continuous consumption sequence $(C_t, C_{t+1} \ldots)$.

This framework is then generalized to evaluate uncertain consumption sequences essentially by replacing the second argument in $V(,)$ by the period t certainty equivalent of the probability distribution over all possible consumption continuations. The resultant class of recursive preferences may be notationally characterized as

$$U_t = V(u(C_t), \mu_t(U_{t+1})), \tag{2}$$

where $\mu_t(\,)$ describes the certainty equivalent function based on the conditional probability distribution over consumption sequences beginning in period $t+1$. Such preferences are dynamically consistent. If the certainty equivalent is obtained via expected utility, the preferences fall into the Kreps and Porteus (1978) family. If it is obtained via a more general risk aggregator of the Chew–Deckel class (e.g., disappointment aversion risk preferences to be considered shortly), the preferences are said to be of the more general Epstein and Zin (1989) variety. These preferences have axiomatic underpinnings as well.

[20]This preference construct has as its intellectual antecedents the time-consistent recursive preference structures of Kreps and Porteus (1978) and Johnsen and Donaldson (1985), and the Chew (1983, 1989) and Dekel (1986) preference generalization, which allows the independence assumption of classical expected utility to be relaxed. The latter is important for at least two reasons. First, there are widely reported violations of independence in the experimental literature; the Allais (1979) paradox is perhaps the most celebrated example. Second, this generalization has permitted the creation of preference structures that allow greater utility weight on bad outcomes than would arise in an expected utility context; see, in particular, Gul (1991). Not only is there empirical support for the hypothesis of high sensitivity to bad outcomes, but it is also a plausible feature that allows consumption-based asset pricing models to better replicate the premium. The reader is directed to Backus et al. (2004) for an elegant and detailed discussion of this preference class.

[21]More recently, Kihlstrom (2007) proposes an alternative dynamic choice model.

[22]We use the same notation as Backus et al. (2004) and Johnsen and Donaldson (1985). See the former study for a careful discussion of all the issues involved.

[23]Preferences over deterministic consumption sequences that have the form (1) satisfy three reasonable properties (by construction): (i) history independence (preferences over consumption sequences $(C_0, C_1, C_2, \ldots)$ do not depend on prior consumption); (ii) future independence (preferences for period t consumption, C_t) are independent of future consumption $({}_{t+1}C)$; and (iii) stationarity (preferences are the same at all dates).

In their most basic analysis, Epstein and Zin explore the CES-like specialized preference ordering that is detailed below:

$$u_t \equiv u(C_t, CE_{t+1}) = \left\{(1-\beta)C_t^{1-\frac{1}{\rho}} + \beta(CE_{t+1})^{1-\frac{1}{\rho}}\right\}^{1-\frac{1}{\rho}} \tag{3}$$

with $0 < \beta < 1, 1 \neq \gamma > 0,\ \rho > 0$, or, in the case of $\rho = 1$,

$$u(C_t, CE_{t+1}) = (1-\beta)\ \log C_t + \beta\ \log(CE_{t+1}) \tag{4}$$

where $CE_{t+1} = CE(\tilde{u}_{t+1})$, the certainty equivalent of next period's utility, is calculated according to

$$[CE\ (\tilde{u}_{t+1})]^{1-\gamma} = E_t\left\{(\tilde{u}_{t+1})^{1-\gamma}\right\},\ 1 \neq \gamma > 0, \text{or} \tag{5}$$

$$\log(CE\ (\tilde{u}_{t+1})) = E_t(\log \tilde{u}_{t+1}),\ \gamma = 1. \tag{6}$$

In the above specification, the elasticity of intertemporal substitution ρ may be specified independently of the CRRA γ. Note that if $\gamma = 1/\rho$, recursive substitution to eliminate u_{t+1} yields

$$u_t = \left[(1-\beta)\ E_t \sum_{j=0}^{\infty} \beta^j\ C_{t+j}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}, \tag{7}$$

which represents the same preference ordering over uncertain consumptions streams as the ordering given by

$$E_t\left(\sum_{j=0}^{\infty} \beta^j\ C_{t+j}^{1-\gamma}\right),$$

a fact that suggests we may unambiguously view the parameter γ of (3) and (4) as the CRRA (Epstein and Zin (1989) provide a more formal argument). Similarly, as the parameter ρ becomes smaller, the agent becomes less willing to substitute utility, and thus consumption, intertemporally. The parameter β can be viewed as the agent's customary time preference parameter.

Weil (1989) uses these preferences in a setting otherwise identical to that of Mehra and Prescott (1985) to compute the resultant premium. Even as specialized to the Epstein–Zin (1989) preference structure, Weil (1989) demonstrates that the Mehra and Prescott (1985) solution for the equilibrium price functions, $q^e\left(Y_t, x^i\right) = \nu_i Y_t$ and $q^b\left(Y_t, x^i\right) = \omega_i$, still applies with the proviso that the constants ν_i, ω_i now must satisfy a system of non-linear equations. Using an endowment process $\{Y_t\}$ otherwise identical to that of Mehra and Prescott (1985), Weil (1989) obtains the results detailed in Table 2.

TABLE 2[i]
Equilibrium Mean Rates of Return Under Epstein–Zin Preferences: Various Parameters/Selections (Rates of return annualized in percent. The first entry is the premium; the second is the risk-free rate.)

EIS	CRRA(γ)						
(ρ)	0	0.5	1	5	10	20	45
∞	0.00	0.05	0.10	0.48	0.94	1.77	3.01
	5.25	5.24	5.21	5.01	4.78	4.40	4.09
2	0.01	0.06	0.11	0.51	1.01	1.89	3.14
	6.20	6.16	6.12	5.79	5.40	4.73	3.93
1	0.01	0.07	0.12	0.55	1.08	2.00	3.27
	7.14	7.08	7.03	6.56	6.02	5.06	3.76
0.2	0.10	0.18	0.26	0.88	1.64	2.91	4.34
	15.02	14.81	14.61	13.02	11.11	7.75	2.45
0.1	0.24	0.35	0.45	1.31	2.33	4.04	5.72
	25.73	25.32	24.96	21.68	17.87	11.23	0.85
0.05	0.56	0.72	0.87	2.12	3.66	6.25	8.66
	50.51	49.55	48.61	41.26	32.80	18.65	−2.23
1/45	1.13	1.36	1.60	3.58	6.22	11.22	17.11
	138.91	135.56	132.25	107.44	80.69	40.39	−9.22

[i] This is Table 1 of Weil (1989).

It is readily apparent that the Epstein–Zin preference construct can, per se, provide some progress toward the equity premium puzzle resolution, but not a complete resolution. In particular, if the CRRA is set at 45 and ρ at 0.1, then a reasonable match to the data (mean values) is obtained. But if a much more reasonable CRRA of 1 is hypothesized, the premium is a scant 0.45 percent, while the mean risk-free rate is 25 percent.[24] Both of these features are counterfactual. As Weil (1989) argues, generalizing preferences in the direction of Epstein and Zin (1989) seems only to deepen the puzzle: (1) in abstract economies such as Mehra and Prescott's (1985), calibrated to observed per capita income dynamics ($C_t = Y_t$), there is insufficient consumption

[24] Consistent with these findings yet using a very different methodological perspective, Kocherlakota (1990) demonstrates that when the growth rate of consumption is i.i.d. (an assumption with substantial empirical support), asset pricing models that separate the CRRA and the EIS as per (1)–(4) have no more explanatory power in empirical tests than those preferences for which these parameters coincide as in the standard Mehra and Prescott (1985) paradigm.

growth risk to justify a substantial premium unless agents are implausibly risk-averse. (2) Furthermore, per capita consumption growth is too high ($\mu = 1.8$ percent) for the equilibrium risk-free rate to be low, when agents are averse to intertemporal consumption substitution to the extent observed. In hindsight, these results are not entirely surprising. Even under Epstein–Zin's (1989, 1991) preferences, there is one parameter alone, γ, which is given the entire role of characterizing the degree of risk aversion. Furthermore, by separating time and risk preferences, their structure tends to strengthen the influence of the intertemporal substitution parameter $1/\rho$. As a result, agents must once again be offered a counterfactually high risk-free rate in order to induce them to save more in an environment when mean consumption growth is exogenously fixed at 1.8 percent. Note that Weil's findings do not disqualify the entire Epstein–Zin family since other approaches to obtaining certainty equivalents may apply. They do suggest, however, that the difficulty may be in separating time and risk preferences where the agent does not need to be too risk-averse to match the data.

Epstein–Zin preference constructs have been employed more successfully by Bansal and Yaron (2004) in a richer economic environment. These authors postulate processes on consumption growth and the dividend which are distinct but share a small, but highly persistent time varying expected growth component. The resultant 'long run consumption risk' is specially onerous for Epstein–Zin style representative agents who desire an early resolution of uncertainty.[25] This latter feature demands a preference specification where the CRRA > (1/EIS), a precondition that the Bansal and Yaron (2004) calibration easily satisfies. For the $\gamma = 10$ and $\rho = 1.5$ case, they are able to replicate all the stylized facts quite well: $Er^e = 6.84\%$, $Er^f = 0.93\%$, $\sigma_{r^e} = 18.65\%$ and $\sigma_{r_r} = 0.57\%$. (Table 4 in Bansal and Yaron (2004)).[26] That the persistent growth component is small lends credibility to their basic consumption growth hypothesis, as it is very difficult, empirically, to distinguish this possibility from the more customary i.i.d. consumption growth assumption. While an EIS as high as 1.5 is disputable, there is no prevailing consensus estimate of this quantity, even as regards to it being greater, equal to, or less than one. (See Attanasio and Vissing-Jorgensen (2007), Hall (1988), Guvenen (2006) and Yogo (2004).

[25]See also the chapter by Ravi Bansal in this volume (Bansal (2008)).

[26]Why the presence of long run consumption risk should be an especially important factor in the Epstein and Zin preference context is more apparent if we examine the corresponding process on the marginal rate of substitution,

$$M_{t+1} = \frac{U_1(C_{t+1})}{U_1(C_t)} : \ln M_{t+1} = \theta \ln \beta + \frac{\theta}{\rho} \ln\left(\frac{C_{t+1}}{C_t}\right) - (1-\theta)\, r^e_{t+1}$$

where $\theta = \dfrac{1-\gamma}{1-1/\rho}$, and $r^e_{t+1} = \ln\left[\dfrac{W_{t+1}+C_{t+1}}{W_t}\right]$ is the period $t+1$ return on the agent's wealth portfolio, W_t. If $\gamma = \frac{1}{\rho}$, the standard power utility case, the expression collapses whereby only consumption growth matters for the MRS variation. If $\gamma \neq 1/\rho$, the wealth component adds another source of risk. Long run persistence in dividend growth affects this latter term.

We view Bansal and Yaron (2004), its predecessors and successors, as important for a number of reasons.[27] First and foremost, it emphasizes the importance of long run growth variation for the asset pricing literature. Their paper also suggests that a full resolution of the equity premium puzzle may demand an improved understanding not only of investor preference structures, but also of the long run statistical characteristics of the principal aggregate macro series.[28]

More recently, Kallenbrunner and Lochstoer (2006) have demonstrated that the consumption growth process postulated by Bansal and Yaron (2004) will arise endogenously in a fairly standard production style real business cycle model, purely as a result of the consumption smoothing efforts of the representative agent armed with the same Epstein-Zin preference specification. While the authors match a fairly limited set of business cycle facts, their study is important not only for lending further credibility to the Bansal and Yaron (2004) perspective but also for the critical issue of cross model verification.

Although standard Epstein and Zin preferences do not necessarily go very far in resolving the equity premium puzzle, it is nevertheless informative to study the full implication of these preferences for the properties of security returns beyond the equity premium. Within the standard Mehra-Prescott (1985) context, Epstein and Zin (1989, 1991) develop an elaborate dynamic programming argument to derive the following asset pricing equation.

$$E_t\left\{\left[\beta\left(\frac{\tilde{C}_{t+1}}{C_t}\right)^{-\frac{1}{\rho}}\right]^{\theta}\left[\frac{1}{1+\tilde{r}^{e}_{t+1}}\right]^{1-\theta}\left(1+\tilde{r}^{j}_{t+1}\right)\right\}\equiv 1, \tag{8}$$

where $\tilde{r}^e_{t+1}$ denotes the period t return on the agent's wealth portfolio and r^j_{t+1} the period t return on some asset j included within it.[29] In the above expression, the pricing kernel is of the form

$$\left[\beta\left(\frac{\tilde{C}_{t+1}}{C_t}\right)^{-\frac{1}{\rho}}\right]^{\theta}\left[\frac{1}{1+\tilde{r}^{e}_{t+1}}\right]^{1-\theta}$$

where, $\theta = (1-\gamma)/(1-1/\rho)$.

[27]We include in the list of predecessors Backus and Zin (1994) and Cochrane and Hansen (1992). Other implementations of the Epstein–Zin preference construct are Epstein and Zin (1989), Yogo (2006) and Malloy et al. (2005).

[28]Bakshi and Chen, (2008) modify the standard model by identifying dividend processes as a fixed fraction of earnings plus an i.i.d error. Potential errors in consumption measurement are there by avoided. They also assume risk free interest rates follow a mean reverting process with constant volatility. As with Bansal & Yaron (2004) they derive closed form solutions and match the premium.

[29]Cochrane derives this relationship in the Appendix to his essay in this volume (2008).

This is a geometric average (with weights θ and $1-\theta$, respectively) of the kernel of the standard CCAPM,

$$\left[\beta\left(\frac{\tilde{C}_{t+1}}{C_t}\right)^{-\frac{1}{\rho}}\right],$$

and the reciprocal of the gross return on the wealth portfolio.[30] Epstein and Zin (1991) next consider a linear approximation to the prior geometric average,

$$\theta\left[\beta\left(\frac{\tilde{C}_{t+1}}{C_t}\right)^{-\frac{1}{\rho}}\right]+(1-\theta)\left[\frac{1}{1+\tilde{r}^e_{t+1}}\right], \tag{9}$$

which, when substituted into (8), yields

$$E_t\left\{\theta\left[\beta\left(\frac{\tilde{C}_{t+1}}{C_t}\right)^{-\frac{1}{\rho}}\right]\left(1+\tilde{r}^j_{t+1}\right)+(1-\theta)\left[\frac{1}{1+\tilde{r}^e_{t+1}}\right]\left(1+\tilde{r}^j_{t+1}\right)\right\}=1 \tag{10}$$

As is well known, the standard CAPM relates the undiversifiable risk of an asset to the covariance of its returns with the returns on the market portfolio. When time and risk preferences are distinct, Eq. (10) suggests that both covariances matter: the covariance of an individual asset's return with the market portfolio captures its essential undiversifiable risk, while the covariance of its returns with the consumption growth rate fundamentally captures its risk across time periods. With separate time and risk preferences, it is natural that both should be separately and individually present. But, as we are aware, this alone does not, in general, solve the puzzle.[31]

1.4. Variation in the CRRA and EIS

But what about the basic Epstein–Zin construct generalized to admit time variation in the CRRA and EIS parameters? Several authors, in particular, have suggested that countercyclical risk aversion may allow for resolving the puzzle (cf. Campbell and Cochrane (1999), Gordon and St. Amour (2000, 2003), and Danthine et al. (1992)). Note, however, that such modifications should still respect time-consistent planning and aggregation while also generating stationary equilibrium security returns.

[30]The difficulty in having a pricing kernel that is dependent on the return to the agent's wealth portfolio is the ambiguity that is introduced in the corresponding empirical tests of the model: what is a reasonable proxy for the wealth portfolio? Epstein and Zin (1991) use the traditional "market portfolio" as that proxy, but it can reasonably be argued that this overstates the covariance of investor wealth and individual asset returns. In particular, there is no labor income in their model, and thus their abstraction implies that human capital wealth is a negligible component of total wealth.

[31]Azeredo (2007) confirms that mild positive autocorrelation in the growth rate of consumption also leads to negative equity premia for reasonably parameterized Epstein–Zin utility (EIS near zero) in a standard Mehra and Prescott (1985) setting. For mild positive serial correlation and EIS near 1, the equity premium is positive but still very small.

Simple generalizations to admit countercyclical risk aversion are, first and foremost, problematic along the return stationarity dimension. Consider a simple expected utility, representative agent preference specification of the form

$$E\left(\sum_{t=0}^{\infty} \beta^t u\left(C_t,\, \gamma_t\right)\right),$$

where

$$u\left(C_t,\, \gamma_t\right) = \frac{C_t^{1-\gamma_t}}{1-\gamma_t}$$

with the stochastic process on γ_t, the period CRRA, chosen so that corr $(\gamma_t,\, g_t) < 0$ (g_t denoting the output growth rate, $g_t = Y_t/Y_{t-1} - 1$). An exploration of this non-trivial generalization of the standard Mehra and Prescott (1985) paradigm is revealing (see Danthine et al. (2004) for details). In particular, returns are non-stationary. This latter assertion can be seen from a straightforward calculation of the representative agent's equilibrium ($C_t = Y_t$) intertemporal marginal rate of substitution (IMRS):

$$\text{IMRS}_{t,\,t+1} = \beta\, \frac{u_1\left(Y_{t+1},\, \gamma_{t+1}\right)}{u_1\left(Y_t,\, \gamma_t\right)} = \beta\, \frac{\left(Y_t\right)^{\gamma_t - \gamma_{t+1}}}{\left(g_{t+1}\right)^{\gamma_{t+1}}}. \tag{11}$$

Clearly, the $\text{IMRS}_{t,t+1}$ depends not only on a transformation of the growth rate of output alone, as in Mehra and Prescott (1985), but also directly on the level of output Y_t. The presence of Y_t in expression (11) guarantees that risk-free asset prices and the equilibrium risk-free return will not be a stationary series (the same is true of the risky asset).

We notice also that the presence of γ_t and γ_{t+1} introduces another source of uncertainty into the agent's economic environment, albeit one that he correctly anticipates (rational expectations). It turns out that this single-parameter curvature variation has big consequences for the equity premium and risk-free rate generated by the model. This is presented in Table 3, where the output growth process is otherwise identical to that found in Mehra and Prescott (1985).

Note that under these parameterizations, the problem is not that returns are insufficiently volatile or that the premium is too small, but that both these quantities are too large! Returns on both securities, but especially the equity security, are also much too large relative to the data. The point is not that the model fails, for the idea that one parameter—even if it now assumes two distinct values—should allow enough flexibility to explain a myriad of financial statistics remains a bit preposterous. Rather, what is of interest is the qualitative fact that parameter uncertainty appears to have an enormous influence on the representative agent's demand for securities. But can this effect play itself out more constructively in an Epstein–Zin context?

Melino and Yang (2003) explore a generalized version of Epstein and Zin (1989) that admits parameter uncertainty in the CRRA for "timeless gambles," the EIS, and the

TABLE 3

Variation in the CRRA: Representative Cases (Summary return statistics in percent, annualized; unless otherwise indicated, $\beta = 0.96$, $\pi = 0.47$, $\mu = 0.018$, $\sigma = 0.036$, $N = 120$, $Y_0 = 1$)[i,ii]

	(1) Mehra and Prescott (1985) $\gamma^1 = \gamma^2 = 3$	(2) $\gamma^1 = 1.0$ $\gamma^2 = 1.5$	(3) $\gamma^1 = 0.5$ $\gamma^2 = 1.0$	(4) $\gamma^1 = 4.0$ $\gamma^2 = 4.5$	(5) $\gamma^1 = 1.5$ $\gamma^2 = 1.0$	(6) $\gamma^1 = 1.0$ $\gamma^2 = 1.5$
Er^e	9.58	16.95	16.14	21.5	19.56	18.77
SDr^e	4.99	53.04	53.17	51.94	59.81	58.36
Er^f	9.10	8.47	7.19	15.74	6.84	7.79
SDr^f	1.61	34.21	33.90	35.78	34.61	35.75
Er^p	0.48	8.48	8.945	5.76	12.72	10.98
SDr^p	4.7	34.92	35.26	32.75	43.21	40.32

[i]Panel 1 reports results from the original Mehra and Prescott (1985) model when the CRRA is fixed at $\gamma = 3$.

[ii]The number N indicates the length of the time series of prices and dividends that was used to compute return statistics. $N = 120$ corresponds to 30 years of data. Reported mean returns and standard deviations represent averages of estimates obtained from 1000 independently generated time series of this length.

agent's subjective discount factor:[32]

$$u_t = \left(c_t^{\rho_t} + \beta_t \left[E_t u_{t+1}^{\gamma_t}\right]^{\frac{\rho_t}{\gamma_t}}\right)^{\frac{1}{\rho_t}}. \tag{12}$$

They also allow $\gamma_t \in \{\gamma^H, \gamma^L\}$, $\rho_t \in \{\rho^H, \rho^L\}$, and $\beta_t \in \{\beta^H, \beta^L\}$, where all parameters superscripted by H occur simultaneously, and so also for those superscripted by L (there are two parameter sets); H and L are also coincident with the high and low growth states, respectively. Roughly speaking, they derive expressions for the equilibrium Er^e, SDr^e, Er^f, and SDr^f in the context of a consumption growth process identical to that of Mehra and Prescott (1985) and then ask what sets of parameters allow the resolution of these equations to satisfy the observed values of Er^e, SDr^e, etc.

Melino and Yang (2003) find that certain reasonable sets of parameters do allow the exact replication of the four required historical moments. Generally speaking, this is accomplished with "modest procyclical variation in the EIS, but strong countercyclical

[32]More precisely, under this specification, $1 - \gamma_t$ is the CRRA_t, EIS_t is defined by $1/(1 - \rho_t)$, and β_t is the customary (though now stochastic) subjective discount factor. Formulation (11) leads to an equilibrium IMRS_t of the form $\text{IMRS}_{t+1} = \beta(s_t)\, g_{t+1}^{\gamma(s_t)-1} \left[Q_t/\beta(s_t)\right]^{1-\gamma(s_t)/\rho(s_t)} [1 + Q_{t+1}]^{[\gamma(s_t)/\rho(s_{t+1})]^{-1}}$, where s_t denotes the underlying state influencing β, γ, and ρ; g_t is the period t growth rate, and $Q_t = q_t^e/c_t$, the "price-to-earnings ratio." From our discussion of Epstein and Zin (1989, 1991), we would expect such a term to be present in the IMRS_t.

TABLE 4[i]
Parameter Values that Replicate Er^e, SDr^e, Er^f, SDr^f (Countercyclical CRRA $1-\gamma$ and procyclical EIS $1/(1-\rho_t)$; constant β)

β	γ^L	γ^H	ρ^L	ρ^H
0.950	−25.00	0.89	1.25	1.31
0.960	−51.89	1.85	0.16	0.17
0.965	−18.91	0.67	−0.38	−0.40
0.970	−21.21	0.76	−0.92	−0.97
0.980	−22.25	0.79	−1.98	−2.10
0.990	−22.57	0.81	−3.04	−3.22

[i]This is Table 4 of Melino and Yang (2003).

risk aversion on the part of the representative agent."[33] Variation in the subjective discount factor is found to have only negligible effects. A sample of the relevant parameters is found in Table 4.

The reader may judge for herself whether any of the reported combinations constitute a reasonable resolution of the equity premium and associated puzzles. We venture only to comment that, for all cases, the degree of risk aversion implied by the low growth state seems high, especially in a context where the probability that the low growth state will continue for more than one period is less than 50 percent. There is also the added complication that risk tolerance cannot be uniquely identified with the γ parameter; rather, it appears that β, γ, and ρ all enter into the preference for random consumption sequences. Melino and Yang (2003) is important, for it presents with clarity the importance of countercyclical risk aversion at least as a (joint with procyclical EIS) sufficient quantity for resolving the first and second moments of the equity and risk-free securities.

1.5. Habit Formation

A second approach to modifying preferences, that of incorporating habit formation, was initiated by Constantinides (1990) and Sundaresan (1989).[34]

1.5.1. The Concept

Under this specification, the agent's period-by-period utility is a function not only of his current consumption but also of his past consumption history. In its simplest form, the representative agent is postulated as maximizing

$$E\sum_{t=0}^{\infty}\beta^t u(C_t-\alpha C_{t-1}), \tag{13}$$

[33]Melino and Yang (2003), Section 7.
[34]Ravn et al. (2004) develop a notion of habits being formed at the level of individual goods. It is a notion that has yet to be applied, to our knowledge, in the equity premium literature.

where C_{t-1} denotes the agent's prior period consumption and α measures the intensity with which the prior period's consumption experience influences utility today. Note that preferences are no longer time separable under this specification: by assumption, an increase in the agent's current consumption, while lowering the marginal utility of his consumption today, increases the marginal utility of his consumption tomorrow. Informally, if the "agent eats more today, he ends up being hungrier when he wakes up tomorrow."

A number of variations on this basic model are considered in the literature. Under the most obvious of these generalizations, the agent's "stock" of habit is represented as some known function of his entire consumption history: $u(\,) = u(C_t, S_t)$, where $S_t = S(C_{t-1}, C_{t-2}, \ldots)$ denotes an aggregator function that summarizes the agent's past consumption experience. Alternatively, the effect of the habit may be captured as the ratio of current consumption to its habit or benchmark level as per

$$u(C_t, S_t) = u\left(C_t, \frac{S_t}{C_t}\right), u_1(\,) > 0, u_2(\,) > 0.$$

Both of these representations are referred to as internal habit models, as the magnitude of the habit is determined by the actions of the individual agent alone.[35]

Pollock (1970) and subsequently Abel (1990) postulate a habit that is external to the agent. Since Abel (1990), this perspective has been referred to as "keeping up with the Joneses."[36] It suggests that the agent is primarily concerned with his own consumption relative to the consumption of others in society; for example, $u(\,) = u(C_t - \bar{\alpha}\bar{C}_{t-1})$, where $\bar{C}_{t-1}$ denotes the societal average consumption level in the previous period and $\bar{\alpha}$ the measure of his sensitivity to it. From the agent's perspective, $\bar{C}_{t-1}$ is exogenous. Under this interpretation the agent wakes up feeling hungrier *if those around him* had all enjoyed a large meal. It is an indirect way of modeling an agent's preference for relative "status." In a representative agent equilibrium context $C_t = \bar{C}_t$, but this does not imply that internal and external habit formations lead to equivalent asset pricing relationships.[37]

[35]Habits that enter as a ratio to current consumption are compatible with balanced growth (see King and Rebelo (1988)), whereas difference representations such as (11) are not. In order to restore balanced growth in the latter case, the difference must be normalized by some series that grows on average at the same rate as consumption. Such a normalization will appear in most preference specifications going forward.

[36]See Andrew Abel's chapter in this volume (Abel (2008)).

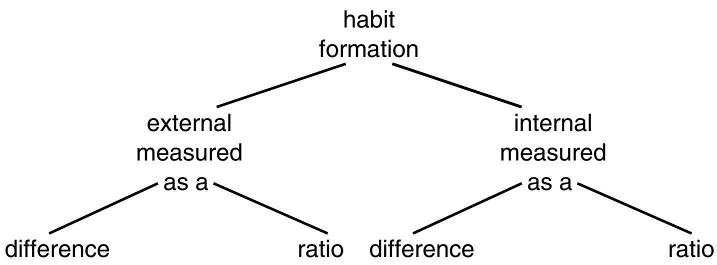
[37]To demonstrate the difference, we need only write out the equilibrium equations that the equity price must satisfy: in the case of an internal habit, $C_{t-1} = \bar{C}_{t-1}$, q_t^e satisfies

a. $u_1(C_t - \alpha C_{t-1})\, q_t^e = \beta \int \left\{ u_1(C_{t+1} - \alpha C_t) \left[q_{t+1}^e + d_{t+1} \right] + u_1(C_{t+1} - \alpha C_t)\, \alpha q_t^e \right\} \, dF(\,),$
where the final term reflects the fact that higher asset purchases in period t reduce the habit in $t+1$. In the case of an otherwise identical external habit,

b. $u_1(C_t - \alpha C_{t-1})\, q_t^e = \beta \int u_1(C_{t+1} - \alpha\, C_t) \left[q_{t+1}^e + d_{t+1} \right] dF(\,).$

With a finite number of possible dividends (and thus a finite number of habit levels), in either case the state-contingent asset prices can be obtained as the solution to a system of linear equations (leading to a level stationary equilibrium). This is directly analogous to the Mehra and Prescott (1985) situation.

We summarize these various alternatives by the tree structure below:



Strictly speaking, preferences structures of this form lie outside the family of recursive preference structures that admit consistent planning for a subtle reason: utility today depends on past consumption, which may depend on consumption plans that are no longer feasible. The way around this is to view past consumption as a habit state variable and view utility today as conditional upon it. In this way, consistency of choice is preserved. Aggregation is also preserved under this simple habit formation, provided $u(\,)$ is of the CRRA form.

1.5.2. Relationship to the Equity Premium

As noted elsewhere in this volume, one source of the equity premium puzzle is the very low volatility in per capita consumption growth. If the risky asset is to command a high rate of return in an environment of very little (per capita) consumption risk, agents must be extremely risk-averse—much more so than micro studies to date have been able to identify.

Incorporating a habit feature sidesteps this issue by postulating that agents are averse to variations in habit-adjusted levels of consumption rather than to consumption variation itself. Especially in circumstances where the current consumption level and the habit do not differ significantly, small changes in consumption growth can generate very large variations in habit-adjusted consumption growth. Even in the presence of modest overall habit-adjusted risk aversion, such variation may contribute to a resolution of the puzzle.

Another way to present the mechanism underlying habit models is to compute the agent's effective (or local) CRRA in the special case where the habit is well proxied by a fixed subsistence level of consumption:[38]

$$u(\,) = u\left(C_t - \bar{C}\right) = \frac{\left(C_t - \bar{C}\right)^{1-\gamma}}{1-\gamma}. \tag{14}$$

[38] See also the discussion in Weil (1989).

Under these circumstances,

$$\frac{-C_t u_{11}(C_t)}{u_1(C_t)} = \frac{\gamma}{\left(1 - \bar{C}/C_t\right)}. \tag{15}$$

Note that if the subsistence level of consumption is high relative to actual consumption (a high threshold for status), then the local CRRA can be very large. For example, if $\gamma = 3$ and $\bar{C}/C_t \approx 0.9$, then the effective CRRA ≈ 30. Under this habit specification, effective risk aversion is time-varying: higher consumption in period $t-1$ provokes the agent to be more risk-averse, ceteris paribus, in period t. It is also countercyclical: when C_t (and output) is low relative to $\bar{C}$ in a recession, $\bar{C}/C_t$ will be high and thus the effective CRRA high. With effective risk aversion not directly observable, these theories will be difficult to test empirically. Nevertheless, these simple relationships give us some idea of what is going on.

They also provide an early warning sign that naive habit formation may not necessarily represent a "cure-all" that allows otherwise standard stochastic equilibrium paradigms to explain the full range of financial phenomena. The basic ingredient of the model, per capita consumption, in particular, does not display much variation in the data. This suggests that under a proper calibration C_t and C_{t-1} will typically be similar in value. With an α close to 1 and standard CRRA utility, the argument $C_t - \alpha C_{t-1}$ will frequently be very small and be on the most concave portion of the agent's utility surface. Marginal utility will be highly volatile and, as a consequence, the risk-free rate is likely to be counterfactually highly volatile as well.[39] This is borne out in many studies for a wide range of settings (see, e.g., Constantinides (1990), Jermann (1998), etc.). Taming this excessive volatility is a particular focus of the habit specification described in the next section.

1.5.3. The Campbell and Cochrane (1999) Mechanism

Campbell and Cochrane (1999) postulate a process on an external habit, X_t, and a period utility function of the simple CRRA form:

$$u(C_t,\, X_t) = \frac{(C_t - S_t)^{1-\gamma}}{1-\gamma}, \tag{16}$$

where C_t in this case denotes the individual agent's consumption and

$$S_t = \left(\frac{C_t - X_t}{C_t}\right),$$

his surplus consumption habit. Note that from the agent's perspective, S_t is exogenous although in equilibrium this will not be so. They next postulate that (1) consumption growth follows an i.i.d. log-normal process,

[39]The reader is also referred to the section on habit formation in Mehra and Prescott (2008b).

$$\Delta c_{t+1} = \log C_{t+1} - \log C_t \equiv g + \tilde{v}_t \text{ where } \tilde{v}_t \sim \text{i.i.d. } N\left(0,\ \sigma_v^2\right), \tag{17}$$

and (2) the log of the surplus consumption ratio $\log S_t = s_t$ also follows a log-normal process to be consistent with (1):

$$s_{t+1} = (1 - \phi)\,\bar{s} + \phi\, s_t + \lambda\,(s_t)\ (\Delta c_{t+1} - g)\,, \tag{18}$$

where ϕ, $\bar{s}$, and g are parameters, $\bar{s} = \log \overline{S}$ ($\overline{S}$ denotes the steady-state surplus consumption ratio), and $\lambda\,(s_t)$ is a prespecified sensitivity function. As a result, the period $t,\ t+1$ pricing kernel is of the form

$$\begin{aligned} \text{MRS}_{t,t+1} &= \beta \frac{u_1(C_{t+1}, S_{t+1})}{U_1(C_t, S_t)} = \beta \left(\frac{S_{t+1}}{S_t} \cdot \frac{C_{t+1}}{C_t} \right)^{-\gamma} \\ &= \beta e^{-\gamma[(1-\phi)(\bar{s}-s_t)+(1+\lambda(s_t))\Delta c_{t+1}-\lambda(s_t)g]} \\ &= \beta e^{-\gamma[(\phi-1)(s_t-\bar{s})+[1+\lambda(s_t)]\tilde{v}_{t+1}]} \end{aligned} \tag{19}$$

by identification (1) above. Under this specification it is apparent that the standard deviation of the MRS, as well as its correlation with consumption growth (both important determinants of the premium), will be determined by the form of the sensitivity function. Campbell and Cochrane (1999) go on to make the following assumption on the form of $\lambda\,(s_t)$:

$$\lambda(s_t) = \begin{cases} \frac{1}{\overline{S}}\sqrt{1 - 2(s_t - \bar{s})} - 1 & s_t \le s_{\max} \\ 0 & s_t \ge s_{\max} \end{cases} \tag{20}$$

where $\overline{S} = \sigma\sqrt{\beta/(1-\phi)}$ and $s_{\max} = \bar{s} + (1/2)\left(1 - \overline{S}^2\right)$.

These requirements allow the habit formulation not only to reflect certain key features of the data but also to generate a plausible habit process.[40] In doing so, the authors attempt to set an upper bound on what can be achieved with a habit model in an exchange setting; that is, they seek to provide a sort of "performance standard" that other competing habit formulations must exceed. No axiomatic foundations for the postulated surplus consumption process are proposed, however.

Lastly, an i.i.d. process on dividend growth is postulated, one that is imperfectly correlated with growth in consumption:

$$\Delta\, d_{t+1} = g + \tilde{\omega}_{t+1}, \quad \tilde{\omega}_{t+1} \sim \text{i.i.d. } N\left(0, \sigma_\omega^2\right) \text{ with } \operatorname{corr}(\tilde{\omega}_t, \tilde{v}_t) = \rho. \tag{21}$$

As before, d_t denotes the log of the actual period t dividend, D_t. With an independently specified dividend process, the $C_t = Y_t$ identity characteristic of Mehra and

[40]To a *first-order* approximation, (14) and (15) imply $X_{t+1} = \phi X_t + (1-\phi)\,(1-\overline{S})C_t$, which appears fairly conventional.

Prescott (1985) is broken, allowing an implicit wage process, levered capital structure, etc. The parameters ρ *and* σ_ω remain to be specified.

The model—and, in particular, the surplus consumption process—is designed to replicate the observed low risk-free rate volatility. Taking advantage of the well-known properties of the log-normal return process in conjunction with model specifications (16)–(19), it becomes a matter of straightforward manipulation to derive the implied expression for log r_t^f:

$$\ln r_t^f = -\ln\beta + \gamma g - \gamma(1-\phi)(s_t - \bar{s}) - \frac{\gamma^2\sigma^2}{2}(1+\lambda(s_t))^2. \qquad (22)$$

Under substitution into (18), expression (22) becomes

$$\ln r_t^f = -\ln\beta + \gamma_g - \frac{\gamma}{2}(1-\phi), \qquad (23)$$

a constant. This clearly is an extreme case of low volatility, and Campbell and Cochrane (1999) offer generalizations by which the precise real risk-free rate volatility can be matched as well.

Plausibility dictates that the habit should not be predetermined, except possibly at the steady state, and formulations (18) and (20) also satisfy this requirement. Indeed, the fact that Δc_{t+1} appears in the expression for the period $t+1$ surplus consumption ratio s_{t+1} dictates that current consumption growth has some influence on the habit.[41] Since the habit is external, it represents the influence on an individual's preference ordering of the average consumption plan of others in the society. Even in an environment of instant communication, it is perhaps unreasonable to presume an immediate and fully external habit adjustment process.

The sensitivity parameter $\lambda(s_t)$ is also increasing in s_t by construction. This guarantees countercyclical risk aversion and, as the authors demonstrate, a countercyclical market price of risk. By making the agent very risk-averse when consumption is already low, this feature accentuates the perceived riskiness of a variable consumption stream. Table 5 presents some sample results of this elaborate exercise.

From these results it is clear that habit formation as articulated by Campbell and Cochrane (1999) is a powerful mechanism for influencing security returns in a dynamic stochastic general equilibrium model. Its flexibility is even greater than what Table 5 suggests since Campbell and Cochrane (1999) subject the model to other tests (e.g., predictability), which it easily transcends. Yet we remain uneasy, as there is no empirical evidence to support the assertion that surplus consumption habits evolve in the way proposed. Furthermore, for the reported simulation, the agent's effective CRRA varies between 1 and 100, which is arguably extreme.

The surplus consumption habit process in Campbell and Cochrane (1999) is highly specialized in other ways. Ljungqvist and Uhlig (1999), in particular, point out that under their specification, the representative agent would experience substantial welfare

[41]Including current consumption as a part of the habit formation process may be used to ensure that period t's consumption never falls below period t's habit. Otherwise, period utility may not be well defined.

TABLE 5
Equilibrium Security Returns and the Premium in a Campbell–Cochrane Model (Rates of return, growth rates annualized in percent; $g = 1.89$, $\sigma = 1.50$, $\ln r^f = 0.94$, $\Phi = 0.87$, $\rho = 0.20$, $\sigma_\omega = 11.2$, $\beta = 0.89$, $\bar{S} = 0.057$, $S_{\max} = 0.094$)

	Postwar data	Simulated moments
Er^f	0.94	0.094
SDr^f	(not reported)	0
Er^e	7.58	7.630
SDr^e	(not reported)	15.200
Er^p	6.69	6.640
SDr^p	15.7	15.200
$E\Delta c_t$	1.89	1.890
$SD\Delta c_t$	1.22	1.220

Source: Campbell and Cochrane (1999), various tables.

gains if 10 percent of her endowment were periodically destroyed. The basic intuition is straightforward: while utility is diminished in the period in which consumption endowment is destroyed, future utility gains result since the habit is correspondingly lower. If the former loss is more than compensated for by the latter gains, the overall result is welfare-enhancing. While this is never the case under standard linear habit evolution, this is possible under the Campbell and Cochrane (1999) construct.[42] These anomalies suggest the need for an axiomatic treatment of habit formation.

1.6. Behavioral Models

Behavioral models are ones for which the postulated agents are not "fully rational." While this can mean that they do not have rational expectations[43] or that they do not

[42] These observations are not as general as would initially appear. Consider the representative agent utility function $\sum \beta^t \{g(x_t) + v(c_t, x_t)\}$, where x_t is the aggregate consumption history, c_t is the agent's time-t consumption and $v(c_t, x_t)$ is increasing and concave in c_t; $g(x_t)$ and $v(c_t, x_t)$ together constitute the agent's period utility function. With an external habit, marginal utility is given by $\partial v(c_t, x_t)/\partial c_t$, independent of $g(x_t)$. The class of utility functions with common $v(c_t, x_t)$ but different $g(x_t)$ support the same equilibrium but may, in general, have different welfare implications. Campbell and Cochrane indeed focus on equilibrium implications driven exclusively by marginal utility $\partial v(c_t, x_t)/\partial c_t$. It is not the case that all utility functions in the class will necessarily exhibit identical welfare implications and therefore the criticism leveled by Ljungquist and Uhlig (1999) against Campbell–Cochrane cannot be viewed as a general statement. We thank George Constantinides for pointing this out to us.

[43] By this we mean, in the context of this essay, that agents know the exact probability distribution governing all random outcomes—returns, consumption, etc.—driving the model.

update their beliefs using Bayes' rule, in the context of this essay it will be taken to mean that the agents make decisions in accordance with preferences that are not expected utility defined over consumption or wealth (indirect utility). In this sense, "habit formation" preferences are behavioral, although their successful use in explaining a wide class of phenomena has given them a degree of acceptance not typical of the broader behavioral finance literature. While the proposed family of behavioral preference structures is otherwise quite large, most have not been applied to the study of the "equity premium" and associated puzzles.[44] For this reason, we restrict our attention to four: (1) the happiness maintenance preferences of Falato (2003), (2) the disappointment aversion and generalized disappointment aversion of Gul (1991) and Routledge and Zin (2004), (3) the prospect theory of Kahneman and Tversky (1979) as applied, most especially, by Barberis et al. (2001), and (4) the rank-dependent expected utility preference structures of Quiggin (1982, 1993).

1.6.1. Happiness Maintenance

Consistent with a large body of experimental evidence (see Isen (1989) for a review), Falato (2003) builds on the assumption that risk aversion may be procyclical. The phenomenon he seeks to capture is the following: investors who are experiencing high utility—due either to a favorable income or to wealth shock—become more risk-averse, as they do not wish these favorable circumstances to be lost; that is, they seek "happiness maintenance." Falato's model postulates that investors derive direct utility from both consumption and the consumption-to-wealth ratio. (As with Melino and Yang (2003), the associated MRS in his model also necessarily contains as an argument the consumption-to-wealth ratio.) In particular, Falato (2003) assumes that the agent's period utility is defined over a composite good

$$u(\,) = u(g_t) = \frac{(g\,(C_t, W_t; \theta_t))^{1-\gamma}}{1-\gamma},$$

where $g(C_t, W_t; \theta_t) = C_t^{1-\theta_t}\, W_t^{\theta_t}$ and W_t, wealth, satisfies

$$W_t = \left(q_t^e + Y_t\right) z_t^e$$

(the setting is otherwise a pure Mehra and Prescott (1985) exchange economy). The agent's full preference ordering over random consumption sequences is thus given by

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} \left(\frac{W_t}{C_t}\right)^{(1-\gamma)\theta_t}, \tag{24}$$

[44]For the wider classes that have been used for macroeconomic studies, see Backus et al. (2004).

where $\theta_t = \theta\, \bar{x}_t(n)$ and

$$\bar{x}_t\,(n) = \frac{1}{n+1} \sum_{t=0}^{n} \left(\frac{Y_{t-\tau}}{Y_{t-\tau-1}} \right). \tag{25}$$

If $n = 0$, Eq. (25) reduces to

$$\theta_t = \theta \left(\frac{Y_t}{Y_{t-1}} \right).$$

It is this latter feature that captures the "happiness maintenance" phenomenon underlying Falato's (2003) work: high average dividend (consumption) growth in the recent past, $\bar{x}_t(n)$, leads the agent to be more risk-averse over wealth in the current period via the exponent $(1-\gamma)\theta_t$ of the wealth-to-consumption ratio.

Falato (2003) proves that equilibrium prices exist for a Mehra and Prescott (1985) economy when the representative agent has "happiness maintenance" preferences.[45] Furthermore, the postulated ordering admits consistent planning and generates stationary return series. Preliminary analysis suggests that a representative agent may be constructed. The issue of full aggregation remains open, however.

As with Melino and Yang (2003), the methodology required to calculate state-contingent asset prices and returns involves the solution to a non-linear system of equations. Table 6 contains the results of this exercise for a sample set of parameters.

We are drawn particularly to the $\gamma = 3$, $\theta = 0.24$ case, which matches nearly all the moments exceptionally well, with the sole qualification that the *SD* of the equity return is about 30 percent too large. This is furthermore accomplished in a model with fewer degrees of freedom than, e.g., Melino and Yang (2003). The presence of the consumption-to-wealth ratio as an argument of the period utility function is not entirely surprising given earlier work by Lettau and Ludvigson (2001). What is perhaps more striking (but not reported in the table) is the ability of Falato's model also to explain the predictable feature of stock returns (see, e.g., Campbell and Shiller (1988)) at least as well as the habit construct of Campbell and Cochrane (1999).[46] Countercyclical risk aversion is clearly not a necessary condition for this class of models to replicate, acceptably, the equity premium and allied statistics. Furthermore, the addition of the (C_t/W_t) ratio does not require the presence of a scaling factor since the consumption-to-wealth ratio series is endogenously stationary.

The remaining behavioral approaches rely directly on countercyclical risk aversion. In both of them, the investing agents are concerned especially with bad outcomes relative to some benchmark, and it is how this latter quantity is defined that sets them apart.[47]

[45] Such an exercise is non-trivial. The main challenge to establishing existence derives from the endogenous nature of the pricing kernel, which is, in turn, the result of the utility function's dependence on wealth.

[46] In Falato's (2003) model, the cost of the business cycle is also about ten times the original estimate of Lucas (1985). See also Alvarez and Jermann (2004).

[47] As such, they are direct descendants of Reitz (1988). Reitz (1988) retains the expected utility representation and postulates a low-probability-very-low-consumption state. In Routledge and Zin (2004) and Barberis et al. (2001), the "disaster" is measured relative to a benchmark. This fact facilitates a more reasonable model calibration.

TABLE 6
Financial Statistics under Happiness Maintenance[i] (All cases: $\beta = 0.99$, $n = 2$, $\pi = 0.43$, $\mu = 0.018$, $\delta = 0.036$; various γ, θ; returns annualized in percentage terms)[ii]

		$E(r^f)$	$E(r^e)$	$E(r^p)$	$SD(r^f)$	$SD(r^e)$	$SD(r^p)$
	$\gamma = 1.1$	2.36	2.56	0.19	0.71	4.04	3.98
$\theta = 0.24$	$\gamma = 2$	2.85	3.75	0.90	1.87	8.12	7.90
	$\gamma = 3$	0.84	6.95	6.11	5.55	23.17	22.57
	$\gamma = 1.1$	2.48	2.66	0.17	0.68	3.94	3.88
$\theta = 0.18$	$\gamma = 2$	3.27	3.89	0.63	1.64	6.66	6.45
	$\gamma = 3$	3.46	5.60	2.15	3.69	12.69	12.14
	$\gamma = 1.1$	2.60	2.76	0.16	0.63	3.85	3.80
$\theta = 0.12$	$\gamma = 2$	3.62	4.09	0.46	1.39	5.59	5.41
	$\gamma = 3$	4.49	5.59	1.09	2.62	8.38	7.95
	$\gamma = 1.1$	2.72	2.87	0.15	0.60	3.77	3.72
$\theta = 0.06$	$\gamma = 2$	3.94	4.30	0.36	1.20	4.83	4.67
	$\gamma = 3$	5.16	5.85	0.69	2.00	6.18	5.85
	$\gamma = 1.1$	2.85	2.99	0.14	0.56	3.69	3.65
$\theta = 0.00$	$\gamma = 2$	4.27	4.55	0.29	1.04	4.23	4.10
	$\gamma = 3$	5.74	6.22	0.48	1.57	4.87	4.60

[i] Table 4.1 in Falato (2003).
[ii] The μ, δ parameters refer to the output growth process in Mehra and Prescott (1985).

We emphasize the Routledge and Zin (2004) generalization of Gul (1991) since the latter model does not satisfactorily explain the premium in the context of a straightforward Lucas (1978)-Mehra and Prescott (1985) tree economy while the former does.[48]

1.6.2. Generalized Disappointment Aversion

Disappointment-averse preferences place a greater utility weight on outcomes that disappoint. For Gul (1991), this threshold is to be identified with consumption realizations that lie below the (conditional) certainty equivalent consumption level, while for Routledge and Zin (2004) outcomes more distant from the certainty equivalent are given special weighting.[49]

[48] Bonomo and Garcia (1993), however, combine a joint Markov switching endowments model with a disappointment version to match the first and second moments of the equity premium and the risk-free rate.
[49] Gul (1991) disappointment aversion (DA) preferences in a standard Mehra–Prescott (1985) economy do not per se resolve the premium. Bonomo and Garcia (1993) enhance the construct by incorporating a joint process on consumption and dividends that follows a Markov switching model—in addition to (DA) preferences. They are able to match the risk-free rate and the first two moments of the premium.

As in the Epstein and Zin (1989) utility construct, time and risk preferences are separated. Their recursive intertemporal utility function $U(C_t, CE_{t+1})$ has the form

$$U(C_t, CE_{t+1}) = \left[\left((1 - \frac{1}{1+\rho}\right) C_t^{\gamma} + \frac{1}{1+\rho} CE_{t+1}^{\gamma}\right]^{1/\gamma}, \tag{26}$$

where CE_{t+1} is the certainty equivalent pertaining to random consumption in the following period. In functional form (26), ρ is the marginal rate of time preference and $1/1-\gamma$ the elasticity of intertemporal substitution.

Risk preferences and the certainty equivalent computation are specified as follows. Let us presume a pure Mehra and Prescott (1985) setting with a finite number of states and suppose, in some period t, that the output level is $Y_t^i = Y^i$. Temporal risk preferences are represented as a period utility function of the form $u(C) = C^{\alpha}/\alpha\ (\alpha \neq 0)$ or $u(C) = \log C (\alpha = 0)$. The period $t+1$ certainty equivalent consumption level $CE_{t+1} = CE_{t+1}\left(\tilde{Y}^i\right)$ is then endogenously determined as the solution to the following definitional equation:

$$\begin{aligned} u\left(CE_{t+1}\left(\tilde{Y}^i\right)\right) &= \frac{\left(CE_{t+1}\left(\tilde{Y}^i\right)\right)^{\alpha}}{\alpha} \\ &= \sum_j \pi_{ij} \frac{\left(Y^j\right)^{\alpha}}{\alpha} - \theta \sum_{Y^j \leq \delta CE_{t+1}\left(Y^i\right)} \pi_{ij} \left[\frac{\left(\delta CE_{t+1}\left(Y^i\right)\right)^{\alpha}}{\alpha} - \frac{\left(Y^j\right)^{\alpha}}{\alpha}\right]. \end{aligned} \tag{27}$$

If $\theta = 0$, structure (27) reduces to the definition of a certainty equivalent in a pure expected utility context.[50] Otherwise, there is a penalty, proportional to θ, for consumption realizations that lie sufficiently below the endogenously determined certainty equivalent. If $\delta = 1$, the certainty equivalent computation follows according to Gul (1991); if $\delta < 1$, the context is that of Routledge and Zin (2004). Note that there is no mention of any external reference point as in the (external) habit formation literature.

Nevertheless, the model generates countercyclical risk aversion, a feature that bodes well for its asset pricing characteristics. Furthermore, GDA preferences are homothetic, are linear in probabilities, and have the same aggregation properties as CRRA expected utility structures. They admit consistent planning, and Routledge and Zin (2004) provide the axiomatic underpinnings.

Table 7 gives a quantitative idea of how well GDA preferences perform in the standard Mehra and Prescott (1985) environment.

A number of these results are worthy of explicit mention. When dividend volatility is low, GDA preferences give only modest premia; they become quite large only when $SD(Y_t) = 0.10$. For the data period underlying the empirical background to the

[50] When $\theta \neq 0$, the Gul-Routledge-Zin structures relax the independence assumption that underlies the expected utility representation. Independence fails because mixing in some arbitrary payoff lottery with two lotteries under comparison may fundamentally change the tail properties of each.

TABLE 7[i]
Equilibrium Security Returns and the Premium: GDA Preferences (Various parameter combinations; all returns in percent)

	Parameter choices					
	A	B	C	D	E	F
α	1	1	0.5000	−2	1	1
γ	1	1	1	1	0.5000	−0.5000
θ	9	24	9	9	9	9
δ	0.9692	0.9692	0.9692	0.9692	0.9692	0.9692
ρ	0.0300	0.0300	0.0300	0.0300	0.0300	0.0300
	$SD(Y_t) = 0.05$					
	A	B	C	D	E	F
Er^f	1.74	1.56	1.72	1.62	2.45	3.89
SDr^f	1.26	1.44	1.25	1.23	1.21	1.11
Er^e	4.30	4.68	4.33	4.46	5.07	6.63
SDr^e	5.72	6.14	5.68	5.52	5.86	6.14
Er^p	2.56	3.12	2.61	3.84	3.86	2.74
SDr^p		not reported				
	$SD(Y_t) = 0.10$					
	A[ii]	B	C	D	E	F
Er^f	1.74	1.56	1.72	1.62	2.45	3.89
SDr^f	1.26	1.44	1.25	1.23	1.21	1.11
Er^e	8.07	9.66	8.17	8.59	8.86	10.46
SDr^e	14.37	16.26	14.27	13.80	14.55	14.91
Er^p	6.33	8.10	6.45	6.97	6.65	9.35
SDr^p		not reported				

[i]Table 7 is an amalgamation of selected results reported in Tables 1 and 2 in Routledge and Zin (2004).
[ii]The return statistics for the risk-free rate are identical for the $SD(Y_t) = 0.05$ and 0.10 cases because risk-free securities do not result in disappointments.

article, de-trended dividend volatility is on the order of 12 percent, which advocates for the model with $SD(Y_t) = 0.10$; Campbell (1999) argues that the *SD* of dividend should be as high as 28 percent. Qualitative results generally match intuition; comparing columns A and B, the result of a greater penalty is to raise the risk-free asset prices and lower equity ones: risky securities are less desirable and the premium rises. Comparing columns B and D, the representative agent becomes atemporally more risk-averse and the premium naturally increases. In general, equity volatilities ($SD(Y_t) = 0.10$ case) match the data quite well; risk-free rate volatilities are generally too small. Although not part of Table 7, it can also be shown that the risk-free rate is procyclical. Overall, the results are very good, and the model is advantaged by retaining a utility specification defined only over the agent's consumption. Note also that countercyclical risk aversion is generated endogenously within the (GDA) construct. To see this, consider the cases of columns A and B: disappointment aversion is the only source of risk aversion ($\alpha = 1$). As a result, the pricing kernel is risk-neutral in the high growth states (no disappointing outcomes) but substantially risk-averse in the low growth states (see Routledge and Zin (2004) for a more detailed analysis).

There is one other major structure that provides countercyclical risk aversion: prospect theory as applied to asset pricing.

1.6.3. Prospect Theories

The major reference is Barberis et al. (2001). See also the essay by Barberis and Huang in this volume. The idea is to postulate investors who derive utility not only from their period-by-period real consumption, but also over *equity* portfolio gains and losses relative to a plausible benchmark.[51] Roughly speaking, these investors max $U(C) + V(G/L)$, where G/L denotes gains or losses. That utility is defined over only equity gains/losses rather than over aggregate gains and losses to total wealth is an illustration of the notion of "narrow framing."

Furthermore, investors are more sensitive to losses than to gains (this is the essence of the notion of "loss aversion"), with the extent of the loss sensitivity depending on the agent's prior portfolio experience. Equity gains and losses are measured with respect to a benchmark, and if the investor's past experience has been to sustain losses relative to his benchmark, then he is modeled as being more acutely sensitive to further ones. If his recent experience has been one of equity portfolio gains, then the agent is modeled as being less sensitive to losses (provided they are not so severe as to negate past gains). These requirements lead to a kink in the agent's utility curve at gain = loss = 0. Roughly speaking, loss aversion as captured in Barberis et al. (2001) implies linear utility over gains and losses to the equity portion of an investor's portfolio of the form

$$V(X_{t+1},\ S_t,\ z_t) = \begin{cases} X_{t+1}, & X_{t+1} \geq 0, \\ \lambda(z_t)\, X_{t+1}, & X_{t+1} < 0, \end{cases}$$

where

$$X_{t+1} = S_t R^e_{t+1} - R^f_t\, S_t.$$

Following our customary notation, R^e_{t+1} and R^f_t, designate, respectively, the gross rates of return on stock and risk-free bonds from period t to $t + 1$ and S_t the value of the stock portion of the investor's portfolio. The latter is subscripted by t alone, as the risk-free

[51]See also Benartzi and Thaler (1995). These authors postulated investors with loss-averse preferences over variations in their financial wealth only. Without any direct connection to consumption, it is impossible to ascertain how their model might describe the joint processes on equilibrium returns and consumption growth, for example.

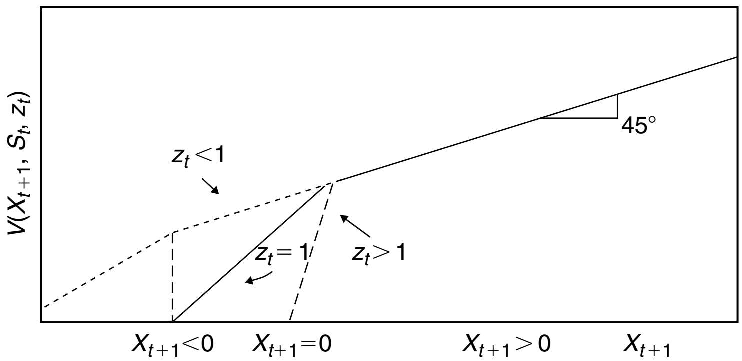


FIGURE 1 $V(X_{t+1}, S_t, z_t)$: Utility of equity portfolio gains or losses. Source: From Barberis et al. (2001).

rate from period t to $t+1$ is solely determined by the risk-free bond's price in period t. Gains or losses are thus measured relative to what would have been earned had the stock portion S_t of his portfolio been invested in risk-free assets. The slope parameter $\lambda(z_t)$ measures the degree of loss sensitivity which is dependent on a variable, z_t, that recalls past portfolio experience, where $z_t = 1$ captures a case of no prior gains or losses, $z_t < 1$ means prior gains, and $z_t > 1$ means prior losses. A representative graphical portrayal is found in Figure 1.[52]

More formally, in a discrete-time asset pricing context, the representative agent undertakes to assemble portfolios of the stock market index with value S_t in conjunction with an aggregate valuation of risk-free bonds, B_t, so as to solve

$$\max\ E\left(\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} + b_t\, \beta^{t+1}\, V(X_{t+1},\, S_t, z_t)\right), \tag{28}$$

where the first term is the standard present discounted utility of consumption under CRRA utility, and the second, $V(\,)$, captures the utility/disutility of gains and losses, X_{t+1}, in the stock market portion of the optimal portfolio, S_t; the z_t-term is as indicated above. The standard period t budget constraint and the constraint governing the evolution of wealth are, respectively,

$$C_t + S_t + B_t \leq W_t, \tag{29}$$

$$W_{t+1} = S_t\, R^e_{t+1} + B_t\, R^f_t, \tag{30}$$

where W_t represents period t wealth and S_t, B_t the amounts of it allocated, respectively, to stock and risk-free bonds post-consumption, C_t, in period t. It remains to detail the precise form of $V(\,)$. Barberis et al. (2001) postulate

[52] An amplified version of this same representation may be found in the companion chapter by Barberis and Huang (2008).

$$V(X_{t+1}, S_t, z)_t = \begin{cases} X_{t+1} & R^e_{t+1} \geq z_t R^f_t, \text{ and } z_t \leq 1, \\ S_t(z_t R^f_t - R^f_t) + \lambda S_t(R^e_{t+1} - z_t R^f_t), & R^e_{t+1} < z_t R^f_t, \text{ and } z_t \leq 1, \\ X_{t+1} & R^e_{t+1} \geq R^f_t, \text{ and } z_t > 1, \\ \lambda(z_t) X_{t+1} & R^e_{t+1} < R^f_t, \text{ and } z_t > 1, \end{cases} \tag{31}$$

where $\lambda(z_t)$ and z_{t+1} evolve according to

$$\lambda(z_t) = \lambda + k(z_t - 1)\,, \tag{32}$$

and

$$z_{t+1} = \eta\left(z_t \frac{\overline{R}^e}{R^e_{t+1}}\right) + (1 - \eta)1, \tag{33}$$

where $\lambda > 1$, $\eta > 0$, and $k > 0$ are constants and z_t is an index that captures whether the investor has recently experienced gains ($z_t \leq 1$) or losses ($z_t > 1$).

Some careful interpretation is in order. Equation (33) describes the evolution of the historical experience variable, z_t, where $\overline{R}^e$ amounts to the average stock return and η is a parameter that captures the memory in the adjustment process.[53] If equity returns are much less than average ($\overline{R}^e / R^e_{t+1} > 1$), this experience tends to increase $z_{t+1} > z_t$ since $\eta \geq 0$. If $\eta = 0$, Eq. (31) suggests that $z_t \approx 1$ or that prior gains and losses have an immediate effect on the investor's risk attitude. If $\eta = 1$, then the benchmark adjusts slowly. Note that $\lambda(z_t)$, the loss sensitivity, is increasing in z_t. Otherwise, $b_t = b_0(\overline{C})^{-\gamma}$, where $(\overline{C})^{-\gamma}$ is purely a scaling factor that is included to maintain stationarity in the equilibrium asset return distributions and b_0 determines the relative utility of gains and losses vs. the utility of consumption. There is, however, no formal proof that investor preferences of this type survive aggregation and no axiomatic foundation for their existence.

It remains to close the economy by specifying the exogenous dividend and consumption processes (as in any exchange model, there is one equity security outstanding, and risk-free bonds are priced in zero net supply). Barberis et al. (2001) consider a number of scenarios. We first report the one that is in the spirit of a pure Mehra and Prescott (1985) setting:

$$\log\left(\frac{C_{t+1}}{C_t}\right) = \log\left(\frac{D_{t+1}}{D_t}\right) = g_C + \sigma_C \tilde{\varepsilon}_{t+1},$$

where $\tilde{\varepsilon}_{t+1} \sim$ i.i.d. $N(1, 0)$, where $g_C = 1.84$ and $\sigma_C = 3.79$ (both expressed in percent) as estimated from data for the period 1926–1995. The results are presented in Table 8.

[53] $\overline{R}^e$ is actually determined within the model in order that, at equilibrium, the median value of $Z_t \equiv 1$. This turns out, not surprisingly, to be approximately the average equity return.

TABLE 8[i]

The Equity Premium in a Model with Loss Aversion ($\gamma = 1, \beta = 0.98, \lambda = 2.25, \eta = 0.9, g_C = 1.84\%, \sigma_C = 3.79\%$; various b_0, k; annualized returns in percent)

	Data	$b_0 = 0.7$ $k = 3$	$b_0 = 2$ $k = 3$	$b_0 = 100$ $k = 3$	$b_0 = 7$ $k = 150$	$b_0 = 2$ $k = 100$	$b_0 = 100$ $k = 50$
$\log r^f$	0.58	3.79	3.79	3.79	3.79	3.79	3.79[ii]
$E \log r^P$	6.03	0.63	0.88	1.26	3.50	3.66	3.28
$SD \log r^P$	20.02	4.77	5.17	5.62	10.43	10.22	9.35
Sharpe ratio	0.30	0.13	0.17	0.22	0.34	0.36	0.35
Average loss aversion		2.25	2.25	2.25	10.70	7.50	4.40

[i]This is Table II of Barberis et al. (2001).
[ii]Under their specification it can be shown that $\log r^f$ is constant.

TABLE 9[i]

The Equity Premium in a Model with Loss Aversion and Distinct Consumption and Dividend Processes ($g_D = g_C = 1.84, \sigma_C = 3.79, \sigma_D = 0.12$, w = 0.15, $\eta = 0.9, \gamma = 1, \beta = 0.98, \lambda = 2.25$; all return measures in percent)

	Data	$b_0 = 0.7$ $k = 3$	$b_0 = 2$ $k = 3$	$b_0 = 100$ $k = 3$	$b_0 = 0.7$ $k = 20$	$b_0 = 2$ $k = 10$	$b_0 = 100$ $k = 8$
$\log r^f$	0.58	3.79	3.79	3.79	3.79	3.79	3.79
$E \log r^P$	6.03	1.30	2.62	3.68	5.17	5.02	5.88
$SD \log r^P$	20.02	17.39	20.87	20.47	25.85	23.84	24.04
Sharpe ratio	0.30	0.07	0.13	0.18	0.20	0.21	0.24
$\rho\left(\frac{C_{t+1}}{C_t}, r^P_{t+1}\right)$	0.10	0.15	0.15	0.15	0.15	0.15	0.15
Average loss aversion		2.25	2.25	2.25	5.80	3.50	3.20

[i]This table is drawn from Table III in Barberis et al. (2001).

Note that for very large k, which leads via Eq. (30) to very high loss aversion in some states of the world, the premium can rise to a level of 3.50 percent in conjunction with about half (10.4 percent) of the observed historical equity return volatility. Increases in the value of the parameter b_0, which place greater overall relative weight on the utility of gains and losses, also lend to increases in the premium, but only to 1.26 percent when $b_0 = 100$. Clearly, it is the loss aversion parameter that has the power.

Table 9 presents results when the consumption and dividend processes are specified independently:

$$\log\left(\frac{C_{t+1}}{C_t}\right) = g_C + \sigma_C \tilde{\xi}_{t+1},$$

$$\log\left(\frac{D_{t+1}}{D_t}\right) = g_D + \sigma_D \tilde{\varepsilon}_{t+1},$$

$$\begin{pmatrix} \xi_t \\ \varepsilon_t \end{pmatrix} \sim \text{i.i.d } N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix}\right).$$

Comparing Tables 8 and 9, it is apparent that separating consumption from dividends has a substantial effect on raising the premium as well as on increasing return volatilities. This is not qualitatively entirely surprising, since there is now much greater dividend volatility–which makes equities more volatile–and this in turn affects an agent's $V(\,)$ utility much more significantly. Although stock returns are now less highly correlated with consumption (Table 9), this does not matter since the investor is concerned about stock market volatility per se. This latter fact more than compensates for the correlation diminution. By way of contrast, this feature is relatively unproductive in habit-style models as the increased volatility of returns is offset by its reduced correlation with consumption; the net consequence is very little change in the premium. An increase in the parameter k, by increasing loss aversion, has a similarly salutary effect on the premium as it does in Table 8.

Loss aversion/narrow framing is an appealing idea, and Barberis et al. (2001) analyze its equilibrium asset pricing implications in a careful and through way. There is, however, a sense in which their study is premature. In particular, we as yet lack choice theoretic underpinnings, and the aggregation properties are as yet unconfirmed (strict aggregation will not hold).

1.7. Beyond One Good and a Representative Agent

All of the models considered in the prior sections of the review have been essentially one good, representative agent models. In this section we first highlight elements of the multi-goods literature, which otherwise retains the representative agent construct. Subsequently, we explore a one good heterogeneous agent model where preferences are postulated to have a non-standard expected utility form (rank-dependent expected utility).

1.7.1. Multiple Goods

A basic reference here is Piazzesi et al. (2007), which considers preferences defined over two goods, all non-housing consumption (this category comprises non-durable goods consumption and the flow of services from non-housing durables) and housing services. Adopting an otherwise standard Lucas-(1978) style exchange setting, the authors postulate the representative agent's preferences to be of the form

$$\max\ E\left(\sum_{t=0}^{\infty} \beta^t u(C_t)\right), \tag{34}$$

where

$$u(C_t) = \frac{C_t^{1-\frac{1}{\sigma}}}{1 - 1/\sigma}, \tag{35}$$

$$C_t = (C_t^{(\varepsilon-1)/\varepsilon} + \omega s_t^{(\varepsilon-1)/\varepsilon})^{\varepsilon/(\varepsilon-1)}. \tag{36}$$

In (34)–(36), c_t denotes non-housing consumption and s_t the flow of housing services; accordingly, σ is the intertemporal elasticity of substitution and ε is the corresponding intratemporal elasticity. For high ε values, the agent is very willing to substitute goods within a period; as $\varepsilon \to \infty$, they become perfect substitutes. In this way, the level of housing services directly affects the marginal utility of non-housing consumption (the numeraire), and therein lies an additional source of consumption risk. In severe recessions where both housing services and non-housing consumption are simultaneously low, for example, the marginal utility of non-housing consumption is especially high as the agent tries to substitute future non-housing consumption for the anticipated shortfall in future housing services. This increases MRS volatility.

An important methodological insight allows the equilibrium pricing kernel to be expressed in terms of non-housing consumption and the share of income going to housing services, an insight that avoids the challenge of constructing a time series for the somewhat imprecise notion of "housing services." For the parameterization $1/\sigma = 5$, $\beta = 0.99$, and $\varepsilon = 1.05$, they obtain a premium of 3.5 percent in conjunction with a risk-free return of 1.8 percent and a premium standard deviation of 11 percent. None of these parameter values is a priori objectionable.[54]

Another study in the multi-good tradition is Ait-Sahalia et al. (2004), which characterizes investor consumption behavior using time series of luxury-goods purchases. Roughly speaking, the idea is that since the "rich" own a vastly disproportionate share of non-pension equity and debt, their consumption behavior should matter for equity pricing. For this group, furthermore, the marginal consumption items are likely to be luxury goods (proxied by sales of expensive automobiles, the sales of high-end Manhattan co-operative apartments, etc.).

The distinction between luxury and basic consumption goods is captured in their utility specification:

$$u(c_t, L_t) = \frac{(c_t - a)^{1-\phi}}{1-\phi} + \frac{(L_t + b)^{1-\varphi}}{1-\varphi} \tag{37}$$

with constants $a, b > 0$ and $\phi > \varphi$.

The former assignment implies that subsistence consumption is positive for basic goods and negative ($-b$) for luxury goods. In equilibrium, luxury goods are not consumed by those with low consumption levels (the poor). In fact, there exists

[54]Piazzesi et al. (2007) rely on two sources of uncertainty: consumption growth and the log expenditure ratio for housing. For that study it is important that they be statistically independent, a property that is assumed. Others have questioned its validity.

a $\underline{c} = a + b^{\varphi/\phi} P_t^{1/\phi} > a$ such that $L_t = 0$ for $c_t \leq \underline{c}$ (P_t is the relative price for luxury goods vis-à-vis basic ones). Finally, the rich consume an increasing fraction of luxury goods as their incomes rise:

$$\lim_{x_t \to \infty} \left(\frac{c_t}{X_t} \right) = 0, \tag{38}$$

where $X_t = c_t + P_t L_t$ denotes total consumption expenditure. Of relevance to this essay is the fact that this specification leads to two unconditional asset pricing equations of the form

$$E\left(\beta \left(\frac{c_{t+1} - a}{c_t - a} \right)^{-\phi} \left(r^e_{t+1} - r^f_{t+1} \right) \right) = 0, \tag{39}$$

$$E\left(\beta \left(\frac{L_{t+1} + b}{L_t + b} \right)^{-\varphi} \left(\frac{P_t}{P_{t+1}} \right) \left(r^e_{t+1} - r^f_{t+1} \right) \right) = 0. \tag{40}$$

Ait-Sahalia et al. (2004) then focus their model on a number of issues related to and inclusive of the equity premium puzzle. Using various self-constructed time series, each corresponding to the aggregate purchase of some type of luxury goods (e.g., luxury cars, wine, apartments, etc.), as proxy consumption sequences, the authors estimate, respectively, the implied CRRAs and equity premia, and explore the ability of the model to replicate the cross-sectional returns of the 25 Fama and French (1993) portfolios. Broadly speaking, they find that certain of these series co-vary strongly with the observed return on equity and lead to high premia (7.8 percent, using an index of Manhattan co-op apartment prices). In most cases, the CRRA estimates are on the order of one-tenth that obtained using standard NIPA aggregate consumption measures: since this latter series hardly varies at all and is little correlated with equity returns. The idea is certainly an appealing one, and the results are impressive. Yet the by-now-familiar questions remain: what are the axiomatic foundations, in what sense is aggregation allowed, etc.?

A less radical foray into multiple-goods territory is found in Yogo (2006). As such, Yogo (2006) provides another illustration of both the potential and the limitations of the Epstein–Zin utility specification. He considers a model in which the representative agent's *intratemporal* utility level is a CES function of non-durable goods consumption, C_t, and the stock of durable goods, D_t:[55]

$$u_t = u\left(C_t, D_t\right) = \left[(1 - \alpha) C_t^{1-\frac{1}{\rho}} + \alpha D_t^{1-\frac{1}{\rho}} \right]^{\frac{1}{1-\frac{1}{\rho}}}. \tag{41}$$

[55]In models of this type, the non-durable good is always the numeraire. Technically, period utility is a function of non-durable goods consumption and the service flow from the stock of durable goods. Yugo (2006) assumes the service flow is directly proportional to the stock.

This is coupled with an Epstein–Zin *intertemporal* utility specification of the form

$$U_t = \left\{ (1-\beta)\left[u\left(C_t, D_t\right)\right]^{1-\frac{1}{\sigma}} + \beta \left(E_t \left[U_{t+1}^{1-\gamma} \right] \right)^{\frac{1}{\kappa}} \right\}^{\frac{1}{1-\frac{1}{\sigma}}}. \tag{42}$$

The corresponding $\text{MRS}_{t,t+1}$ is thus

$$\text{MRS}_{t,t+1} = \left[\rho \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\sigma}} \frac{v(D_{t+1}/C_{t+1})^{\frac{1}{\rho}-\frac{1}{\sigma}}}{v(D_t/C_t)} \cdot R_{w,t+1}^{1-\frac{1}{\kappa}} \right]^{\kappa}, \tag{43}$$

where

$$\kappa = (1-\gamma)/(1-1/\sigma)$$

$$v\left(D_t/C_t\right) = \left[1 - \alpha + \alpha \left(D_t/C_t\right)^{1-\frac{1}{\rho}}\right]^{\frac{1}{1-\frac{1}{\rho}}}$$

and $R_{w,t+1}$ is the return on wealth from the optimal portfolio.

Following Hansen and Singleton (1982), Yogo (2006) then estimates the parameter values $(\alpha, \beta, \rho, \sigma, \gamma)$ that are consistent with (1) the above model and (2) the historical returns on the 25 Fama–French portfolios.[56] He obtains the estimates $\sigma = 0.024$, $\alpha = 0.827$, $\rho = 0.52$, $\beta = 0.9$ and $\gamma = 191$, all of which seem plausible except for the risk aversion estimate and, possibly, the relative intratemporal emphasis on durable goods ($\alpha = 0.827$).

That these values can, via (41)–(43), match the data quite well is attributable to the additional terms in the $\text{IMRS}_{t,t+1}$ expression, which are dependent on the (D_t/C_t) ratio. In the data, this ratio is highly procyclical and thus magnifies the countercyclicality of the marginal utility of consumption. Marginal utility becomes very high at business cycle troughs where both consumption itself and the (D_t/C_t) ratio are low. Nevertheless, because durable consumption is so smooth, the required CRRA estimate is still excessive relative to that obtained in standard microeconomic applications. While the Epstein and Zin construct is flexible in accommodating many phenomena, it does not necessarily do so for reasonable parameter choices. Note that for this story to work, $\rho < \sigma$. In comparing financial statistics, it is the low value of the EIS that allows the model to replicate the historically observed risk-free rate.

Multiple goods are not, however, any guarantee of puzzle resolution. Although Yogo (2006) obtains generally attractive results in a model with durable and non-durable goods within the Epstein–Zin preference specification, Giannikos (2007) derives equilibrium premia similar to that obtained in Mehra and Prescott (1985) using standard CES preferences defined over non-durable goods consumption and various representations of the durable goods service flow. His results are driven by positive autocorrelation in the estimated non-durable consumption dividend, a fact that leads to a negative premium on the non-durable goods firm (recall Azeredo (2007)). This negative premium overwhelms the positive premium to the durable goods—producing firm, giving an overall market premium that is counterfactually small and is negative for some parameterizations.

[56] As such, Yogo's (2006) model also displays a form of composition risk (via the D_t/C_t ratio).

1.7.2. Agent Heterogeneity

There already exists a substantial literature that explores the magnitude of the premium in incomplete market settings where heterogeneity is captured via the experience of idiosyncratic income shocks. In this section, we focus alternatively on preference heterogeneity, its implications for the equity premium in a trading environment, and, ultimately, the usefulness of the representative agent construct. The literature is extremely small: we found only two directly relevant papers, Chan and Kogan (2002) and Chapman and Polkovnichenko (2006).[57]

Each studies an exchange equilibrium where financial markets are complete. Chan and Kogan (2002) postulate agents with ratio habit preferences of the form

$$u\left(c_t, X_t; \gamma\right) = \frac{1}{1-\gamma}\left(c_t/X_t\right)^{1-\gamma}$$

where c_t is the consumption of an individual agent with CRRA parameter value γ. The X_t term, which is their measure of the economy wide living standard, is defined as a weighted average of past aggregate endowment realizations:

$x_t = \ln(X_t)$ satisfies

$$x_t = x_0 e^{-\lambda t} + \lambda_t \int_0^t e^{-\lambda(t-s)} y_s \, \mathrm{ds}$$

where $y_t = \ln(Y_t)$, and Y_t is the period t economy wide endowment.

The latter follows a standard Brownian motion process. Note that the choice of the parameter λ determines how strongly past endowment experience influences the living standard measure.

Chan and Kogan (2002) construct a representative agent and solve for equilibrium by solving the associated planning problem. Changes in the cross sectional distribution of wealth lead to changes in the economy wide level of risk aversion in the following way. Since the risk averse agents hold most of their wealth in stocks, when endowment is low and the stock's price low, the fraction of the economy's wealth controlled by these agents diminishes, leading to increased economy wide risk aversion. The resultant countercyclical risk aversion allows their model to match nearly all of the Campbell and Cochrane (1999) results except for the volatility of the risk free rate. In Chan and Kogan (2002) it is too high relative to what is observed in the data; in Campbell and Cochrane (1999) it matches the data by construction. Chan and Kogan (2002) is important particularly for generating countercyclical risk aversion endogenously in a way that works to resolve the puzzle.

In Chapman and Polkovnichenko (2006), preference heterogeneity works against these same stated goals. They postulate a setting where investors have different rank dependent expected utility preferences (RDEU). With its surprising implications, the paper demands a somewhat detailed review. Rather than weighing various utility outcomes by their objective probabilities, RDEU preferences weight outcomes by some

[57]We thank Nick Barberis for bringing these papers to our attention.

non-linear weighting function defined over the cumulative probabilities of outcomes that have been utility ranked from worst to best. Typical parameterizations end up over weighting bad outcomes. In general, a RDEU welfare functional will have kinks at allocations where consumption outcomes are equal across any two states.

Consider an environment of $i = 1, \ldots, n$ states of nature and K agents where all agents have the same CRRA utility function over consumption outcomes in particular states,

$$U^k\left(c_i^k\right) = \frac{\left(c_i^k\right)^{1-\gamma}}{1-\gamma},$$

where c_i^k represents agent k's consumption in state i. Let $\left(C^k, P^k\right)$ denote a consumption lottery where, respectively, consumption across states is ranked from worst to best, and P denotes the vector of cumulative probabilities. More specifically, $\left(C^k, P^k\right) = \left(c_1^k \ldots, c_N^k, P_1^k, \ldots, P_N^k\right)$ where $c_i^k < c_{i+1}^k$ and $P_i^k = \sum_{j \le i} \pi_j$ where π_j is the objective probability of the state corresponding to the consumption allocation c_j^k. Note that different outcomes will have different associated vectors of cumulative probabilities. In the Chapman and Polkovnichenko (2006) formulation, the RDEU value function of agent k is given by

$$U^k\left(\{C, P\}\right) = \sum_{i=1}^{N} q_k\left(P\right) u^k\left(c_i^k\right), \tag{44}$$

where $q_k\left(P\right) = \left(P_i^k\right)^{\phi_k} - \left(P_{i-1}^k\right)^{\phi_k}$ is the weighting function. Note that if $\phi_k \equiv 1$, the expression collapses to one of expected utility, and if $\phi_k < 1$, the agent is pessimistic. In general equilibrium, the $q_k\left(P\right)$ weights will be endogenously determined, just as the consumption outcomes across states are themselves determined endogenously in equilibrium. The authors then go on to consider a two-agent, two-date equilibrium where the agents differ in their pessimism parameters, ϕ_k. With $\phi_1 = 1$ (expected utility) and $\phi_2 = 0.4$ (substantial pessimism), they find that the equity risk premium is only 17 percent of the risk premium in an economy where the consumers are identical ($\phi_1 = \phi_2 = 0.4$).

It is, however, twice the level obtained for the corresponding homogeneous preference economy, where everyone has expected utility ($\phi_1 = \phi_2 = 1$). In solving for the equilibrium, they construct a representative agent and solve accordingly. In equilibrium, REDU economies display limited risk sharing since the preference structure implies that for small risk premia, the more pessimistic agent may not choose to invest in the risky asset at all. Aggregate risk is then priced only by the less pessimistic agent. What is present is endogenous participation in the equity markets. In a dynamic equilibrium multi-agent RDEU model, we could imagine a situation of some agents entering and exiting the equity markets voluntarily as the economy passes from expansions to recessions. Would such an economy endogenously exhibit countercyclical risk aversion? Would the attractive business cycle/equity premium characteristics of, e.g., the Guvenen (2005) and Danthine and Donaldson (2002) models be preserved, despite the fact that participation in equity markets is not fixed a priori? These are reasonable issues to explore. From the perspective of this paper, the results are especially striking. In particular, Chapman and Polkovnichenko (2006) remind us that using the representative agent construct in a non-standard heterogeneous preference setting may be uninformative as it ignores important endogenous risk-sharing effects.

Before leaving this body of material, some comparative remarks might be useful. Table 10 summarizes the properties of the various preference specifications considered thus far.

We first note that a number of open questions remain. Many involve aggregation. Without this property, the results of Chapman and Polkovnichenko (2006) may recur in other contexts. The habit formation construct of Campbell and Cochrane (1999), the prospect theory of Barberis et al. (2001), and the GDA preferences of Routledge and Zin (2004) all yield effective countercyclical risk aversion, and in many respects, their results with regard to the equity premium are similar. A number of features of the GDA construct would seem to give it an edge, however. Relative to simple expected utility, it

TABLE 10
Characteristics of Principal Asset Pricing Paradigms

Attribute	Preference type								
	A Conventional CRRA	B Epstein–Zin, Chew–Deckel	C Melino–Yang	D Constantinides–Sundaresan (linear habit)	E Campbell–Cochrane	F Falato	G Zinn–Routledge	H Barberis–Huang	I RDEU
1. Does a representative agent exist?	Yes	Yes	?[i]	Yes	No	Yes	Yes	No	Yes
2. Is strict aggregation a possibility?	Yes	No	No	?	No	No	Yes	No	Yes
3. Does its use in portfolio problems lead to time consistent planning?	Yes	Yes	?	Yes	?	?	Yes	?	?
4. Do choice-theoretic axiomatic foundations exist?	Yes	Yes	No	No	No	No	Yes	No	?
5. Is counter-cyclical risk aversion a principal feature?	No	Yes	Yes	Yes	Yes	No	Yes	Yes	?

[i]The question mark indicates that the answer to the question is unknown or not fully understood.

generates the fewest additional parameters and is the only structure of the three (to date) with axiomatic foundations. Relative to the two alternatives cited above, they are the only structure that allows the separation of risk and time preferences. (More recent work by Barberis and Huang (2001) allows this as well.) Furthermore, there is no need for external scaling or an external reference point to maintain return stationarity. In short, GDA preferences have the advantage of economy and simplicity. It is unclear, however, if any of these constructs will be compatible with the stylized financial and business cycle facts—or, indeed, with generally accepted consumption processes—when placed in a production setting.

Numerous other behavioral representations are modeled in the literature. Constantinides et al. (2007), for example, argue for a period utility representation of the form

$$U(c_t) + LV(B_t),$$

where c_t denotes the period consumption and $V(B_t) = B_t^{1-\gamma_B}/(1-\gamma_B)$ the utility of bequests, B_t. The L parameter assumes a value of 0 except in the final period of the agent's life when $L = 1$ (the context is obviously one of finitely lived overlapping generations). Since B_t measures the value of bequests, it is a quantity that will be highly correlated with the return on the stock market. As a result, agents who are substantially bequest risk-averse will demand a high premium to hold equity securities in their final period of life. While this modeling device does lead to a high premium, other counterfactuals arise.

2. PRODUCTION ECONOMIES

The production-based asset pricing literature is oriented to constructing models that successfully replicate both the stylized facts of the business cycle and the financial markets. To review how the business cycle is characterized in the data and in dynamic, stochastic, general equilibrium models, see the section by Danthine, Donaldson, and Siconolfi in this volume (Danthine et al. (2008)). Relevant papers with a primary focus on the equity premium puzzle include Jermann (1998), Boldrin et al. (2001), Danthine et al. (1992), Danthine and Donaldson (2002), and Guvenen (2005). Lettau and Uhlig (2000) focus only on the relationship of habit formation to consumption volatility, although the implications of their results for the premium are straightforward.

Production models display the feature that the dividend series and the asset pricing agent's consumption series are endogenously determined within the model.[58] All the mechanisms reviewed in this section will have consequences for the statistical properties of these series. Habit formation in a production context such as Hansen (1985), in particular, will induce the model's representative agent to tailor his investment

[58] Accordingly, assets are then priced using the standard Lucas (1978) asset pricing perspective.

and labor service decisions in such a way as to reduce, dramatically, his endogenous consumption volatility vis-à-vis an otherwise identically parameterized construct without the habit feature: consumption as a pure macroeconomic quantity becomes counterfactually smooth, and the associated premium falls essentially to zero. A successful model must thus display not only a strong desire on the part of agents for consumption smoothing but also severe (but plausible) technological restrictions that inhibit their ability to do so.

Within the production literature, most models require some variation on habit formation to generate a high and countercyclical level of effective risk aversion. The complementary features that interfere with the consequent consumption smoothing are myriad. We consider Jermann (1998) as an important illustrative case.

Jermann's equilibrium allocations of consumption, investment, etc. solve

$$\max_{\{I_t\}} E\left(\sum_{t=0}^{\infty} \beta^t u\,(C_t - \alpha C_{t-1})\right),$$

subject to

$$\begin{aligned} K_{t+1} &= (1-\Omega)\,K_t + \varphi\,(I_t/K_t)\,K_t \\ C_t + I_t &\leq f\,(K_t,\ P_t N_t)\,\lambda_t \\ \lambda_t &= \rho\lambda_{t-1} + \tilde{\varepsilon}_t, \end{aligned}$$

where $f(K_t,\ P_t N_t)\,\lambda_t = K^{\theta}\,(P_t N_t)^{1-\theta}\,\lambda_t$, with

$$\theta = 0.64,\ \ \beta = 0.99,\ \ \rho = 0.99,\ \ u\,(C_t - \alpha C_t) = \frac{(C_t - \alpha C_{t-1})^{1-\gamma}}{1-\gamma},$$

with $\gamma = 5$, $\Omega = 0.025$, and $P_t = (1.005)\ P_{t-1}$; i.e., a constant growth rate of labor productivity of 0.5 percent per quarter (the calibrated time period).[59] Clearly, it is the habit formation feature most prominently that causes the agent to display a strong desire for consumption smoothing.

But what features stand in his way from doing so? First and foremost is the cost of capital adjustment function,

$$\phi\left(\frac{I_t}{K_t}\right) K_t = \frac{a_1}{1-\xi}\left(\frac{I_t}{K_t}\right)^{1-1/\xi} + a_2,$$

where $\xi = 0.23$, and a_1, a_2 are constants chosen so that the balanced growth path is invariant to ξ. Second, and less transparently, there is no ability of the agent to vary his labor supply, despite the fact that his period preference ordering manifests no disutility from doing so. The results of Jermann's (1998) exercise are presented in Table 11.

All the claims we have made in this section are manifest in Table 11. Consider column A's statistics. This corresponds to an unfettered production model: no habit and

[59]Under a growth scenario $\beta = 0.99 = \hat{\beta}\,(1.005)^{1-\gamma}$, where $\hat{\beta}$ is the subjective discount factor.

TABLE 11[i]

The Equity Premium in the Production Economy of Jermann (All parameter values as in text, except as noted; various scenarios; all return statistics in percent, annualized)

	Parameter choices					
	A	B	C	D	E	F[iii]
Er^f	4.26	3.36	4.20	3.91	0.82	1.580
SDr^f	0.62	0.76	0.59	0.61	11.46	0.460
Er^e	4.28	3.66	4.23	4.58	7.00	1.590
SDr^e	1.02	2.90	1.21	6.09	19.86	0.480
Er^p	0.02	0.26	0.03	0.67	6.18	0.001
Er^p	0.02	0.26	0.03	0.67	6.18	0.001
SDr^p	not reported					
$SD\ \Delta C/SD\ \Delta Y^{(ii)}$	0.77	0.78	0.33	1.14	0.49	

A. Standard model: $\alpha = 0$, $a_1 = a_2 = 0$ – no habit, no COA.
B. Same as case A but with $\gamma = 10$.
C. Habit formation, no COA.
D. Costs of adjustment, no habit formation.
E. Habit and adjustment costs.
F. Habit and adjustment costs but variable labor supply.

[i] Columns A–E are from Table 1 in Jermann (1998), while column F is from Boldrin et al. (2001).
[ii] This is the *SD* of the consumption growth rate relative to the *SD* of the output growth rate.
[iii] Preferences as noted in text; $\alpha = 0.9$, $\xi = 0.23$; otherwise as in Jermann (1998).

no adjustment costs, so the agent is totally free to adjust his investment to smooth out his consumption. As a result, the premium is a miserly 0.02 percent, which is even lower than its exchange economy counterpart. Column B corresponds to an otherwise identical case but for $\gamma = 10$. The premium does rise, but by an insignificant 0.24 percent, and return standard deviations remain much too low (compare this result with those in Table 1). In column C, the agent displays habit formation with a habit parameter $\alpha = 0.82$. This greatly amplified effective risk aversion induces the agent to tailor his investment plan to elicit greater consumption smoothing. As a result, the *SD* of the growth rate of consumption declines by more than one-half. Despite the added local risk aversion, the smoother consumption series draws the premium—and the return volatilities—more or less back to their column A values.

Costs of adjustment (column D) make it substantially more expensive for the agent to effect consumption smoothing via investment volatility. With this obstruction, the premium rises but still falls short even of 1 percent. When habit formation is incorporated in conjunction with COA (column E), the premium matches its historical benchmark, confirming the earlier intuition; equity return volatility is quite reasonable as well, as is the mean risk-free rate. Only the risk-free rate volatility is mismatched, with the model producing an *SD* more than twice the historical level. This is not entirely surprising given the knife-edged nature of the habit formation mechanism and the enormous swings in risk aversion it tends to generate.

Column F, from an early version of Boldrin et al. (2001), contains the results for a slightly modified Jermann (1998) model enriched by incorporating an agent whose utility is linear in leisure (risk-neutral), essentially as in Hansen (1985). Such an agent is unconcerned with variations in his labor supply; under this specification, the agent's period utility function is of the form

$$u(C_t, C_{t-1}, 1 - n_t) = \log(C_t - \alpha C_{t-1}) + Bn_t,$$

Unfortunately, this added flexibility causes the premium to collapse to a mere 0.001 percent, with the equity and risk-free standard deviations not far behind at, respectively, 0.46 and 0.48 percent. This example fully illustrates the difficulty in incorporating standard preference mechanisms into production settings. As of this writing, we are unaware of any attempts to incorporate "loss aversion" or "happiness maintenance" into a production context, but the consequences are likely to be similar.

Note that nearly all recent production models rely in some way on habit formation. Eschewing this feature, Campanale et al. (2006) explore the business cycle and agent pricing properties of a production economy very similar to Jermann (1998) (for instance, there is a cost to adjusting capital) where the representative agent possesses Epstein–Zin preferences and the certainty equivalent is obtained using GDA risk preferences and special cases thereof (expected utility and (DA), disappointment aversion preferences). Solving the model using traditional discrete state-space methodologies, they are able to replicate a very wide class of stylized financial and business cycle facts, thereby effectively challenging the habit formation device. Perhaps most strikingly, they do so for implied risk aversion that is similar to a logarithmic expected utility agent. As with the earlier habit-based literature, risk-free rate volatility remains excessive.[60] It would also be of interest to know the consequences for their results of admitting a variable labor supply. The paper is important not only for introducing these richer preference structures into a production context but also for more broadly re-energizing the asset pricing production research program.

Production-based asset pricing models are important because they afford an opportunity for cross-model verification. Since the actions of the same agents give rise to both macroeconomic and financial phenomena, a comprehensive, reliable model should be able to explain both sets of stylized facts equally well. At the current level of understanding, this is not yet the case.

3. DISASTER EVENTS AND SURVIVORSHIP BIAS

Reitz (1988) proposed a solution to the equity premium puzzle that incorporates a very small probability of a very large drop in consumption (a "disaster state"). Beyond the incorporation of this third growth state, the structure of his model is identical to that of

[60]Kallenbruner and Lochstoer (2004) demonstrate that a reasonably calibrated production model with Epstein–Zin preferences can jointly explain important macroeconomic and financial stylized facts without the excessive risk-free rate volatility characteristic of habit formulations.

Mehra and Prescott (1985). Since disaster states are ones of extremely high marginal utility of consumption, we might expect their incorporation to push up risk-free asset prices and diminish risky ones. As a result, the premium should rise. Table 12 presents a sample of scenarios drawn from Reitz (1988).

The incorporation of the disaster scenario, even with small conditional or unconditional probability, does allow for a resolution of the premium. This is not in dispute although the plausibility of such a huge drop in consumption is open to question. The model requires a 1-in-100 chance of a 25 percent decline in consumption to reconcile the equity premium with a risk aversion parameter of 10. Following on these results, Barro (2006) details all major cross-country consumption declines post-1913. While certain European countries experienced consumption declines of as much as 64 percent during World War II (e.g., Germany and Greece), even in the worst years of the Great Depression, U.S. consumption only declined by about 30 percent. In light of these figures, the percentage declines postulated by Reitz (1988) do not seem so extreme. One must assess other potential implications of the model, however. One is that the real interest rate and the probability of the occurrence of the extreme event move inversely. For example, the perceived probability of a recurrence of a depression was probably very high just after World War II and subsequently declined over time. If real interest rates rose significantly as the war years receded, that evidence would support the Rietz hypothesis. Similarly, if the low probability event precipitating the large decline in consumption were a nuclear war, the perceived probability of such an event has surely

TABLE 12
The Equity Premium and Risk-Free Rate: Possible Disaster State (All returns in percent, annualized)

Panel A[i]				
Disaster: Output falls to half its previous value				
Probability of entering	γ	β	Er_f	Er_p
Disaster				
0.0008	7.00	0.997	0.83	6.18
0.0030	5.30	0.980	0.89	6.15
Panel B[ii]				
Disaster: Output falls to three-fourths its previous value				
Probability of entering	γ	β	Er_f	Er_p
Disaster				
0.010	9.80	0.999	0.74	6.95
0.014	8.85	0.999	0.84	6.49

[i] Taken from Table 3 of Reitz (1988).
[ii] Taken from Table 6 of Reitz (1988).

varied over the last 100 years. It must have been low before 1945, the first and only year the atom bomb was used. And it must have been higher before the Cuban Missile Crisis than after it. If real interest rates had moved as predicted, that would have supported Rietz's disaster scenario. But they did not. The reader is referred to Mehra and Prescott (1988) for a detailed response to Reitz (1988) along these lines.

Barro (2006) presents a model in the same spirit as Reitz (1988) but in a context where more attention may be given to calibration and the historical record. The historical justification for his calibration is, in fact, fascinating to read. The point of departure of the paper is a postulated stochastic process for aggregate output growth that allows for rare events in a generalized Rietz setting. Specifically, he models aggregate output growth as a random walk with drift, whose innovations are of three possible types:

1. "diffusive" shocks (i.i.d. normal shocks),
2. jump shocks of "type v"—shocks that represent situations where output contracts sharply but there is no occurrence of a default on debt, and
3. jump shocks of "type w"—shocks that represent situations where output contracts sharply and a default on debt ensues.[61]

In particular, Barro assumes that the log of output $Y_t \equiv C_t$ grows according to the random walk with drift process:

$$\log Y_{t+1} = \log Y_t + \overline{g} + \tilde{u}_{t+1} + \tilde{v}_{t+1}, \tag{45}$$

where $\tilde{u}_{t+1}$ is an i.i.d random variable distributed $N(0, \sigma_u^2)$ and $\tilde{v}_{t+1}$ (jump shock "type v") captures low probability downward jumps in GDP. If a disaster occurs, with probability p, output declines by the fraction b. Very roughly speaking, a reasonable calibration requires b to be large and p small.

More formally,

$$\tilde{v}_{t+1} = \begin{cases} 0, & \text{with probability} \quad e^{-p}, \\ \log\left(1 - \tilde{b}\right), & \text{with probability} \quad 1 - e^{-p}. \end{cases} \tag{46}$$

In his calibration exercise, $\tilde{b}$ is a random variable whose probability density function coincides with the frequency distribution reflecting the size of contractions in 35 countries in the 20th century ($\tilde{b}$ ranges from 0.17 to 0.62). Barro also admits the possibility of default on government debt of proportion d, with probability q whenever a "type v" jump occurs. In other words, whenever a jump shock of "type v" occurs, there is a probability q of a "type w" jump. Table 13 presents a sample of results for the calibration exercises $\left(\tilde{b} = \tilde{d}\right)$. In all cases, preferences are as in Mehra and Prescott (1985) with CRRA = 4 and time preferences parameter $\beta = e^{-\rho} = e^{-0.03} = 0.9704$. The results presented here are intriguing and lend support to a resurrection of the disaster scenario as a viable justification for the premium.

The distinction between "type v" events and "type w" events is theoretically appealing in terms of their different implications for quantities and prices. However, we feel

[61] Barro (2006) does not explicitly define a type w shock. We introduce it here as an aid to exposition.

TABLE 13[i]
Financial Statistics: Barro's (2006) Model (All rates of return annualized in percent; all cases $\theta = 4$, $\sigma_u = 0.02$, $q = 0.4$; time preference $e^{-\rho}$, $\rho = 0.03$)

	$\bar{g} = 0.025$ $p = 0.017$	$\bar{g} = 0.025$ $p = 0.025$	$\bar{g} = 0.020$ $p = 0.017$	$\bar{g} = 0.025$ $p = 0.017$
Er^e	7.1	4.4	5.1	6.1
Er^f	3.5	−0.7	1.5	2.5
Er^p	3.6	5.2	3.6	3.6

this distinction is largely a consequence of the specific functional forms used to model the two types of rare events and the assumption that they are independent—something we intuitively think is not very plausible.

To support the observed equity premium (in the baseline model), a decline in GDP growth of 50 percent is needed. The plausibility of such a substantial decline in per capita GDP growth is justified by the figures in Barro (2006, Table 1), where this decline is shown to be 64 percent in Germany and Greece (during World War II) or 31 percent in the U.S. (during the Great Depression). These observations, however, are not at an annual frequency.[62] Looking at data at an annual frequency, we see little evidence of such a substantial decline either in GDP growth or in the growth rate of consumption. If the equity premium is estimated at an annual frequency, then the relevant growth rates should also be at annual frequencies. We argue that calibrations should be done at the same frequency as observations.

As is standard in the equity premium literature, Barro calibrates his model using output growth rates rather than consumption growth rates since in a pure exchange setting they are identical. While this distinction may not matter in a stylized economy without disaster states, the two quantities might empirically be quite different in the event of an extreme disaster. Consumption per capita will probably be a much smoother series than GDP per capita. In the event that such states occur, consumption is unlikely to remain a constant proportion of the output. Capital investment and capital utilization rates will change and act as channels to absorb the negative productivity shocks (see also Section 2), enabling consumption smoothing over time and across states. For this reason, since we do live in a production economy, it would be interesting to explore the robustness of Barro's (2006) results using cross-country data on the growth rates of per capita consumption instead of output within that more general context.

[62]In a sense the results are obtained using the 4-year drop in consumption as if it is a 1-year effect. The claimed equity premium should be compared to the observed 4-year effect (4 × 6 = 24 percent) and not the 1-year 6 percent.

A closely related literature concerns “peso problems.” This means that returns are higher than justified by current experience because investors assign a positive probability to some as-yet-unrealized disaster state. In the context of this essay, it suggests the possibility that the observed equity premium in the U.S. may result from expectations of disaster events that happen not to have materialized in the sample period of observations. Recent work by Goetzmann and Jorion (1999) supports this perspective: after concluding that the high U.S. historical premium is in some sense unique, they suggest that it may be attributable to the fact that disastrous events (most especially WWII) affecting many financial markets have largely spared the U.S. economy.

In a production setting similar to Jermann’s (1988) but with a variable labor supply, no costs of adjustment and no habit (a model that is in some sense set up with a bias against resolving the puzzle), Danthine and Donaldson (2002) replicate all first moments of returns, nearly exactly, in a peso-style model where the disaster is anticipated but never realized. The required ingredients mandate a debt-to-equity ratio of 0.3, an anticipated disaster state where output falls by 50 percent, a CRRA of 3, and a (conditional) probability of entering the disaster state of 0.008. Return volatilities are much too low, however (this is to a lesser extent true for Barro (2006) as well). A similar but more detailed exercise is undertaken by Cecchetti et al. (1993, 2000) in an exchange setting. All this is to suggest that unrealized beliefs may be critical to the puzzle’s resolution and that the results in Reitz (1988) and others can be strengthened if reported returns are generated from time series in which the actual disaster state does not appear. Veronesi (2004) goes one step further by postulating a peso-style model with representative agent learning. Dividends follow a simple linear stochastic process where the drift coefficient assumes one of two values, a normal level of $\bar{\theta}$ and a low, recession level of $\bar{\theta}$; there is a small probability of shifting from $\bar{\theta}$ to $\bar{\theta}$ and a high probability of returning. The drift coefficient is not observed directly, however, and Veronesi (2004) derives the period-by-period posterior probability of being in the normal growth state. This estimate itself evolves stochastically. Veronesi (2004) then explores the asset pricing characteristics of this economy for time paths where the low growth state is never actually realized. He derives high price sensitivity to dividend changes, increased return volatility in perceived recessions, etc. as well as significant equity premia. Unfortunately, theories built on beliefs other than pure rational expectations are difficult to test empirically.

Another attempt at resolving the puzzle proposed by Brown et al. (1995) focuses on the related notion of survival bias. The central thesis here is that the ex-post measured returns reflect the premium, in the U.S., on a stock market that has successfully weathered the vicissitudes of fluctuating financial fortunes. Many other exchanges were unsuccessful, and hence the ex-ante equity premium was low. Since it was not known a priori which exchanges would survive, for this explanation to work, stock and bond markets must be differentially influenced by a financial crisis. Governments have expropriated much of the real value of nominal debt by the mechanism of unanticipated inflation. Five historical instances come readily to mind: during the German hyperinflation, holders of bonds denominated in Reichmarks lost virtually all value invested in those assets. During the Poincaré administration in France in the 1920s, bond holders lost nearly 90 percent of the value invested in nominal debt. And in the 1980s, Mexican holders of dollar-denominated debt lost a sizable fraction of its value when the Mexican government, in a period of rapid inflation, converted the debt to pesos and limited the rate at which these funds could be withdrawn. Czarist bonds in Russia and Chinese debt holdings (subsequent to the fall of the Nationalists) suffered a similar fate under communist regimes.

The above examples demonstrate that in times of financial crisis, bonds are as likely to lose value as stocks. Although a survival bias may impact on the *levels* of both the return on equity and debt, there is no evidence to support the assertion that these crises impact *differentially* on the returns to stocks and bonds; hence, the equity premium should not necessarily be materially affected. In every instance where trading equity has been suspended, due to political upheavals, etc., governments have either reneged on their debt obligations or expropriated much of the real value of nominal debt through the mechanism of unanticipated inflation.

The difficulty that several model classes have collectively had in explaining the equity premium as a compensation for bearing risk leads us to conclude that perhaps it is *not* a "risk premium" but rather due to other factors. We consider these in the next section.

4. MARKET INCOMPLETENESS AND TRADING FRICTIONS

Market incompleteness assumes one of two different forms: (1) certain securities are not traded, and/or (2) certain individuals are for some reason excluded from financial market participation. A complete market structure where all investors can trade any security is necessary for the construction of the representative agent. In this case, the equilibrium in a heterogeneous full information economy is isomorphic in its asset pricing implications to the equilibrium of a representative agent, full information economy, if households have von Neumann–Morgenstern preferences. It is this paradigm that has provided the foundation to the essay so far. In this section we briefly review context (2), above, since the first is more than adequately considered in the essays by Constantinides and Heaton and Lucas. In our discussion of (2), we focus on restricted participation resulting from borrowing constraints imposed on certain agents.

4.1. Restricted Participation

In infinite horizon models, the effect of borrowing constraints and transaction costs is to force investors to hold an inventory of bonds (precautionary demand) to smooth their consumption. As a result, agents come close to equalizing their marginal rates of substitution with little effect on the equity premium: the economy is essentially one of complete markets.[63] Recent attempts to resolve the puzzle by incorporating both

[63]This is true unless the supply of bonds is unrealistically low. See Aiyagari and Gertler (1991).

borrowing constraints *and* consumer heterogeneity have thus focused on finite horizon OLG economies. One such example is proposed in Constantinides, Donaldson, and Mehra (2002).[64]

The authors construct an overlapping-generations (OLG) exchange economy in which consumers live for three periods. In the first period, a period of human capital acquisition, the consumer receives a relatively low endowment income. In the second period, the consumer is employed and receives wage income subject to large uncertainty. In the third period, the consumer retires and consumes the assets accumulated in the second period.

The implications of a borrowing constraint are explored by deriving and contrasting the stationary equilibria in two versions of the economy. In the *borrowing-constrained* version, the young are prohibited from borrowing or from selling equity short. This is not an artificial constraint since loans and short sales require collateral (accumulated wealth), which the young typically lack. The *borrowing-unconstrained* economy differs from the borrowing-constrained one only in that the borrowing constraint and the short-sale constraint for the young are absent.

There are two types of securities in the model, *bonds*, and *equity* with *ex-coupon* and *ex-dividend* prices q_t^b and q_t^e, respectively. Bonds are a claim to a coupon payment b every period, while the equity is a claim to the dividend stream $\{d_t\}$. The consumer born in period t receives deterministic wage income $w^0 > 0$ in period t, when young; stochastic wage income $w_{t+1}^1 > 0$ in period $t+1$, when middle-aged; and zero wage income in period $t+2$, when old. The consumer purchases $z_{t,0}^e$ shares of stock and $z_{t,0}^b$ bonds when young. The consumer adjusts these holdings to $z_{t,1}^e$ and $z_{t,1}^b$, respectively, when middle-aged. The consumer liquidates his/her entire portfolio when old. Thus, $z_{t,2}^e = 0$ and $z_{t,2}^b = 0$.

An *unconstrained* representative agent's maximization problem is formulated as follows. The agent born in period t solves

$$\max_{\left\{z_{t,i}^e,\, z_{t,i}^b\right\}} E\left[\sum_{i=0}^{2} \beta^i U(c_{t,i})\right]$$

subject to

$$\begin{aligned} c_{t,0} + q_t^e z_{t,1}^e + q_t^b z_{t,1}^b &\leq w^0, \\ c_{t,1} + q_{t+1}^e z_{t,2}^e + q_{t+1}^b z_{t,2}^b &\leq \left(q_{t+1}^e + d_{t+1}\right) z_{t,1}^e + \left(q_{t+1}^b + b\right) z_{t,1}^b + w_{t+1}^1, \\ c_{t,2} &\leq \left(q_{t+2}^e + d_{t+2}\right) z_{t,2}^e + \left(q_{t+2}^b + b\right) z_{t,2}^b, \end{aligned}$$

where $c_{t,j}$ is the consumption in period $t+j$ $(j = 0, 1, 2)$ of a consumer born in period t. When considering the borrowing constrained equilibrium, the following additional constraints are imposed: $z_{t,0}^e > 0$ and $z_{t,0}^b > 0$.

The model introduces two forms of restricted participation. First, consumers of one generation are prohibited from trading claims against their future wage income

[64]See also the companion papers by Constantinides et al. (2005, 2007) and the paper by Storesletten et al. (2008) in this volume.

with consumers of another generation.[65] Second, consumers of one generation are prohibited from trading bonds and equity with consumers of an unborn generation. As discussed earlier in Section 2, absent a complete set of contingent claims, consumer heterogeneity in the form of *uninsurable*, *persistent,* and *heteroscedastic* idiosyncratic income shocks, with *countercyclical* conditional variance, can potentially contribute to the resolution of empirical difficulties encountered by representative-consumer models.[66]

The novelty lies in incorporating a life-cycle feature to study asset pricing. As discussed earlier, the attractiveness of equity as an asset depends on the correlation between consumption and equity income. If equity pays off in states of high marginal utility of consumption, it will command a higher price (and consequently a lower rate of return) than if its payoff is in states where marginal utility is low. Since the marginal utility of consumption varies inversely with consumption itself, equity will command a high rate of return if it pays off in states when consumption is high, and vice versa.[67] In and of itself, this is the standard intuition.

A key insight of the paper is the observation that as the correlation of equity income with consumption *changes* over the life cycle of an individual, so does the attractiveness of equity as an asset. Consumption can be decomposed into the sum of wages and equity income. A young person looking forward has uncertain future wage *and* equity income; furthermore, the correlation of his equity income with his consumption will not be particularly high, as long as stock and wage income are not highly correlated. This is empirically the case, as documented by Davis and Willen (2000). Equity will thus be a hedge against fluctuations in wages and a "desirable" asset to hold as far as the young are concerned.

The same asset (equity) has a very different characteristic for the middle-aged. Their wage uncertainty has been entirely resolved: their future retirement wage income is either zero or deterministic, and the innovations (fluctuations) in their consumption occur from fluctuations in equity income. At this stage of the life cycle, equity income is highly correlated with consumption. Consumption is high when equity income is high, and equity is no longer a hedge against fluctuations in consumption; hence, for this group, it requires a higher rate of return.

The characteristics of equity as an asset therefore change, depending on who the predominant holder of the equity happens to be. Life-cycle considerations thus become crucial for asset pricing. If equity is a "desirable" asset for the marginal investor in the economy, then the observed equity premium will be low, relative to an economy where the marginal investor finds it unattractive to hold equity. The *deus ex machina* is the *stage* in the life cycle of the marginal investor.

[65] Being homogeneous within their generation, consumers have no incentive to trade claims with consumers of their own generation.

[66] See Mankiw (1986) and Constantinides and Duffie (1996).

[67] This is precisely the reason, why high-beta stocks in the simple CAPM framework have a high rate of return. In that model, the return on the market is a proxy for consumption. High-beta stocks pay off when the market return is high, i.e., when marginal utility is low; hence, their price is (relatively) low and their rate of return high.

The authors argue that the young, who should be holding equity in an economy without frictions and with complete contracting, are effectively shut out of this market because of borrowing constraints. The young are characterized by low wages; ideally, they would like to smooth lifetime consumption by borrowing against future wage income (consuming a part of the loan and investing the rest in higher-return equity). However, they are prevented from doing so because human capital alone does not collateralize major loans in modern economies for reasons of moral hazard and adverse selection.

In the presence of borrowing constraints, equity is thus exclusively priced by the middle-aged investors, since the young are effectively excluded from the equity markets, and we observe a high equity premium. If the borrowing constraint is relaxed, the young will borrow to purchase equity, thereby raising the bond yield. The increase in the bond yield induces the middle-aged to shift their portfolio holdings from equity to bonds. The increase in demand for equity by the young and the decrease in the demand for equity by the middle-aged work in opposite directions. On balance, the effect is to increase both the equity and the bond return while simultaneously shrinking the equity premium. Furthermore, the relaxation of the borrowing constraint reduces the net demand for bonds, and the risk-free rate puzzle re-emerges.[68]

In a subsequent paper, Da Silva (2006) generalizes the Constantinides et al. (2002) restricted participation model to admit habit formation where the habit coefficients of the old and middle-aged differ.[69] Using the notation of this section, the objective function of the period t-born representative agent is

$$\max_{\{z^e_{t,i},\, z^b_{t,i}\}} E\left(\sum_{i=1}^{2} \beta^i u\left(c_{t,i} - \eta_i c_{t-1}, i\right) \right), \tag{47}$$

with η_1, η_2 the habit parameters.

It follows that the MRS and the effective level of risk aversion of the middle-aged agent are, respectively,

$$\frac{\beta u_1\left(c_{t,2} - \eta_2 c_{t,1}\right)}{u_1\left(c_{t,1} - \eta_1 c_{t,0}\right) - \eta_2 \beta E\left(u_1\left(c_{t,2} - \eta_2 c_{t,1}\right)\right)}, \tag{48}$$

$$\frac{\gamma c_{t,1}}{c_{t,1} - \eta_1 c_{t,0}}. \tag{49}$$

the latter expression corresponding to the standard CRRA period utility function, $u(c) = (c^{1-\gamma}/1-\gamma)$, interacting with the habit.

It is immediately apparent from (48) that both η_1 and η_2 appear in the MRS and that, by (49), the effective CRRA is thus time-varying.

[68]Bear in mind, however, that the returns reported for this construct are annualized 20-year returns.

[69]Da Silva and Giannikos (2006) explore the same construct where the CRRA of the old agent is greater than the corresponding value for the middle-aged agent.

Generally speaking, Da Silva and Giannikos (2006) plausibly hypothesize that $\eta_1 > \eta_2$, suggesting that the elderly more readily accept consumption levels altered relative to that of their middle-aged working years than do the middle-aged relative to their youth.

The model yields surprisingly good results. In particular, for $\gamma = 2$ and $(\eta_1, \eta_2) = (0.7, 0.2)$, the premium exceeds 4 percent. This figure is associated with standard deviations that match their empirical counterparts quite well (20 and 7 percent for the equity and risk-free securities, respectively); average returns, however, are somewhat too high across the board. By (49), the effective CRRA of the middle-aged agent is 3.37 (2.48 for the old agent), figures that, per se, will not raise eyebrows. Compared to the effective risk aversion coefficient in Campbell and Cochrane (1999), these figures are astonishingly low. Generally speaking, a high equilibrium premium can result even if η_1 and η_2 are both small, a fact that is not necessarily true of infinite horizon constructs. A premium of 2.96 is obtained, for example, from a parameter combination of $\gamma = 2, \eta_1 = 0.30$ and $\eta_2 = 0.05$.[70] Da Silva's work is an example of the case where a small amount of agent heterogeneity goes a long way toward matching the data. As in representative agent habit models, risk aversion is effectively stochastic. Nevertheless, returns are stationary since the MRS depends only on consumption growth rates, which themselves are stationary series.[71] The level dependence effects noted in Danthine et al. (2004) are thus absent.

Guvenen (2005) invokes limited participation but within a traditional infinite horizon production context. In his model, shareholders can trade stocks and bonds to smooth out their own consumption, while workers can only trade risk-free bonds. Thus limited in their trading opportunities, workers can only self-insure by acquiring bonds, thereby bidding up their prices, and leading to a low risk-free rate. Although shareholders have multiple consumption smoothing devices (capital shares, bonds), in equilibrium they end up issuing bonds to workers and thus indirectly provide insurance to them. Not only does Guvenen's (2005) abstraction lead to a replication of all the basic financial stylized facts, but he is also able to demonstrate a near equivalence with the habit formation construct of Campbell and Cochrane (1999).

The restricted participation construct of Guvenen (2005) is related to the hypothesis that firms' shareholders directly provide partial income insurance to their workers in the context of a labor contract (see Danthine et al. (2008) in this volume). In either case, the equilibrium outcome is one in which the wealthier agents with greater access to consumption smoothing devices end up insuring the less wealthy, restricted-access agents. With regard to replicating the stylized facts, it does not seem important whether the insurance is provided within the firm or in the financial market. In the former case,

[70] All figures are obtained in a context where the endowment process is identical to that hypothesized in Constantinides et al. (2002).

[71] For CRRA preferences, $u(c) = c^{1-\gamma}/1-\gamma$, the MRS is

$$\beta \frac{u'(c_{t+1} - c_t)}{u'(c_t - c_{t-1})} = \beta \frac{(c_{t+1} - c_t)^{-\gamma}}{(c_t - c_{t-1})} = \left(\frac{g_{t,t+1} - 1}{1 - \left(\frac{1}{g_{t-1,t}} \right)} \right)^{-\gamma}$$

shareholder consumption is more variable as they are residual claimants: wages essentially leverage their dividend income streams. In the latter case (Guvenen (2005)), shareholders have, on net, negative bond positions: they hold leveraged portfolios and receive as a result equally high risk consumption streams.

Surely, these alternative mechanisms can be assessed on empirical grounds—Mehra and Prescott's (2007) balance sheet analysis of the U.S. economy emphasizes the latter alternative since wealthy shareholders are found to be net borrowers in the economy. The actual fact is likely to be a bit of both. Both alternatives are more appealing—based on micro-foundations—than the somewhat higher-level abstraction, e.g., of Campbell and Cochrane (1999).

In an important, very recent paper, Gomes and Michaelides (2006) go beyond Constantinides et al. (2005) by endogenizing stock market participation (there is a fixed cost to participation and only as agents become sufficiently wealthy are they willing to pay it) in a production based OLG model where agents experience idiosyncratic labor income shocks and borrowing constraints (they cannot borrow against their stochastic life cycle earnings profile). They find that the equity risk premium is relatively unaffected by the limited participation, and is almost entirely determined by incomplete risk sharing within the shareholder class. Non stock market participants are more realistically less wealthy than their counterparts in Guvenen (2005). In their baseline parameterization they derive an equlibrium premium of around 4% with a 2.5% risk free rate. The model is itself a new benchmark and reminds us how the progressive endogenization of previously fixed decisions can change our perspective on the fundamental drivers of financial phenomena. It lacks only a standard macro-style labor leisure choice, something that experience has shown will not be helpful for the task at hand. Nevertheless, Gomes and Michaelides (2006), appears to be the most comprehensive production model to date.

5. MODEL UNCERTAINTY

The original Mehra and Prescott (1985) model, as well as all the modifications considered heretofore in this essay, are rational expectations formulations. As such, the agents in the model know exactly the environment in which they operate, including precise knowledge of the stochastic process governing the evolution of output (also the dividend and consumption) and any other state variables that arise in more elaborate versions. At best, this is a very strong assumption.

In reality, the exact form of the relevant stochastic processes is not known to investors, at least with regard to distribution parameters, and it is reasonable to expect that this "model uncertainty" would itself command a premium. Three recent papers in this tradition are Mehra and Sah (2002), Barillas et al. (2006), and Weitzman (2007).

Mehra and Sah (2002) analyze the consequences of human mood fluctuations on financial markets. The consequences turn out to be surprisingly large. They address a very specific question: can small fluctuations in the subjective discount factor (β) and the risk aversion coefficient (γ) of agents result in large price movements in equity markets? They consider the type of fluctuations that occur over a short period of time rather than as secular changes in individual tastes. On a given day, for example, an individual may be in one of several possible preference states, but he is unaware of these taste fluctuations and thus does not implicitly incorporate their consequences into his optimization decisions.

In a variant of the Mehra-Prescott (1985) paradigm, they show that even changes of small magnitude in γ and β can have a significant impact on the evolution of asset prices. Their analysis assumes that taste fluctuations occur around temporally stable preferences, and it employs two simplifying assumptions. First, the growth rate of dividends is assumed to evolve according to a geometric Brownian motion. This allows them to derive a closed-form representation for the equity price in terms of the subjective parameters of interest and the drift and volatility parameters of the Brownian motion. Second, they assume that although the parameters change stochastically, the agent's demands are based on the assumption that the current realization will prevail forever. Thus, the agent, in his demand for securities, behaves myopically with respect to changes in these parameters; their analysis has been criticized on this account.

Weitzman (2007) addresses the equity premium, the risk-free rate, and the excess volatility puzzles in a pure exchange economy where the representative agent has standard CRRA preferences. The continuously compounded growth rate of consumption is assumed to be i.i.d with a normal distribution. Its mean and variance, however, are *uncertain*. While the first three specifications are standard, the last differs from the setting in Mehra and Prescott (1985), where the mean and variance are treated as known parameters that are calibrated to past sample averages.

The structural uncertainty in the mean and variance of the growth rate of aggregate consumption introduces a form of Bayesian posterior background risk that is inherited from the prior and does not converge to zero as the number of subsequent observations increases to infinity. This background risk fattens the tails of the posterior distribution of future consumption growth rates and increases significantly the value of both the equity premium and excess volatility, while simultaneously decreasing the risk-free rate (as the motivation for precautionary savings becomes relatively more important than intertemporal substitution).

In his framework, expected utility fails to exist except when preferences are logarithmic ($\gamma = 1$).[72] Weitzman circumvents this problem by assuming that the precision $w = 1/\sigma^2$ is a truncated-gamma p.d.f. with truncation parameter $\delta > 0$, which represents a lower bound for the support of the prior distribution of w. The assumption that $1/\delta > \sigma^2$ guarantees both the existence of finite expected utility and that the posterior distribution is in the same form as the prior and is subject to the same bounding constraint.

[72]See Geweke (2001).

In general, with Bayesian learning it should not matter what initial values are assigned to the parameters characterizing the priors since, as the sample size grows, they converge to the true values. In Weitzman's (2007) context, this holds true for all model parameters except δ, which controls the convergence properties of expected utility. The structural parameter δ never loses its critical impact on the subsequent behavior of agents, regardless of the amount of data accumulated over time.

One natural concern (and a potential criticism) about the model is that there is no learning about fundamentals in this setting. If δ does not decline as the sample size grows, then in what sense is the representative agent learning through time? In this model the representative agent can never precisely assess the volatility of consumption growth even if the sample size is infinity.

A second concern is that there is, as yet, no refutable hypothesis. Is the calibrated optimal value of δ a reasonable one? More importantly, how can one assess whether a value of δ is reasonable? For instance, to account for the equity premium, it might well be the case that Weitzman is using a low (reasonable) value of the risk aversion coefficient, but an unrealistic value of δ.

In Barillas et al. (2006), the agent is in the possession of an approximating model that he does not entirely trust. This latter fact manifests itself in the agent postulating a set of unspecified models—which are statistically similar, as measured by entropy—to his benchmark and which he believes will govern the data. The consequence of this model uncertainty is to allow an otherwise standard construct to satisfy the Hansen and Jagannathan (1991) bounds.

6. CONCLUDING COMMENTS

Is the resolution of the equity premium and other financial puzzles dependent on a wise choice of preference structures alone? There is ample indication, in many of the models considered in this essay, that certain behavioral style constructs can come close to such resolution: they work. In many cases, however, we do not know the axiomatic foundations underlying them or their behavior in more general settings. In that sense, we have very incomplete knowledge. Yet, the significance of the equity premium and related puzzles cannot be overstated. The consistency of neoclassical intertemporal economics would seem to rest, in large measure, on its eventual resolution.

Our sense is that the context of infinitely lived agents may not ultimately be the most fruitful setting. If one lives forever in a stationary return environment of complete markets, there is an intuitive sense that small risks should not matter. Properly measured and accounted for, they should average out over long horizons. Investors should therefore be largely indifferent to them, implying a very low premium.

For finitely lived agents nearing retirement, however, the same risks are critical: a downturn in the stock market is potentially crippling for life, for example. It is our view that a fully satisfactory resolution of the puzzle is likely to await more tractable multi-agent models with a richer demographic structure.

References

Abel, A. B. Asset prices under habit formation and catching up with the Joneses. *American Economic Review, Papers and Proceedings* 80 (1990): 38–42.

Abel, A. B. Equity premia with benchmark levels of consumption: Closed-form results, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Ait-Sahalia, Y., J. Parker, and M. Yogo. Luxury goods and the equity premium. *Journal of Finance* 59 (2004): 2959–3004.

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

Allais, M. The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School, in M. Allais, and O. Hagen, eds., *Expected Utility Hypothesis and the Allais Paradox*. D. Reidel Publishing Co., Dordrecht, Holland (1979).

Alvarez, F., and U. Jermann. Asset pricing when risk sharing is limited by default. *Econometrica* 48 (2000): 775–797.

Alvarez, F., and U. Jermann. Using asset prices to measure the cost of business cycles. *Journal of Political Economy* 112 (2004): 1223–1256.

Attanasio, O. P., J. Banks, and S. Tanner. Asset holding and consumption volatility. *Journal of Political Economy* 110 (2002): 771–92.

Attanasio, O. P., and A. Vissing-Jorgensen. Stock market participation, intertemporal substitution and risk aversion. *American Economic Review, Papers and Proceedings* (2007).

Azeredo, F. Essays on aggregate economics and finance. Doctoral dissertation, University of California, Santa Barbara (2007).

Backus, D. K., B. R. Routledge, and S. E. Zin. Exotic preferences for macroeconomists, in M. Gertler, and K. Rogoff, eds., *NBER Macroeconomics Annual 2004*. MIT Press, Cambridge, MA (2004).

Backus, D. K., and S. E. Zin. Reverse engineering the yield curve. NBER Working paper no. 4676 (1994).

Bakshi, G., and Z. Chen. Can compensation for cash flow risk and discounting risk reconcile the equity premium puzzle: A quantitative analysis, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Bansal, R. Risk compensation in equity markets, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Bansal, R., and J. W. Coleman. A monetary explanation of the equity premium, term premium and risk free rate puzzles. *Journal of Political Economy* 104 (1996): 1135–1171.

Bansal, R., and A. Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59 (2004): 1481–1509.

Barberis, N., and M. Huang. Mental accounting, loss aversion, and individual stock returns. *The Journal of Finance*, LVI(4) (2001): 1247.

Barberis, N., and M. Huang. The loss aversion/narrow framing approach to the stock market pricing and participation puzzles, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Barberis, N., M. Huang, and T. Santos. Prospect theory and asset prices. *The Quarterly Journal of Economics*, CXVI (1) (2001): 1–53.

Barrillas, F., L. Hansen, and T. Sargent. Risk-sensitive valuations of aggregate consumption fluctuations: Risk or model uncertainty? Mimeo, New York University (2006).

Barro, R. Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics* 121 (2006): 823–866.

Basak, S., and D. Cuoco. An equilibrium model with restricted stock market participation. *The Review of Financial Studies* 11 (1998): 309–341.

Benartzi, S., and R. H. Thaler. Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics* 110 (1995): 73–92.

Bewley, T. F. Thoughts on tests of the intertemporal asset pricing model. Working paper, Northwestern University (1982).

Boldrin, M., L. Christiano, and J. Fisher. Habit persistence, asset returns, and the business cycle. *American Economic Review* 91 (2001): 149–166.

Bonomo, M., and R. Garcia. Disappointment aversion as a solution to the equity premium and the risk-free rate puzzles. Working paper, Université de Montréal (1993).

Brav, A., G. M. Constantinides, and C. C. Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy* 110 (2002): 793–824.

Brav, A., and C. C. Geczy. An empirical resurrection of the simple consumption CAPM with power utility. Working paper, University of Chicago (1995).

Brown, S., W. Goetzmann, and S. Ross. Survival. *Journal of Finance* 50 (1995): 853–873.

Campanale, C., R. Castro, and G. L. Clementi. Asset pricing in a production economy with Chew-Dekel preferences. Working Paper, New York University Stern School of Business (2006).

Campbell, J. Y. Asset pricing at the millennium. *Journal of Finance* 55 (2001): 1515–1567.

Campbell, J. Y. Consumption-based asset pricing, in G. M., Constantinides, M. Harris, and R. Stulz, eds., *Handbook of the Economics of Finance*. North-Holland Elsevier, Amsterdam, (2003) pages 803–887.

Campbell, J. Y., and J. H. Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behaviour. *Journal of Political Economy* 107 (1999): 205–251.

Campbell, J. Y., and R. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1 (1988): 195–228.

Cecchetti, S. J., P. Lam, and N. C. Mark. The equity premium and the risk free rate: Matching the moments. *Journal of Monetary Economics* 31 (1993): 21–46.

Cecchetti, S. J., P. Lam, and N. C. Mark. Asset pricing with distorted beliefs: Are equity returns too good to be true? *American Economic Review* 90 (2000): 787–805.

Chan, Y. L., and L. Kogan. Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset pricing. *Journal of Political Economy* 110 (2002): 1255–1285.

Chapman, D., and V. Polkovnichenko. Heterogeneity in preferences and asset market outcomes. Working paper, Boston College (2006).

Cochrane, J. H. Where is the market going? Uncertain facts and novel theories. *Economic Perspectives* 21 (1997): 3–37.

Cochrane, J. H. Financial markets and the real economy, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Cochrane, J., and L. Hansen. Asset pricing explorations for macroeconomics, in *NBER Macroeconomics Annual* (1992) pages 115–165.

Constantinides, G. M. Intertemporal asset pricing with heterogeneous consumers and without demand aggregation. *Journal of Business* 55 (1982): 253–267.

Constantinides, G. M. Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98 (1990): 519–543.

Constantinides, G. M. Understanding the equity risk premium puzzle, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Constantinides, G. M., J. B. Donaldson, and R. Mehra. Junior can't borrow: A new perspective on the equity premium puzzle. *Quarterly Journal of Economics* 118 (2002): 269–296.

Constantinides, G. M., J. B. Donaldson, and R. Mehra. Junior must pay: Pricing the implicit put in privatizing social security. *Annals of Finance* 1 (2005): 1–34.

Constantinides, G. M., J. B. Donaldson, and R. Mehra. Junior is rich: Bequests as consumption. *Economic Theory*, 32 (2007) 125–155.

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

Chew, S. H. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica* 51(4) (1983): 1065–1092.

Chew, S. H. Axiomatic utility theories with the betweenness property. *Annals of Operations Research* 19 (1989): 273–298.

Daniel, K., and D. Marshall. The equity premium puzzle and the risk-free rate puzzle at long horizons. *Macroeconomic Dynamics* 1 (1997): 452–484.

Danthine, J.-P., and J. B. Donaldson. Labour relations and asset pricing. *Review of Economic Studies* 69 (2002): 41–64.

Danthine, J.-P., J. B. Donaldson, and R. Mehra. The equity premium and the allocation of income risk. *Journal of Economic Dynamics and Control* 16 (1992): 509–532.

Danthine, J.-P., J. B. Donaldson, C. Giannikos, and H. Guirguis. On the consequences of state dependent preferences for the pricing of financial assets. *Financial Research Letters* 1 (2004): 143–153.

Danthine, J.-P., J. B. Donaldson, and P. Siconolfi. Distribution risk and equity returns, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Da Silva, A. Habit formation in an overlapping generations model with borrowing constraints. Working paper, Columbia University Department of Economics (2006).

Da Silva, A., and C. Giannikos. Higher risk aversion in older agents: Its asset pricing implications. Working paper, Columbia University Department of Economics (2006).

DeLong, B., and K. Magin. The U.S. equity premium: Past, present and future. Working paper, U.C. Berkeley, Department of Economics (2007).

Davis, S. J., and P. Willen. Occupation-level income shocks and asset returns: Their covariance and implications for portfolio choice. CRSP working papers 523, Center for Research in Security Prices, Graduate School of Business, University of Chicago (2000).

Dekel, E. An axiomatic characterization of preferences under uncertainty. *Journal of Economic Theory* 40 (1986): 304–318.

Dimson, E., P. Marsh, and R. Staunton. Global evidence on the equity risk premium, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Epstein, L. G., and S. E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* 57(4) (1989): 937–969.

Epstein, L. G., and S. E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy* 99 (1991): 263–286.

Epstein, L. G., and S. E. Zin. The independence axiom and asset returns. *Journal of Empirical Finance* 8 (2001): 537–572.

Ferson, W. E., and G. M. Constantinides. Habit persistence and durability in aggregate consumption. *Journal of Financial Economics* 29 (1991): 199–240.

Falato, A. Happiness maintenance and asset prices. Mimeo, Columbia University (2003).

Fama, E. F., and K. R. French. Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics* 33 (1993): 3–56.

Gabaix, X., and D. Laibson. The 6D bias and the equity premium puzzle, in B. Bernanke, and K. Rogoff, eds., *NBER Macroeconomics Annual 2001*. MIT Press, Cambridge, MA (2001).

Geweke, J. A note on some limitations of CRRA utility. *Economic Letters* 71 (2001): 341–345.

Giannikos, C. Asset pricing in the presence of durable goods. Working paper, Columbia University, Department of Economics (2007).

Goetzmann, W., and P. Jorion. Global stock markets in the twentieth century. *Journal of Finance* 54 (1999): 953–980.

Goetzmann, W., and R. Ibbotson. History and the equity risk premium, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Gomes, F., and A. Michaelides. Asset pricing with limited risk sharing and heterogeneous agents. Working paper, London Business School, (2006) forthcoming *Review of Financial Studies*.

Gordon, S., and P. St-Amour. A preference regime model of bull and bear markets. *American Economic Review* 90(4) (2000): 1019–1033.

Gordon, S., and P. St-Amour. Asset returns and state-dependent preferences. CIRPEE Working Paper No. 03–16 (2003).

Gul, F. A theory of disappointment aversion. *Econometrica* 59(3) (1991): 667–686.

Guvenen, F. A parsimonious macroeconomic model for asset pricing: Habit formation of cross-sectional heterogeneity. Mimeo, University of Texas, Dept. of Economics (2005).

Guvenen, F. Reconciling conflicting evidence on the elasticity of intertemporal substitution: A macroeconomic perspective. *Journal of Monetary Economics* 53(7) (2006): 1451–1472.

Hall, R. E. Intertemporal substitution in consumption. *Journal of Political Economy* 96(2) (1988): 339–357.

Hansen, G. Indivisible labor and the business cycle. *Journal of Monetary Economics* 16 (1985): 309–327.

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

Hansen, L. P., and K. J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50 (1982): 1269–1288.

He, H., and D. M. Modest. Market frictions and consumption-based asset pricing. *Journal of Political Economy* 103 (1995): 94–117.

Heaton, J. An empirical investigation of asset pricing with temporally dependent preference specifications. *Econometrica* 66 (1995): 681–717.

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

Heaton, J. C., and D. J. Lucas. Market frictions, savings behavior and portfolio choice. *Journal of Macroeconomic Dynamics* 1 (1997): 76–101.

Heaton, J. C., and D. J. Lucas. Portfolio choice and asset prices: The importance of entrepreneurial risk. *Journal of Finance* 55 (2000).

Heaton, J. C., and D. J. Lucas. Can heterogeneity, undiversifiable risk, and trading frictions explain the equity premium? in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).

Isen, A. M. Positive affect and decision making, in M. Lewis, and J. M. Haviland, eds., *Handbook of Emotion*. The Guilford Press, New York (1989).

Jermann, U. Asset pricing in production economies. *Journal of Monetary Economics* 41 (1998): 257–275.

Johnsen, T., and J. B. Donaldson. The structure of intertemporal preferences under uncertainty and time consistent plans. *Econometrica* 53 (1985): 1451–1458.

Kahneman, D., and A. Tervsky. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (1979): 263–292.

Kallenbrunner, G., and L. Lochstoer. Long run risk through consumption smoothing. Working paper, London Business School (2006).

Kihlstrom, R. Risk aversion and the elasticity of substitution in general dynamic portfolio theory: Consistent planning by forward looking, expected utility maximising investors. Working paper, Finance Department, The Wharton School, University of Pennsylvania (2007).

King, R. G., and S. Rebelo. Public policy and economic growth: Developing neoclassical implications. RCER Working Papers 225, University of Rochester-Center for Economic Research (RCER) (1988).

Kocherlakota, N. Disentangling the coefficient of relative risk aversion from the elasticity of intertemporal substitution: An irrelevance result. *Journal of Finance* 45 (1990): 175–190.

Kocherlakota, N. The equity premium: It's still a puzzle. *Journal of Economic Literature* 34 (1996): 42–76.

Koopmans, T. C. Stationary ordinal utility and impatience. *Econometrica* 28 (1960): 287–309.

Krebs, T. Consumption-based asset pricing with incomplete markets. Working paper, Brown University (2000).

Kreps, D. M., and E. L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46 (1978): 185–200.

Lettau, M., and S. C. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 53 (2001): 815–850.

Lettau, M., and H. Uhlig. Can habit formation be reconciled with business cycle facts? *Review of Economic Dynamics* 3 (2000): 79–99.

Ljungqvist, L., and H. Uhlig. On consumption bunching under Campbell-Cochrane habit formation. SFB 649 Discussion Paper, Berlin (1999). http://sfb649.wiwi.hu-berlin.de/discussionPapers_de.php.

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

Lucas, R. E., Jr. Asset prices in an exchange economy. *Econometrica* 46 (1978): 1429–1445.

Lucas, R. E., Jr. *Models of Business Cycles*. Yrjo Jahnsson Lectures, Basil Blackwell, Oxford (1985).
Lustig, H., and S. G. van Nieuwerburgh. Housing collateral, consumption insurance and the risk premia: An empirical perspective, *Journal of Finance* 60 (2005): 1167–1219.
Luttmer, E. G. J. Asset pricing in economies with frictions. *Econometrica* 64 (1996): 1439–1467.
Malloy, C., T. Moskowitz, and A. Vissing-Jorgensen. Long run stockholders consumption risk and asset returns. Working paper (2005).
Mankiw, N. G. The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics* 17 (1986): 211–219.
Mankiw, N. G., and P. Z. Stephen. The consumption of stockholders and nonstockholders. *Journal of Financial Economics* 29 (1991): 97–112.
McGrattan, E. R., and E. C. Prescott. Taxes, regulations, and asset prices. Working paper 610, Federal Reserve Bank of Minneapolis (2001).
Mehra, R., and E. C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (1985): 145–161.
Mehra, R., and E. C. Prescott. The equity premium: A solution? *Journal of Monetary Economics* 22 (1988): 133–136.
Mehra, R., and E. C. Prescott. The equity premium: ABCs, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008a).
Mehra, R., and E. C. Prescott. Non risk based explanations of the equity premium puzzle, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008b).
Mehra, R, and E. C. Prescott. Intermediated quantities and returns. UCSB working paper. (2007).
Mehra, R., and R. Sah. Mood fluctuations, projection bias, and volatility of equity prices. *Journal of Economic Dynamics and Control* 26 (May) (2002): 869–887.
Melino, A., and A. X. Yang. State dependent preferences can explain the equity premium puzzle. *Review of Economic Dynamics* 6(2) (2003): 806–830.
Piazzesi, M., M. Schneider, and S. Tuzel. Housing, consumption and asset prices. *Journal of Financial Economics* 83 (2007): 531–569.
Pollack, R. Habit formation and dynamic demand functions. *Journal of Political Economy* 78 (1970): 745–763.
Prescott, E. C., and R. Mehra. Recursive competitive equilibrium: The case of homogeneous households. *Econometrica* 48 (1980): 1365–1379.
Quiggin, J. A theory of anticipated utility. *Journal of Economic Behavior and Organisation* 3 (1982): 323–343.
Quiggin, J. *Generalized Expected Utility Theory/The Rank-Dependent Expected Utility Model*, Kluwer-Nijhoff, Amsterdam (1993).
Ravn, M., S. Schmitt-Groh, and M. Uribe. Deep habits. CEPR Working paper no. 4269 (2004).
Reitz, T. The equity premium: A solution. *Journal of Monetary Economics* 22 (1988): 117–131.
Routledge, B., and S. Zin. Generalized disappointment aversion and asset prices. Mimeo, Carnegie-Mellon University (2004).
Storesletten, K., C. I. Telmer, and A. Yaron. Consumption and risk sharing over the lifecycle. *Journal of Monetary Economics* 51 (2004): 609–633.
Storesletten, K., C. I. Telmer, and A. Yaron. Asset pricing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics* (2007).
Storesletten, K., C. I. Telmer, and A. Yaron. Asset prices and intergenerational risk sharing: The role of idiosyncratic earnings shocks, in R. Mehra, ed., *Handbook of the Equity Risk Premium*. Elsevier, Amsterdam (2008).
Sundaresan, S. M. Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies* 2 (1989): 73–89.
Telmer, C. I. Asset-pricing puzzles and incomplete markets. *Journal of Finance* 49 (1993): 1803–1832.

Veronesi, P. The peso problem hypothesis and stock market returns. *Journal of Economic Dynamics and Control* 28 (2004).

Vissing-Jorgensen, A. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy* 110 (2002): 825–853.

Weil, P. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics* 24 (1989): 401–421.

Weitzman, M. Subjective expectations and asset return puzzles (2007). Forthcoming in *American Economic Review*.

Yogo, M. Estimating the elasticity of intertemporal substitution when instruments are weak. *The Review of Economics and Statistics* 86(3) (2004): 797–810.

Yogo, M. A consumption based explanation of expected stock returns. *Journal of Finance* 61 (2006): 539–580.