The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression?*

Timothy Cogley[†] Thomas J. Sargent[‡]

Revised: January 25, 2008

Abstract

By positing learning and a pessimistic initial prior, we build a model that disconnects a representative consumer's subjective attitudes toward risk from the high price of risk that a rational-expectations econometrician would deduce from financial market data. We follow Friedman and Schwartz (1963) in hypothesizing that the Great Depression heightened fears of economic instability. We use a robustness calculation to elicit a pessimistic prior for a representative consumer and let him update beliefs via Bayes' law. Learning eventually erases pessimism, but while it persists, pessimism contributes a volatile multiplicative component to the stochastic discount factor that a rational-expectation econometrician would detect. With sufficient initial pessimism, the model generates substantial values for the market price of risk and equity premium and predicts high Sharpe ratios and forecastable excess stock returns.

Key words: Asset pricing, learning, market price of risk, robustness.

1 Introduction

1.1 Conflicting measures of risk aversion

A risk premium depends on how much risk must be borne and how much compensation a risk-averse representative agent requires to bear it. From the Euler equation

*For useful suggestions, we thank Martin Ellison, Narayana Kocherlakota, two referees, and seminar participants at the Bank of England, Columbia, the European Central Bank, FRB New York, FRB Philadelphia, George Washington University, NYU, Universitat Autònoma de Barcelona, and Universitat Pompeu Fabra. The appendices to this paper are available online at Science Direct.

[†]University of California, Davis. Email: twcogley@ucdavis.edu

[‡]New York University and Hoover Institution. Email: ts43@nyu.edu

for excess returns and the Cauchy-Schwartz inequality, Hansen and Jagannathan (1991) deduce an upper bound on expected excess returns,

$$E(R_x) \leq \frac{\sigma(m)}{E(m)}\sigma(R_x). \tag{1}$$

Here R_x represents excess returns, m is a stochastic discount factor, and $E(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation, respectively, of a random variable. The term $\sigma(R_x)$ represents the amount of risk, and the ratio $\sigma(m)/E(m)$ is the market price of risk.

Hansen and Jagannathan characterized the equity-premium puzzle in terms of a conflict between the outcomes of two ways of measuring the market price of risk. The first way contemplates thought experiments that conclude that most people would be willing to pay only small amounts to avoid some well-understood gambles.[1] When stochastic discount factor models are calibrated to those mild levels of risk aversion, the implied price of risk is small.

The second way is to use asset market data on prices and returns along with equation (1) to estimate a lower bound on the market price of risk and a function of a risk-free rate. This can be done without imposing any model for preferences. Estimates reported by Hansen and Jagannathan and Cochrane and Hansen (1992) suggest a price of risk that is so high that it can be attained in conventional models only if agents are very risk averse. Thus, although people seem to be risk tolerant when confronting well-understood gambles, their behavior in securities markets suggests a high degree of risk aversion.

There have been several reactions to this conflict. Some economists, like Kandel and Stambaugh (1991), Cochrane (1997), Campbell and Cochrane (1999), and Tallarini (2000), dismiss the relevance of the thought experiments used to deduce low risk aversion and propose models with high risk aversion. Others put more credence in the thought experiments and introduce distorted beliefs to explain how a high price of risk can emerge in securities markets inhabited by risk-tolerant agents. This paper contributes to the second line of research.

Within the rational-expectations tradition, many asset-pricing models assume a representative consumer who knows the parameters of a Markov transition kernel that determines consumption growth. Calibration strategies for such models stress the assumption that subjective beliefs about the transition kernel coincide with the actual kernel. We ask how things would change if the representative agent did not know the parameter values for the transition kernel and instead updated beliefs using Bayes' Law. How would that affect asset prices and the market price of risk?

[1] For instance, see the Pratt calculations in Cochrane (1997, p. 17) or Ljungqvist and Sargent (2004, pp.258-260). Kocherlakota (1996, p. 52) summarizes by stating that "a vast majority of economists believe that values for [the coefficient of relative risk aversion] above ten (or, for that matter, above five) imply highly implausible behavior on the part of individuals."

1.2 Objects in play

To answer this question, we study a version of the Mehra and Prescott (1985) model that features the following objects.

1. A risk-neutral representative consumer whose intertemporal marginal rate of substitution is constant and equal to β.

2. A subjective Euler equation that determines asset prices,

$$E_t^s \beta R_{t+1} = \int \beta R_{t+1}(g_{t+1}, C_t) f(g_{t+1}|g^t, p) dg_{t+1} = 1, \tag{2}$$

where C_t represents consumption, g_t is consumption growth, and $g^t = [g_t, g_{t-1}, \ldots, g_0]$ is the history of consumption growth. This condition involves the consumer's $IMRS$, β, and subjective predictive density, $f(g_{t+1}|g^t, p)$. Because the consumer has a constant discount factor, the price of risk is zero from his point of view.

3. The consumer's one-step ahead predictive density,

$$f(g_{t+1}|g^t, p) = \int f(g_{t+1}|g_t, F) f(F|g^t, p(F)) dF, \tag{3}$$

where F is a transition matrix for g_t whose true value F_0 is unknown, $f(g_{t+1}|g_t, F)$ is the perceived transition density for g_t conditioned on a particular value for F, and $f(F|g^t, p(F))$ is the consumer's posterior over F. This posterior is formed by combining the consumer's prior $p(F)$ with the likelihood function $f(g^t|F)$,

$$f(F|g^t, p(F)) \propto f(g^t|F) p(F). \tag{4}$$

The inclusion of p on the left side highlights the impact of the consumer's initial prior.

4. Procedures for selecting the representative consumer's prior $p(F)$. We adopt two alternative procedures that represent pessimism: (1) we truncate a pre-1933 sample of consumption growth rates to oversample discouraging 1929-1933 observations; and (2) we use a robustness calculation to twist an initial prior pessimistically.

5. A rational-expectations Euler equation, i.e., one that correctly prices assets with respect to the actual transition density $f(g_{t+1}|g_t, F_0)$,

$$E_t^a m_{t+1}^* R_{t+1} = \int m_{t+1}^* R_{t+1}(g_{t+1}, C_t) f(g_{t+1}|g^t, F_0) dg_{t+1} = 1. \tag{5}$$

The pricing kernel in this equation is not the consumer's $IMRS$, but

$$m^*_{t+1} = \beta \frac{f(g_{t+1}|g^t, p)}{f(g_{t+1}|g_t, F_0)}. \qquad (6)$$

If the initial prior $p(F)$ dogmatically put probability one on $F = F_0$, the stochastic discount factor in (6) would be β and would agree with the consumer's IMRS. But if the representative consumer has a nondogmatic prior over F, the Radon-Nikodým derivative $f(g_{t+1}|g^t, p)/f(g_{t+1}|g_t, F_0)$ contributes volatility to what a rational-expectations econometrician would measure as the stochastic discount factor. The volatility of that learning wedge disconnects the representative consumer's subjective assessments of risk from prices of risk that a rational-expectations modeler deduces from financial market data.

1.3 Language

In the remainder of this paper, when we say 'rational-expectations model' or 'rational-expectations econometrician,' we refer to the standard practice of equating objective and subjective distributions that in the context of our model would equate the stochastic discount factor to β.

1.4 Motivation

Friedman and Schwartz (1963) said that the Great Depression of the 1930s created a mood of pessimism that for a long time affected markets for money and other assets:

> "The contraction after 1929 *shattered* beliefs in a 'new era,' in the likelihood of long-continued stability ... . The contraction instilled instead an *exaggerated* fear of continued economic instability, of the danger of stagnation, of the possibility of recurrent unemployment." (p. 673, emphasis added).
>
> "[T]he climate of opinion formed by the 1930s ... [was] further strengthened by much-publicized predictions of 'experts' that war's end would be followed by a major economic collapse. ...[E]xpectations of great instability enhanced the importance attached to accumulating money and other liquid assets." (p. 560).

Friedman and Schwartz attributed some otherwise puzzling movements in the velocity of money in the U.S. after World War II to the gradual working off of pessimistic views about economic stability that had been inherited from the 1930s.

> "The mildness and brevity of the 1953-54 recession must have strongly reinforced the lesson of the 1948-49 recession and reduced still further the fears of great economic instability. The sharp rise of velocity of money from 1954 to 1957 – much sharper than could be expected on cyclical grounds alone – can be regarded as a direct reflection of the growth of confidence in future economic stability. The brevity of the 1957-58 recession presumably further reinforced confidence in stability, but, clearly, each such episode in the same direction must have less and less effect, so one might suppose that by 1960 expectations were approaching a plateau. ... If this explanation should prove valid, it would have implications for assets other than money." (pp. 674-675.)

Our story also posits that the Depression shattered confidence in a 'normal' set of beliefs, making them more pessimistic in terms of their consequences for a representative consumer's utility functional, then explores how asset markets were affected as Bayes' law caused pessimism gradually to evaporate. But instead of studying velocity, we explore how pessimism and learning affect the market price of risk.

1.5 Related literature

The idea that pessimism can help explain the behavior of asset prices has been used before in quantitative studies. Some papers study the quantitative effects on asset prices of exogenously distorting agents' beliefs away from those that a rational expectations modeler would impose; e.g., see Rietz (1988), Cecchetti et al. (2000), and Abel (2002). Other papers endogenously perturb agents' beliefs. Thus, Hansen et al. (1999), Cagetti et al. (2002), Hansen et al. (2002), and Anderson et al. (2003) study representative agents who share but distrust the same model that a rational expectations modeler would impute to them. Distrust inspires agents to make robust evaluations of continuation values by twisting their beliefs pessimistically. This decision-theoretic model of agents who want robustness to model misspecification is thus one in which pessimistic beliefs are endogenous outcomes of the analysis.

All of these papers assume pessimism that is *perpetual*, in the sense that the authors do not give agents the opportunity to learn their ways out of their pessimism by updating their models as more data are observed.

In contrast, this paper assumes only transitory pessimism by allowing the representative consumer to update his model via Bayes' Law.[2] We push the representative agent's initial ideas about transition probabilities away from those that a rational expectations modeler who ignores learning might impose. We study the effects of two alterations: one that comes from endowing the representative agent in the learning

[2]Kurz and Beltratti (1997), and Kurz et al. (2004) also study models with transitory belief distortions that they restrict according to the notion of a rational-beliefs equilibrium.

model with knowledge of parameter values that is based on less data than a rational-expectations-without-learning econometrician would give him, the other from giving the representative agent a pessimistic prior that we construct by applying a risk-sensitivity operator of Hansen and Sargent (2005, 2007a). Then we give the representative consumer Bayes' Law, which via a Bayesian consistency theorem eventually erases his pessimism. We ask: How do asset prices behave in the meantime?

A number of papers also argue that learning is helpful for understanding asset prices. For instance, Barsky and DeLong (1993), Timmermann (1993, 1996), and Adam et al. (2006) study the implications of learning for stock-return volatility and predictability. In work closely related to ours, Bossaerts (2002, 2004) studies how learning alters rational-expectations pricing conditions.[3] In the spirit of the 'efficient learning market' of (Bossaerts (2002, ch. 5)), our story is about a wedge or pricing shock that Bayesian learning puts into Euler equations for pricing assets from the point of view of the probability measure that a rational-expectations econometrician imputes to his representative agent.

This paper also complements work of Hansen and Sargent (2006) that focuses on robustness in a learning context and that features another likelihood ratio:

$$\frac{\hat{f}(g_{t+1}|g^t, p)}{f(g_{t+1}|g^t, p)}, \tag{7}$$

where $\hat{f}(g_{t+1}|g^t, p)$ is a worst-case conditional density that comes from solving a robust filtering problem. When it is evaluated, not with respect to $f(g_{t+1}|g^t, F_0)$, but with respect to the predictive density $f(g_{t+1}|g^t, p)$ of an ordinary Bayesian with initial prior p who completely trusts his model, this likelihood ratio is the multiplicative contribution to the stochastic discount factor for a robust representative consumer who is learning.

Hansen and Sargent (2006) focus on alternative ways to specify how an agent's desire for decision rules that are robust to misspecification influence a worst-case predictive density $\hat{f}(g_{t+1}|g^t, p)$ and the implied contribution to his valuations of risky assets under the predictive density of someone who learns but does not fear model misspecification. In contrast, the focus in this paper is simply how a representative consumer who is learning about F values risky assets relative to one who knows that $F = F_0$. The models developed in this paper are Bayesian; robust control tools are used here only to elicit Bayesian priors.

2 The model

Following Mehra and Prescott (1985), we study an endowment economy populated by an infinitely-lived representative agent. Our consumer has time-separable,

[3]See especially Bossaerts (2002, p. 134) for a summary of his approach.

isoelastic preferences,

$$U = E_0^s \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\alpha} - 1}{1 - \alpha}, \tag{8}$$

where C_t represents consumption, β is the subjective discount factor, α is the coefficient of relative risk aversion, and E_t^s denotes the mathematical expectation with respect to the representative consumer's time t predictive density (3). We calibrate $\alpha = 0$ and $\beta = 1.023^{-1}$, so that the consumer is risk neutral and reasonably patient. To make our message transparent, we set $\alpha = 0$. Among other things, this implies that the price of risk from the consumer's point of view is always zero. Nevertheless, a high price of risk will emerge from a rational expectations Euler equation.

The consumption good arises exogenously and is nonstorable, so current-period output is consumed immediately. Realizations for gross consumption growth follow a two-state Markov process with high and low-growth states, denoted g_h and g_l, respectively. The Markov chain is governed by a transition matrix F, where $F_{ij} = \text{Prob}[g_{t+1} = j | g_t = i]$. Shares in the exogenous consumption stream are traded, and there is also a risk-free asset that promises a sure payoff of one unit of consumption in the next period. Asset markets are frictionless, and asset prices reflect expected discounted values of next period's payoffs,

$$P_t^e = E_t^s[m_{t+1}(P_{t+1}^e + C_{t+1})], \tag{9}$$

$$P_t^f = E_t^s(m_{t+1}). \tag{10}$$

The variable m_{t+1} is the consumer's intertemporal marginal rate of substitution, P_t^e is the price of the productive unit, which we identify with equities, and P_t^f is the price of the risk-free asset. Notice that we follow the Mehra-Prescott convention of equating dividends with consumption.[4] With $\alpha = 0$, the IMRS is constant and equal to β.

As already mentioned, E_t^s denotes the representative agent's subjective conditional-expectations operator. We will use the notation E_t^a to denote the expectations operator under the true transition probabilities $f(g_{t+1}|g_t, F_0)$. It is well known, however, that a rational expectations version of this model cannot explain asset returns unless α and β take on values that many economists regard as implausible.[5] Therefore, we borrow from Cecchetti et al. (2000) (CLM) the idea that distorted beliefs ($E_t^s \neq E_t^a$) may help to explain asset-price anomalies. In particular, they demonstrate that a number of puzzles can be resolved by positing pessimistic consumers who over estimate the probability of the low-growth state.

[4]An asset entitling its owner to a share of aggregate consumption is not really the same as a claim to a share of aggregate dividends, so the equity in our model is only a rough proxy for actual stocks. That is one reason why we focus more on the market price of risk.

[5]For an account of attempts to model asset markets in this way, see Kocherlakota (1996).

Our approach differs from that of CLM in one important respect. Their consumers have permanently distorted beliefs, never learning from experience that the low-growth state occurs less often than predicted. In contrast, we assume that the representative consumer uses Bayes' theorem to update estimates of transition probabilities as realizations accrue. Thus, we also incorporate the idea that learning is helpful for understanding asset prices.

A Bayesian consistency theorem holds for our model, so the representative consumer's beliefs converge to rational expectations. That means the market price of risk would eventually vanish because it is zero in the rational expectations version of the model. We study how long this takes. Our story begins in 1933 with consumers who are about to emerge from the Great Contraction. We endow them with prior beliefs that exaggerate the probability of another catastrophic depression. Then we explore how their beliefs evolve and whether their pessimism lasts long enough to explain the price of risk over a length of time comparable to our sample of post-Contraction data.

2.1 Objective Probabilities

We start with a hidden Markov model for consumption growth estimated by CLM. They posit that log consumption growth evolves according to

$$\Delta \ln C_t = \mu(S_t) + \varepsilon_t, \tag{11}$$

where S_t is an indicator variable that records whether consumption growth is high or low and ε_t is an identically and independently distributed normal random variable with mean 0 and variance σ_ε^2. Applying Hamilton's (1989) Markov switching estimator to annual per capita US consumption data covering the period 1890-1994, CLM compute the estimates in table 1.

Table 1: Maximum Likelihood Estimates of the Consumption Process

	F_{hh}	F_{ll}	μ_h	μ_l	σ_ε
Estimate	0.978	0.515	2.251	-6.785	3.127
Standard Error	0.019	0.264	0.328	1.885	0.241

Note: Reproduced from Cecchetti, et. al. (2000)

As CLM note, the high-growth state is quite persistent, and the economy spends most of its time there. Contractions are severe, with a mean decline of 6.785 percent per annum. Furthermore, because the low-growth state is moderately persistent, a run of contractions can occur with nonnegligible probability, producing something like the Great Contraction. For example, the probability that a contraction will last 4 years is 7.1 percent, and if that were to occur, the cumulative fall in consumption would amount to 25 percent. In this respect, the CLM model resembles the crash-state scenario of Rietz (1988). An advantage relative to Rietz's calibration is that the magnitude of the crash and its probability are fit to data on consumption growth.

Notice also that much uncertainty surrounds the estimated transition probabilities, especially F_{ll}, the probability that a contraction will continue. This parameter is estimated at 0.515 with a standard error of 0.264. Using a normal asymptotic approximation, a 90 percent confidence interval ranges from 0.079 to 0.951, which implies that contractions could plausibly have median durations ranging from 3 months to 13 years.Thus, even with 100 years of data, substantial model uncertainty endures. The agents in our model cope with this uncertainty.

We simplify the endowment process by suppressing the normal innovation ε_t, assuming instead that gross consumption growth follows a two-point process,

$$\begin{aligned} g_t &= 1 + \mu_h/100 \quad \text{if } S_t = 1, \\ &= 1 + \mu_l/100 \quad \text{if } S_t = 0. \end{aligned} \tag{12}$$

We retain CLM's point estimates of μ_h and μ_l as well as the transition probabilities F_{hh} and F_{ll}. We assume that this model represents the true but unknown process for consumption growth.[6]

2.2 Subjective Beliefs

To represent subjective beliefs, we assume that the representative consumer knows the two values for consumption growth, g_h and g_l, but not the transition probabilities F. Instead, he learns about the transition probabilities by applying Bayes' theorem.

The representative agent adopts a distorted beta-binomial probability model for learning about consumption growth. A binomial likelihood is a natural representation for a two-state process such as this, and a beta density is the conjugate prior for a binomial likelihood. In some of our simulations, we distort the beta prior by the T^2 risk-sensitivity operator defined by Hansen and Sargent (2007a). This induces additional pessimism by tilting prior probabilities towards the low-growth state.

We assume the agent has a prior of the form

$$p(F_{hh}, F_{ll}) \propto p(F_{hh})p(F_{ll})\zeta(F_{hh}, F_{ll}; \theta), \tag{13}$$

where the function $\zeta(F_{hh}, F_{ll}; \theta)$ is a pessimistic distortion that results from applying

[6]We do this primarily to make the learning problem tractable. Under rational expectations, suppressing the noise term has only a slight effect on asset prices (see appendix A on Science Direct). One of CLM's key assumptions is that agents can observe the state, although econometricians cannot. The noise term matters more in models in which consumers cannot observe the state because agents must solve a signal extraction problem. Brandt et al. (2004) study that case when F is known. Adding learning about F to their environment would be much more challenging.

the T^2 risk-sensitivity operator with parameter θ, and $p(F_{hh})$ and $p(F_{ll})$ are independent beta densities,

$$\begin{aligned} p(F_{hh}) &\propto {F_{hh}}^{n_0^{hh}-1}(1-F_{hh})^{n_0^{hl}-1}, \\ p(F_{ll}) &\propto {F_{ll}}^{n_0^{ll}-1}(1-F_{ll})^{n_0^{lh}-1}. \end{aligned} \tag{14}$$

The variable n_t^{ij} is a counter that records the number of transitions from state i to j through date t, and the parameters n_0^{ij} represent prior beliefs about the frequency of transitions. As θ grows large, $\zeta(\cdot,\theta)$ converges to 1, and the prior collapses to the product of beta densities. We explain below how we elicit worst-case priors. At this point, we just want to describe Bayesian updating.

The likelihood function for a batch of data, $g^t = \{g_s\}_{s=1}^t$, is proportional to the product of binomial densities,

$$f(g^t|F_{hh},F_{ll}) \propto {F_{hh}}^{(n_t^{hh}-n_0^{hh})}(1-F_{hh})^{(n_t^{hl}-n_0^{hl})}{F_{ll}}^{(n_t^{ll}-n_0^{ll})}(1-F_{ll})^{(n_t^{lh}-n_0^{lh})}, \tag{15}$$

where $(n_t^{ij}-n_0^{ij})$ is the number of transitions from state i to j observed in the sample.[7] Multiplying the likelihood by the prior delivers the posterior kernel,

$$\begin{aligned} k(F_{hh},F_{ll}|g^t) &= {F_{hh}}^{n_t^{hh}-1}(1-F_{hh})^{n_t^{hl}-1}{F_{ll}}^{n_t^{ll}-1}(1-F_{ll})^{n_t^{lh}-1}\zeta(F_{hh},F_{ll};\theta), \\ &= f(F_{hh}|g^t)f(F_{ll}|g^t)\zeta(F_{hh},F_{ll};\theta), \end{aligned} \tag{16}$$

where

$$\begin{aligned} f(F_{hh}|g^t) &\propto beta(n_t^{hh},n_t^{hl}), \\ f(F_{ll}|g^t) &\propto beta(n_t^{ll},n_t^{lh}). \end{aligned} \tag{17}$$

Hence, the prior and likelihood form a conjugate pair. The posterior is also a distorted beta density, and the counters are sufficient statistics.

To find the posterior density, we must normalize the kernel so that it integrates to 1. The normalizing constant is

$$M(g^t) = \iint f(F_{hh}|g^t)f(F_{ll}|g^t)\zeta(F_{hh},F_{ll};\theta)dF_{hh}dF_{ll}. \tag{18}$$

We lack a closed-form expression for $M(g^t)$, but it can be evaluated by quadrature. Hence, the posterior density is $f(F_{hh},F_{ll}|g^t) = k(F_{hh},F_{ll}|g^t)/M(g^t)$.

This formulation makes the updating problem manageable. Agents enter each period with a prior of the form (13). We assume that they observe the state, so to update their beliefs they just need to update the counters, incrementing by 1 the

[7] According to this notation, n_t^{ij} represents the sum of prior plus observed counters.

element n_{t+1}^{ij} that corresponds to the realizations of g_{t+1} and g_t. The updating rule can be expressed as

$$\begin{aligned} n_{t+1}^{ij} &= n_t^{ij} + 1 \quad \text{if } g_{t+1} = j \text{ and } g_t = i, \\ n_{t+1}^{ij} &= n_t^{ij} \quad \text{otherwise.} \end{aligned} \tag{19}$$

Substituting the updated counters into (16) and normalizing delivers the new posterior, which then becomes the prior for the following period.

Notice the absence of a motive for experimentation to hasten convergence. Our consumers are learning about an exogenous process that their behavior cannot affect,[8] so they engage in passive learning, waiting for 'natural experiments' to reveal the truth. The speed of learning depends on the rate at which these experiments occur. Agents learn quickly about features of the Markov chain that occur often, more slowly about features that occur infrequently.

For CLM's endowment process, that means agents learn quickly about F_{hh} as the economy spends most of its time in the high-growth state and there are many transitions from g_h to g_h. Because this is a two-state model and rows of F must sum to one, it follows that agents also learn quickly about $F_{hl} = 1 - F_{hh}$, the transition probability from the high-growth state to the contraction state. Even so, uncertainty about expansion probabilities is important for our story. In theory, the key variable is not the estimate $F_{ij}(t)$ but the ratio $F_{ij}(t)/F_{ij}$.[9] Even though $F_{hh}(t)$ moves quickly into the neighborhood of F_{hh}, uncertainty about $F_{hl}(t)/F_{hl}$ endures, simply because F_{hl} is a small number. Seemingly small changes in $F_{hl}(t)$ remain influential for a long time because a high degree of precision is needed to stabilize this ratio.

Learning about contractions is even more difficult. Contractions are rare, yet one must occur in order to update estimates of F_{ll} or $F_{lh} = 1 - F_{ll}$. Indeed, because the ergodic probability of a contraction is 0.0434,[10] a long time must pass before a large sample of contraction observations accumulates. The persistence of uncertainty about the contraction state is also important in the simulations reported below, for that also retards learning.

2.3 How Asset Prices are Determined

In this section, we put the subjective stochastic discount factor (6) and the other objects mentioned in section 1.2 to work. After updating beliefs, the representative consumer makes investment decisions and market prices are determined from no-arbitrage conditions involving Bayesian beliefs.

[8] Even if consumption were a choice variable, atomistic consumers would not experiment because actions that are decentralized and unilateral have a negligible influence on aggregate outcomes.

[9] How the ratio comes into play is explained below.

[10] A contraction is not an ordinary recession; it is more like a deep recession or a depression.

The risk-free rate is trivial. Because $\alpha = 0$, the price of a risk-free bond satisfies $P_{ft} = E_t^s m_{t+1} = \beta$, regardless of beliefs about F. Hence the risk-free rate is constant and equal to β^{-1}.

Equity returns can be calculated from the price-consumption ratio,

$$R_{t,t+1} = \frac{1+\rho_{t+1}}{\rho_t} g_{t+1}, \tag{20}$$

where $\rho_t \equiv P_t^e / C_t$. For the risky asset, arbitrage opportunities exist with respect to subjective expectations unless equation (9) is satisfied. That condition can also be expressed in terms of the price-consumption ratio as

$$\rho_t = E_t^s [m_{t+1} g_{t+1} (1 + \rho_{t+1})]. \tag{21}$$

After iterating forward, using the law of iterated expectations,[11] and imposing a no-bubbles condition, we find

$$\rho_t = E_t^s \sum\nolimits_{j=1}^{\infty} \left(\prod\nolimits_{s=1}^{j} m_{t+s} g_{t+s} \right), \tag{22}$$

where expectations are taken with respect to the Bayesian predictive density over paths for g_{t+s}. If the consumer knew the current state and the true transition probabilities, the price-consumption ratio could be calculated using a formula of Mehra and Prescott (1985), which we label $\rho(S_t, F)$. Because the transition probabilities are unknown, the price-consumption ratio is calculated by integrating $\rho(S_t, F)$ with respect to the posterior for F,[12]

$$\rho_t = \int \rho(S_t, F) f(F|g^t) dF. \tag{23}$$

Geweke (2001) and Weitzman (2007) warn that integrals such as this need not be finite because Bayesian predictive tails often have fat tails.[13] In our setting, however, convergence is guaranteed if the representative consumer discounts the future at a sufficiently high rate. Because β^{-1} is the risk-free rate, one approach to calibrating

[11] Appendix B on Science Direct proves that the law of iterated expectations holds under Bayesian updating.

[12] For a derivation, see appendix C on Science Direct.

[13] Geweke (2001) studies an endowment economy with time-separable *CRRA* preferences. For a standard Bayesian learning model for consumption growth, the expected utility integral fails to exist when $\alpha > 1/2$. Savage (1954) ensures the existence of expected utility by adopting axioms that imply that period utility is bounded (Fishburn 1970, 206-207), a condition that *CRRA* preferences violate. DeGroot (1970, ch. 7) allows for unbounded period utility but ensures existence of expected utility by restricting the family of predictive densities. Weitzman (2007) adopts DeGroot's strategy and studies implications for a variety of asset pricing puzzles. Among other things, he finds a calibration that sustains a permanently high equity premium. In contrast, our model predicts a declining equity premium.

β would be to match the *ex post* real return on Treasury bills. For the period after 1933, that comes to around 40 basis points per annum. However, a discount rate that low will blow up the price-consumption ratio because expected long-run consumption growth eventually exceeds the discount rate with positive probability. Since expected consumption growth cannot exceed maximum consumption growth, a sufficient condition for a finite price-consumption ratio is $\beta^{-1} > g_h$. In that case, $\rho(S_t, F)$ is finite for all possible values of F; hence the integral in (23) is also finite. We calibrate $\beta = 1.023^{-1}$, a value that represents a compromise between attaining a low risk-free rate and a finite price-consumption ratio.

2.4 Initial Beliefs: Bayesian Posteriors and Worst-Case Alternatives Circa 1933

All that remains is to describe how we specify the representative consumer's prior. We start by describing Bayesian posteriors conditional on data through 1933, the year our simulations begin. Then we inject different doses of pessimism by using a procedure from the robust control literature to deduce worst-case transition models relative to those posteriors.

To describe beliefs in 1933, we must take a stand on date zero. One possibility is that date zero occurred very far in the past. If the consumer had been updating beliefs for a very long time prior to 1933, he would have converged to the neighborhood of rational expectations by then, and our model would add nothing to Mehra and Prescott (1985). In particular, the market price of risk and equity premium would both be nil.

The assumption that date zero occurred very far in the past is reasonable for a stationary environment but less plausible when there are structural breaks.[14] In this paper, we entertain a vision of a punctuated learning equilibrium. We imagine long spans of time during which the consumer updates beliefs in the conventional Bayesian manner. But we also imagine that infrequent traumatic events occasionally shatter the agent's confidence, causing him to re-evaluate his model and re-set his prior. These recurrent rare events arrest convergence to a rational expectations equilibrium and initialize a new learning process. Following Friedman and Schwartz, we posit that the Great Contraction was one such event. We analyze asset prices in its aftermath.

After a traumatic shock, we assume the consumer reformulates beliefs by discounting data from the distant past and putting more weight on recent observations. Exactly how this is done is hard to say, so we investigate alternative scenarios.

[14]In fact, because sustained growth is a relatively recent phenomenon, a large sample of observations of growing consumption would not have been available in 1933.

2.4.1 Pessimism induced by a short sample ending in 1933

In our benchmark model, we assume the consumer re-estimates a Bayesian posterior in 1933 using a short training sample. We consider two training samples, one running from 1890-1933 and the other from 1919-1933. The first conditions on all the consumption data we possess, and the second begins after World War I.

The priors at the beginning of the respective training samples are recorded in Table 2. Although we simulate a two-point Markov process, we must estimate a hidden Markov model when processing actual data. For our purposes, the key parameters are the transition probabilities. Briefly, for μ_h, μ_l, and σ_ε^2, the prior is weakly informative and centered on CLM's estimates. For the key parameters F_{hh} and F_{ll}, the prior is proper but close to being uniform on $(0, 1)$. The prior mean and standard deviation for F_{hh} and F_{ll} are 0.5 and 0.289, respectively.

Table 2: Priors at the Beginning of the Training Sample

$\mu_h \sim N(0.02251, 0.005^2)$
$\mu_l \sim N(-0.06785, 0.015^2)$
$\sigma_\varepsilon^2 \sim IG(0.03^2, 4)$
$F_{hh} \sim beta(1, 1)$
$F_{ll} \sim beta(1, 1)$

Next, we fast forward to 1933, assuming that the consumer has updated beliefs using Bayes theorem.[15] Tables 3 and 4 summarize his posterior, which becomes the prior for 1934. In most cases, the posterior mean is within one standard deviation of CLM's estimates. The most notable difference concerns F_{ll}, which is higher than in CLM's sample. Consumption growth was sharply negative during 1930-1933; with a short training sample, this experience would have made a Bayesian pessimistic about contraction-state persistence. In addition, for the 1919-1933 training sample, F_{hh} is lower than CLM's estimate. In this case, the consumer is pessimistic about both the onset and the persistence of contractions.

Table 3: Posterior Means and Standard Deviations in 1933

	F_{hh}	F_{ll}	μ_h	μ_l	σ_ε
1890-1933	0.977	0.693	2.03	-6.40	4.34
	(0.023)	(0.269)	(0.41)	(1.45)	(0.50)
1919-1933	0.915	0.805	2.32	-5.98	3.54
	(0.081)	(0.191)	(0.48)	(1.25)	(0.85)

[15] We simulate the posterior using a Markov Chain Monte Carlo algorithm described in Kim and Nelson (1999, ch. 9).

Table 4 describes the implied distribution for the ergodic mean of consumption growth, which we label $\mu_{\Delta c}$. We summarize this distribution by reporting the median estimate of $\mu_{\Delta c}$ as well a centered 95-percent credible set.[16] Pessimism about the contraction state reduces $\mu_{\Delta c}$ relative to the DGP. While CLM's model implies $\mu_{\Delta c} =$ 1.86 percent, the median Bayesian estimate is 1.43 and -0.13 percent, respectively, for the two training samples. In addition, uncertainty about the transition probabilities creates fat tails. Furthermore, since contractions are rare, F_{ll} is more uncertain, and the lower tail is quite a bit fatter than the upper tail. Not only does the predictive density put a lot of probability on negative values for $\mu_{\Delta c}$, the credible sets also include values that are large in magnitude, ranging down to -4.06 and -5.71 percent, respectively. Because contractions are infrequent, one cannot easily rule out high values for F_{ll}, and this leaves open the possibility that the economy could spiral downward. The possibility of a downward spiral accounts for the long lower tail.

Table 4: Bayesian Credible Sets for Mean Consumption Growth

Percentiles	0.025	0.5	0.975
1890-1933	-4.06	1.43	2.59
1919-1933	-5.71	-0.13	2.53

To represent the benchmark prior for our simulation model, we adopt specification (13), but set $\theta = +\infty$ to eliminate the distortion (i.e., so that $\zeta(\cdot, \theta) = 1$). Then we calibrate beta densities for F_{hh} and F_{ll} so that they have the same means and degrees of freedom as in table 3. We refer to this as our beta prior.

2.4.2 Using robustness to add initial pessimism

The preceding models of beliefs instill some initial pessimism to a representative consumer relative to what a typical rational expectations econometrician who ignores a learning process would impute to him. As we demonstrate below, these models make modest progress toward explaining the price of risk and equity premium. To obtain better quantitative results, we need to inject an additional dose of pessimism. We do that by using techniques from the robust control literature to distort the beta priors. The details are given in appendix D on Science Direct. Here we provide some intuition and describe an alternative pair of worst-case priors.

The undistorted beta prior envisions a structural break. After the Great Contraction, the consumer suspects a shift in the transition probabilities, discounts data from the old regime, and re-initializes his prior in order to expedite learning about the new regime. But the consumer retains complete confidence in his model specification. In the alternative worst-case models, we assume the consumer not only suspects a

[16]I.e., we cut off 2.5 percent of the draws in the upper and lower tails.

shift in the parameters but also begins to doubt that a beta probability model is correctly specified. To guard against model misspecification, a robust consumer makes an additional risk adjustment that tilts the beta prior. For our two-state endowment process, that makes the consumer more pessimistic about the onset and persistence of contractions.

The literature on robust control emphasizes that the degree of robustness must be restrained in order to keep a risk-sensitive objective function well defined.[17] Proceeding in the spirit of Anderson et al. (2003) and Hansen et al. (2002), we restrain the degree of robustness by insisting that the worst-case alternative is statistically difficult to distinguish from the benchmark model given data in a training sample. We do this by calculating the Bayes' factor for the beta and worst-case representations. This is defined as the ratio of the probability of the training sample viewed through the lens of the beta model relative to the probability viewed through the lens of the worst-case alternative:

$$B = \frac{f(g^{training}|beta)}{f(g^{training}|worst\ case)}. \tag{24}$$

According to Bayesian conventions, $0 < 2\log B < 2$ is viewed as weak evidence against the worst-case alternative, but 'barely worth mentioning.' Values of $2\log B$ between 2 and 5 are regarded as moderate evidence, values between 5 and 10 are considered strong evidence, and values greater than 10 are interpreted as decisive evidence (Raftery (1996)). We consider specifications at the upper edge of the 'moderate' and 'strong' regions, respectively. In this way, we rule out worst-case scenarios in which the representative consumer guards against specification errors that he could decisively dismiss based on outcomes in the training sample.

In what follows, the worst-case models are distortions of beta densities as in equation (13), where the beta piece is the same as in the benchmark prior and the distortion is calculated to deliver values of $2\log B$ equal to 5 and 10, respectively. We do this by varying θ. The magnitude of the distortion – and hence the Bayes factor – depend on a single free parameter θ that we adjust to deliver the desired value for the Bayes factor.

Table 5 records the mean of the worst-case transition probabilities. Relative to the beta prior, the worst-case priors underestimate F_{hh} and exaggerate F_{ll}. In other words, the robust consumer initially believes that contractions occur more often and are longer when they do occur. Since long contractions have the character of Great Depressions, our robust consumer is initially wary of another crash.

[17] See Hansen and Sargent (2007b) for a discussion of breakdown values for the risk-sensitivity parameter θ.

Table 5: Worst-Case Priors

$2\log B = 5$	F_{hh}^{WC}	F_{ll}^{WC}	$\mu_{\Delta c}$	$E(E_t \Delta c_{t+1})$
1890-1933	0.956	0.907	-1.65	1.36
1919-1933	0.866	0.915	-3.77	0.98
$2\log B = 10$				
1890-1933	0.955	0.971	-3.95	1.00
1919-1933	0.854	0.952	-4.95	0.97

Note: B is the Bayes factor for the beta and worst-case models.

In column 4, we also record the consumer's worst-case prior for the ergodic mean of consumption growth. Because he exaggerates the probability of contractions, $\mu_{\Delta c}$ is negative and large in magnitude. Notice, however, that the worst-case value for $\mu_{\Delta c}$ always remains within the Bayesian credible set. Although the robustness calculations drive us toward the lower portion of the Bayesian credible set, they do not drive us outside. The worst-case scenarios resemble parameterizations contained within the credible set. This is just another way to say that the worst-case scenarios are statistically hard to distinguish from the reference model.

Another way to assess the degree of pessimism is to calculate the mean one-step ahead forecast for consumption growth. In a time-invariant probability model, this would be identical to the ergodic mean, but the two can differ along a learning transition. Column 5 records the average one-step ahead forecast along each transition path. The numbers range from 0.97 to 1.36 percent per annum. Although these are substantially lower than the true mean of 1.86 percent, they are much less pessimistic than the prior for the ergodic mean. Along a learning path, the consumer worries a lot about things that scarcely ever happen. He fears that disaster looms just over the horizon, but those disasters rarely materialize.

3 Simulation Results

We simulate asset returns by drawing 1000 paths for consumption growth from the true Markov chain governed by F_0. To imitate the period following the Great Contraction, we initialize each trajectory in the low-growth state and then simulate 70 years of consumption. The consumer is endowed with one of the priors described above and applies Bayes' law to each consumption-growth sequence. Prices that induce the consumer to hold the two securities follow from the subjective Euler equations.

3.1 Market Prices of Risk

Hansen and Jagannathan calculate a market price of risk in two ways. The first, which we label the 'required' price of risk, is inferred from security market data without reference to a model discount factor. According to equation (1), the price of risk must be as least as large as the Sharpe ratio for excess stock returns

$$\frac{\sigma(m_t)}{E(m_t)} \geq \frac{E(R_{xt})}{\sigma(R_{xt})}. \tag{25}$$

Thus, the Sharpe ratio represents a lower bound that a model discount factor must satisfy in order to reconcile asset returns with an *ex post*, rational expectations Euler equation. Hansen and Jagannathan find that the required price of risk is quite large, on the order of 0.23.Table 6 reproduces estimates using Robert Shiller's annual data series for stock and bond returns.[18]

Table 6: The Mean, Standard Deviation, and Sharpe Ratio for Excess Returns

	1872-2003	1872-1928	1934-2003	1934-1968	1969-2003
$E(R_{xt})$	0.039	0.027	0.068	0.099	0.036
$\sigma(R_{xt})$	0.174	0.151	0.166	0.177	0.151
$E(R_{xt})/\sigma(R_{xt})$	0.224	0.179	0.409	0.559	0.238

Note: Estimates are based on Robert Shiller's annual data.

Shiller's sample runs from 1872 to 2003, and for that period excess stock returns averaged 3.9 percent per annum with a standard deviation of 17.4 percent, implying a Sharpe ratio of 0.224. Before the Depression, however, the unconditional equity premium and Sharpe ratio were both lower. For the period 1872-1928, the mean excess return was 2.7 percent, the standard deviation was 15.1 percent, and the Sharpe ratio was 0.179. In contrast, after the Great Contraction the equity premium and Sharpe ratio were 6.8 percent and 0.409, respectively. Furthermore, if the post-Contraction period is split into halves, we find that the equity premium and Sharpe ratio were higher in the first half, at 9.9 percent and 0.559, and lower in the second, at 3.6 percent and 0.238. Thus, table 6 points to a high but declining market price of risk after the Great Contraction.[19]

Hansen and Jagannathan also compute a second price of risk from discount factor models in order to check whether the lower bound is satisfied. They do this by substituting consumption data into a calibrated model discount factor and then computing its mean and standard deviation. For model prices of risk to approach the required price of risk, the degree of risk aversion usually has to be set very high. When it is set at more 'plausible' values, i.e., ones consistent with the though experiment mentioned above, the model price of risk is quite small, often closer to 0.02 than to 0.2.

[18]The data can be downloaded from http://www.econ.yale.edu/~shiller/data.htm

[19]For formal evidence on the declining equity premium, see Blanchard (1993), Jagannathan et al. (2000), Fama and French (2002), and DeSantis (2004).

Thus, the degree of risk aversion needed to explain security market data is higher than values that seem reasonable a priori. That conflict is evident in the rational expectations version of our model. Because the consumer's IMRS is constant, the model price of risk under rational expectations is zero.

Recall the objects set forth in section 1.2. In a rational expectations model, there is a unique model price of risk because subjective beliefs coincide with the actual law of motion. But that is not the case in a learning economy. In our model, subjective beliefs eventually converge to the actual law of motion, but they differ along the transition path, so when we speak of a model price of risk we must specify the probability measure with respect to which moments are evaluated. At least two prices of risk are conceivably relevant in a learning economy, depending on the probability measure that is used to evaluate the mean and standard deviation of m_t.

If we were to ask the representative consumer about the price of risk, his response would reflect his beliefs. We call this the 'subjective' price of risk,

$$PR_t^s = \frac{\sigma_t^s(m_{t+1})}{E_t^s(m_{t+1})}. \tag{26}$$

Here a superscript s indicates that moments are evaluated using the representative consumer's subjective probabilities encoded in the time t predictive density (3). Because $m_t = \beta$ is constant, the subjective price of risk is always zero irrespective of the growth state or the agent's beliefs.

Next, we imitate Hansen and Jagannathan by seeking the market price of risk needed to reconcile equilibrium returns with rational expectations E^a, by which we mean expectations with respect to the actual transition density $f(g_{t+1}|g_t, F_0)$ from section 1.2. In the learning economy, the price-consumption ratio satisfies the subjective no-arbitrage condition

$$\rho_t = E_t^s[m_{t,t+1} g_{t+1}(1 + \rho_{t+1})]. \tag{27}$$

In contrast, under rational expectations (i.e., taking the mathematical expectation E^a with respect to the actual transition law $f(g_{t+1}|g_t, F_0)$), arbitrage opportunities exist unless

$$\rho_t = E_t^a[m^*_{t,t+1} g_{t+1}(1 + \rho_{t+1})], \tag{28}$$

for the non-negative stochastic discount factor mentioned in section 1.2, namely,[20]

$$m^*_{t,t+1} \equiv m_{t,t+1} \cdot \frac{f(g_{t+1}|g^t, p)}{f(g_{t+1}|g^t, F_0)}. \tag{29}$$

Thus, the rational expectations discount factor $m^*_{t,t+1}$ consists of the product of the consumer's IMRS and the Radon-Nikodým derivative of the subjective transition

[20] Also see appendix E on Science Direct.

probabilities with respect to the actual transition probabilities. In what follows, we label the extra term RN_{t+1}.

Equation (28) is a rational expectations no-arbitrage condition that explains asset prices in the learning economy. Therefore, the price of risk that reconciles returns with rational expectations is

$$PR_t^{RE} = \frac{\sigma_t^a(m_{t+1}^*)}{E_t^a(m_{t+1}^*)}. \tag{30}$$

In a learning economy, subjective and rational expectations prices of risk can differ. The existence of two different prices of risk is the key to our resolution of the price-of-risk paradox. In our model, the consumer's IMRS is constant and the subjective price of risk is zero. This accords roughly with thought experiments and surveys. But in our simulations, RE prices of risk are often large, reflecting the change of measure needed to reconcile returns from a learning economy with rational expectations.

Notice that the definition of $m_{t,t+1}^*$ implies $\sigma_t^a(m_{t,t+1}^*) = \beta\sigma_t^a(RN_{t+1})$. In addition, the bond-pricing condition implies $E_t^a(m_{t,t+1}^*) = \beta$. Therefore, the price of risk that would be measured by a rational expectations econometrician who ignores learning reduces to

$$PR_t^{RE} = \sigma_t^a(RN_{t+1}), \tag{31}$$

the conditional standard deviation of the Radon-Nikodým derivative. In appendix F on Science Direct, we show that RN_{t+1} is the ratio of the posterior mean $F_{ij}(t)$ to the true transition probability,

$$RN_{t+1}(i,j) = \frac{F_{ij}(t)}{F_{ij}}, \tag{32}$$

The appendix also shows that its conditional standard deviation is

$$\sigma_t^a(RN_{t+1}|g_t = i) = \left[F_{ih}\left(\frac{F_{ih}(t)}{F_{ih}}\right)^2 + F_{il}\left(\frac{F_{il}(t)}{F_{il}}\right)^2 - 1\right]^{1/2}. \tag{33}$$

Hence, disagreement between actual and subjective probabilities contributes to a high rational expectations price of risk. Notice that parameter uncertainty per se does not, for $\sigma_t^a(RN_{t+1})$ would be zero if priors were centered on the true parameters. Pessimism plays an essential role in this model.[21]

Notice also that the model predicts a declining rational expectations price of risk. As observations accrue, the estimates $F_{ij}(t)$ move toward the true values F_{ij}, and subjective beliefs accord more closely with actual probabilities. In the limit,

[21] In a closely related model, Azeredo and Shah (2007) demonstrate that this also holds for the equity premium.

disagreement vanquishes, and the rational expectations price of risk converges to zero.

These calculations deliver conditional prices of risk. In contrast, table 6 reports bounds on the unconditional price of risk under rational expectations. A final adjustment is needed to reconcile the two. By 'unconditional,' we mean that the mean and standard deviation of $m^*_{t,t+1}$ do not depend on the growth state at date t. To marginalize with respect to the growth state, we take a weighted average of the conditional risk prices using the ergodic probabilities as weights,

$$MPR_t = \sigma_t^a(RN_{t+1}|g_t = g_h) \cdot F_h + \sigma_t^a(RN_{t+1}|g_t = g_l) \cdot F_l. \tag{34}$$

MPR_t is the unconditional market price of risk under rational expectations.

Figure 1 portrays simulations of MPR_t from our learning economy. The panels depict simulations initialized with different priors. The columns refer to the two training samples, and the rows correspond to beta and worst-case priors. The prior becomes more diffuse as we move from left to right, and the degree of initial pessimism increases as move down the rows. In each panel, the solid curve illustrates the unconditional rational expectations price of risk, and the dotted and dashed curves represent the rational expectations price of risk in expansions and contractions, respectively. Each line represents the cross-sectional average of 1000 sample paths in a given year.

The upper left panel depicts results for a beta prior conditioned on the training sample 1890-1933. For this prior, MPR is about 0.05 to 0.07, well short of the Hansen-Jagannathan bound. In this scenario, the consumer is initially pessimistic about the persistence of contractions, but evidently more disagreement is needed to attain a high MPR.

One way to make the consumer more pessimistic is to shorten the training sample. The upper right panel portrays results for a beta prior conditioned on data for 1919-1933. As shown in table 3, shortening the training sample makes the consumer more pessimistic both about the onset and persistence of contractions. Since there is more initial disagreement, the unconditional MPR is higher, ranging from 0.3 to 0.4 in the first decade of the simulation. But with a short training sample, the prior is also diffuse, and expansion-state pessimism evaporates quickly, causing the MPR to fall. After 70 years, the RE price of risk dwindles to 0.13. This represents progress, but we would like a high MPR to last longer.

Another way to increase initial pessimism is to elicit worst-case priors using the robustness correction described above. The second row illustrates results for priors calibrated so that $2\log B = 5$. For the longer training sample, the unconditional MPR starts around 0.23 and declines gradually to 0.18. For the shorter training sample, the MPR exceeds 0.65 for the first decade, falls to 0.4 after 35 years, and remains above 0.3 at the end of the simulation. Higher prices of risk are obtained in simulations with priors calibrated so that $2\log B = 10$, shown in the third row. For the 1890-1933 training sample, the unconditional MPR starts around 0.28 and falls slightly to 0.25 in year 70. For the 1919-1933 training sample, the MPR remains above or close to 0.5 for the first half of the simulations and then declines slowly to 0.36 at the end. Thus, even after 70 years, the rational expectations price of risk is quite a bit larger than the subjective price of risk, and in many cases it exceeds Hansen and Jagannathan's bound.

Next, we explore whether uncertainty about F_{hh} or F_{ll} contributes more to the unconditional MPR. According to equation (34), MPR is a weighted average of the conditional RE prices of risk. According to equation (33), the expansion- and contraction-state prices of risk depend, respectively, on uncertainty about F_{hh} and F_{ll}. Figure 1 shows how RE prices of risk vary across expansions and contractions. From that we can infer how uncertainty about the respective transition probabilities influences the MPR.

The RE price of risk is substantially higher in contractions, and the contraction price of risk falls more slowly. The persistence of the contraction value follows from the fact that the representative consumer learns more slowly about the contraction-state transition probabilities F_{lj}. Contractions are observed less often, so the consumer has fewer opportunities to learn about them. Although the contraction-state price of risk is higher and more persistent, the unconditional price of risk more closely resembles the expansion-state price. This reflects the unequal weights attached to expansion- and contraction-state values when forming unconditional moments. The conditional moments in (34) are weighted by the ergodic probabilities F_h and F_l, so expansion-state moments get a weight roughly 20 times that of the contraction-state values. Thus, the expansion-state price of risk is more influential for the unconditional price of risk.

To summarize, we explain the disconnect between subjective assessments of well understood risks and the price of risk deduced from financial market data by making the risk less well understood, thereby inserting a learning wedge into a rational expectations Euler equation. In our model, the subjective price of risk is zero because the consumer's IMRS is constant. Nevertheless, the price of risk needed to reconcile returns with rational expectations is often quite large. The modified discount factor m^*_{t+1} that appears in a rational expectations Euler equation is the product of the consumer's IMRS and a Radon-Nikodým derivative that encodes the effects of learning. The Radon-Nikodým derivative can be highly volatile along a learning path. In our simulations, we get large rational expectations prices of risk when the consumer's prior is based on a short training sample and involves a moderate concern for robustness.

3.2 The Risk-free Rate and the Equity Premium

Thus far we have concentrated on model prices of risk. Now we turn to the properties of asset returns. As noted above, our model implies a constant risk free rate equal to $\beta^{-1} - 1 = 0.023$. Although the learning wedge amplifies the volatility of the stochastic discount factor, it does not make the risk-free rate excessively volatile. As shown in appendix F, the conditional mean of the Radon-Nikodým derivative is 1.[22]

Next we turn to the equity premium. Table 7 reports the average excess return across 1000 sample paths. Results are reported for the full 70-year simulation and also for the first and second halves. The top row reproduces estimates from actual data.

Table 7: Mean Excess Returns

Data: 1934-2003	Full Sample	First Half	Second Half
	0.0680	**0.0990**	**0.0360**
	Prior: 1890-1933		
	Full Sample	First Half	Second Half
Beta	0.0090	0.0153	0.0027
$2 \log B = 5$	0.0317	0.0485	0.0148
$2 \log B = 10$	0.0515	0.0747	0.0283
	Prior: 1919-1933		
	Full Sample	First Half	Second Half
Beta	0.0267	0.0425	0.0109
$2 \log B = 5$	0.0421	0.0566	0.0277
$2 \log B = 10$	**0.0511**	**0.0654**	**0.0369**

In the data, the equity premium averaged 9.9 percent in the first half of the sample and declined to 3.6 percent in the second. The average for the full sample was 6.8 percent. When the representative consumer's prior is sufficiently pessimistic, the model approximates reasonably well both the level and decline in the equity premium. For instance, for the 1919-1933 training sample and $2 \log B = 10$ (highlighted in boldface), the average excess return was 6.54 percent in the first half of the simulation, 3.69 in the second, and 5.11 percent over the full sample. Although the model understates the premium in the first half, it is right on target for the second, and it explains roughly three-quarters of the observed premium in the full sample.

[22]The price of long-term real discount bonds is also constant at β. Hence the model implies a flat real term structure. To make contact with the nominal term structure, we would have to say something about inflation and how agents learned about changing inflation dynamics after World War II. We prefer not to address that here in order to retain the simplicity of Mehra and Prescott's framework.

The model is less successful for lower degrees of initial pessimism. For example, for the 1919-1933 prior, the equity premium falls from 5.11 to 4.21 percent as $2 \log B$ declines from 10 to 5. The equity premium declines further to 2.67 percent when a concern for robustness is withdrawn and returns are simulated for the undistorted beta prior. Similar patterns are evident in the subsamples.

Comparable results emerge when priors are elicited from the 1890-1933 training sample. Indeed, the model equity premium is actually a bit higher in this case. However, as we explain below, this model is less successful explaining Sharpe ratios. That is why we emphasize results for the 1919-1933 training sample.

Thus, to approximate the equity premium, our model must assume that the Depression marked a severe break with the past, shattering the representative consumer's beliefs about the likelihood of expansions and contractions. After a sharp structural break, it is plausible that agents would radically revise beliefs. That they would adopt a diffuse prior is also credible, because they would have had little previous experience in the new environment. It is also at least conceivable that they would initially guard against model uncertainty by forming a robust prior. But whether the Depression shattered beliefs to the extent envisioned in our most successful model is hard to say.

Since the consumer is risk neutral and the subjective price of risk is zero, it should be clear that the average excess return does not represent compensation for bearing risk. On the contrary, it is due to forecast errors that are systematic with respect to the rational expectations probability model. Ex ante, the consumer always expects equity returns to equal the risk-free rate. Ex post, a high average excess return emerges because the consumer consistently underestimates the return on equities.

High unexpected equity returns stem from two sources. First, because the consumer has pessimistic priors, he routinely underestimates consumption (dividend) growth one step ahead. Thus, the dividend yield is consistently higher than anticipated. Second, equities are initially undervalued relative to the true transition probabilities because the consumer over-estimates the likelihood of another Depression. The consumer becomes less pessimistic as he learns, and the equity price rises as pessimism evaporates, correcting the initial undervaluation. Thus, along the transition path, the consumer also consistently reaps unanticipated capital gains.

Of the two, the revaluation effect dominates. A quick calculation shows why the contribution of unanticipated dividend yields is small. The innovation in the dividend yield can be expressed as

$$\frac{C_{t+1}}{P_t^e} - E_t^s\left(\frac{C_{t+1}}{P_t^e}\right) = \frac{C_t g_{t+1}}{P_t^e} - E_t^s\left(\frac{C_t g_{t+1}}{P_t^e}\right) = \frac{g_{t+1} - E_t^s g_{t+1}}{\rho_t}. \tag{35}$$

For our most successful scenario, the mean one-step forecast error for consumption growth is 89 basis points per annum.[23] In view of the fact that mean consumption

[23] Recall that the average one-step ahead forecast need not coincide with the unconditional mean along a learning path.

growth is 1.86 percent, that is a substantial downward bias. To calculate its contribution to unexpected returns, we must divide by the price-consumption ratio, and that reduces its contribution substantially. Thus, the direct contribution of dividend forecast errors is quite small. The high unexpected return mostly reflects unanticipated capital gains.

Of course, the two terms are not independent; innovations in dividend growth also drive learning and hence revaluations. Since there is really only one force at work, one could also say that unexpected dividend growth accounts for the entire equity premium. But the direct channel through dividend yields is small relative to the indirect channel through capital gains.

The revaluation effect is illustrated in figure 2, which portrays the simulation average of the price-consumption ratio for worst-case priors constructed from the 1919-1933 training sample. For our preferred scenario, the price-consumption ratio starts around 15 and increases gradually to 40 as pessimism evaporates. The overall increase is broadly consistent with Shiller's data on the price-dividend ratio, which rises from the mid teens in the 1930s to around 50 in 2003. However, the model's timing is off. In the model, the price-consumption ratio increases steadily along the learning path, while in actual data the increase occurs mostly in the 1990s.

3.3 Volatility and Sharpe Ratios

Next we consider evidence on return volatility and Sharpe ratios. Table 8 records the standard deviation for equity returns averaged across 1000 simulations. Once again, we report results both for the full 70-year simulation and the first and second halves. The top row reproduces estimates from actual data. Our preferred scenario is again highlighted in boldface.

Table 8: Standard Deviation of Excess Returns

Data: 1934-2003	Full Sample	First Half	Second Half
	0.166	**0.177**	**0.151**
	Prior: 1890-1933		
	Full Sample	First Half	Second Half
Bayesian	0.130	0.161	0.066
$2 \log B = 5$	0.174	0.226	0.071
$2 \log B = 10$	0.241	0.305	0.115
	Prior: 1919-1933		
	Full Sample	First Half	Second Half
Bayesian	0.147	0.179	0.080
$2 \log B = 5$	0.128	0.146	0.078
$2 \log B = 10$	**0.136**	**0.148**	**0.092**

In the data, the standard deviation for excess returns was 0.166 for the period 1934-2003. Returns were slightly more volatile in the first half of the sample and slightly less in the second, with standard deviations of 0.177 and 0.151, respectively. In our preferred simulation, the standard deviation was 0.148 in the first half of the sample, 0.092 in the second, and 0.136 for the sample as a whole. Thus, the model accounts for slightly more than 80 percent of return volatility over the full sample and in the first half, but overstates the decline that occurred in the second half. Results for the Bayesian and other worst-case priors are similar. Indeed, the model's implications for volatility are less sensitive across priors than are implications for other features of the data.

Table 9 combines the numbers in tables 7 and 8 to calculate Sharpe ratios. As before, sample values and our favorite scenario are highlighted in boldface. In Shiller's data, the Sharpe ratio declined from 0.559 in the years 1934-1968 to 0.235 for the period 1969-2003, averaging 0.409 for the sample as a whole. For the full sample, our preferred scenario generates a Sharpe ratio of 0.376, which amounts to 92 percent of the observed value. However, the model understates the Sharpe ratio in the first half of the simulation and overstates it in the second, predicting values of 0.442 and 0.401, respectively.

Table 9: Sharpe Ratios for Excess Returns

Data: 1934-2003	Full Sample	First Half	Second Half
	0.409	**0.559**	**0.235**
	Prior: 1890-1933		
	Full Sample	First Half	Second Half
Bayesian	0.069	0.095	0.041
$2 \log B = 5$	0.182	0.215	0.208
$2 \log B = 10$	0.214	0.245	0.246
	Prior: 1919-1933		
	Full Sample	First Half	Second Half
Bayesian	0.182	0.237	0.136
$2 \log B = 5$	0.329	0.388	0.355
$2 \log B = 10$	**0.376**	**0.442**	**0.401**

Two aspects of the consumer's prior influence the Sharpe ratio. Holding constant the length of the training sample, the degree of initial pessimism varies directly with the Bayes factor. The larger the Bayes factor, the larger is the set of transition matrices consistent with the available observations, and the more alarming is the worst-case member of that set. Greater initial pessimism contributes to higher average excess returns early in the simulation by making the consumer shy away from the risky asset, thus increasing the Sharpe ratio. For both training samples, Sharpe ratios increase as we move down the table.

The length of the training sample influences the consumer's prior in two ways. First, with few prior degrees of freedom, the consumer is initially more pessimistic, but his prior is also more diffuse and his outlook is more flexible. As prior degrees of freedom decrease, the consumer is less capable of discriminating among competing models, and the gap between the worst-case and true models widens, thus increasing the degree of initial pessimism. Second, a small training sample also makes priors more diffuse, so that beliefs are more responsive to new observations and less firmly anchored to the initial transition matrix. A small training sample therefore promotes both pessimism and learning.

These two features of the prior – pessimism and diffusion – have countervailing effects on the Sharpe ratio. Greater initial pessimism contributes to higher average excess returns, but a more diffuse prior magnifies variation in the Radon-Nikodým derivative, thus contributing to higher volatility of returns. The former increases the Sharpe ratio, while the latter diminishes it.

3.4 Forecasting Excess Returns and Dividend Growth

Finally, we consider the implications of our learning model for predictability of excess returns and dividend growth. The literature on this subject is vast, and we address only a small part of it. [24] To establish some stylized facts, we run two sets of regressions. In the first, we project the average excess return from year t to $t+k$ onto a constant and the log of the current price-dividend ratio:

$$\frac{R_x(t, t+k)}{k} = \alpha_k + \beta_k \log(P_t^e/D_t) + u_{t+k}. \tag{36}$$

In the second, we replace excess returns with real dividend growth,

$$\frac{\log(D_{t+k}/D_t)}{k} = \alpha_k + \beta_k \log(P_t^e/D_t) + u_{t+k}. \tag{37}$$

We use Shiller's data to measure dividends and returns. Both regressions are estimated by OLS, and Newey and West (1987) standard errors are calculated to correct for serial correlation and heteroskedasticity in the residuals. Table 10 records the results for the period 1934-2003, with allowances at the end for calculating multiperiod returns and dividend growth.

Table 10 reproduces a number of results reported elsewhere. The price-dividend ratio predicts excess returns but not real dividend growth. In the excess return regressions, the coefficients on $\log(P_t^e/D_t)$ are negative and statistically significant; thus, a high price-dividend ratio portends lower excess returns in the future. In addition, the fit improves with the forecasting horizon. The R^2 statistics are not as

[24]For an interpretative survey, see Cochrane (2001, ch. 20.1). For previous research on learning and stock return predictability, see Timmermann (1993, 1996) and Adam et al. (2006).

large as those reported in other studies, but that is due to differences in the sample period. Many studies on return predictability focus on the period after the end of World War II until the late 1980s or mid 1990s. For example, using quarterly data spanning 1947-1996, Cochrane (2001, ch. 20) reports R^2 statistics ranging from 0.15 at the 1-year horizon to 0.6 at the 4-year horizon. We also get higher R^2 statistics when focusing on that period.

Table 10: Predictability of Excess Returns and Dividend Growth

	Excess Returns			Real Dividend Growth		
Horizon	$\hat{\beta}$	$SE(\hat{\beta})$	R^2	$\hat{\beta}$	$SE(\hat{\beta})$	R^2
1 year	-0.127	0.043	0.099	-0.018	0.024	0.006
2 years	-0.112	0.043	0.147	-0.016	0.021	0.008
3 years	-0.104	0.041	0.187	-0.018	0.018	0.018
4 years	-0.102	0.040	0.204	-0.019	0.017	0.027

Note: Estimates are based on Shiller's data, 1934-2003.

Tables 11 and 12 summarize estimates of the same regressions run on simulated data. Once again, the entries represent the mean across 1000 sample paths. Our preferred scenario is highlighted in boldface. Broadly speaking, the regressions run on simulated data reproduce the patterns reported in table 10. For excess returns, the coefficients on the log price-dividend ratio are negative and statistically significant, and the R^2 statistic increases with the forecast horizon. In the dividend-growth regressions, the coefficients are close to zero and statistically insignificant, and in many cases the R^2 statistic is less than 0.1. The main problem with the simulations is that the regressions fit too well. In the data, the R^2 statistics range from around 0.1 for 1 year ahead forecasts to 0.2 for 4 year ahead forecasts. In contrast, for our preferred model, the R^2 statistics range from 0.23 to 0.5, respectively, for 1 to 4-year ahead forecasts.

Cochrane (2001) summarizes the standard interpretation of this evidence, viz. that it is a manifestation of time-varying risk premia. Assuming a time-invariant probability law, he decomposes variation in the price-dividend ratio into movements arising from changes in expected returns and revisions in expected dividend growth. As an accounting matter, variation in the price-dividend ratio must reflect one or the other under rational expectations. Since price-dividend ratios forecast excess returns but not dividend growth, Cochrane infers that most of the variation in price-dividend ratios reflects time-varying expected returns. Assuming absence of arbitrage as well, time-varying expected returns must be due to time-varying risk premia. This points toward stochastic discount factor models that give rise to a lot of variation in risk premia. A high degree of risk aversion is typically needed to achieve this.

Table 11: Predictive Regressions Based on 1890-1933 Training Sample

Beta Prior

	Excess Returns			Real Dividend Growth		
Horizon	$\hat{\beta}$	$SE(\hat{\beta})$	R^2	$\hat{\beta}$	$SE(\hat{\beta})$	R^2
1 year	-0.159	0.119	0.107	0.007	0.013	0.059
2 years	-0.130	0.070	0.186	0.001	0.011	0.059
3 years	-0.110	0.050	0.245	-0.003	0.010	0.070
4 years	-0.110	0.039	0.293	-0.005	0.009	0.086

$2 \log B = 5$

	Excess Returns			Real Dividend Growth		
Horizon	$\hat{\beta}$	$SE(\hat{\beta})$	R^2	$\hat{\beta}$	$SE(\hat{\beta})$	R^2
1 year	-0.531	0.287	0.343	0.034	0.019	0.175
2 years	-0.385	0.122	0.508	0.022	0.014	0.137
3 years	-0.296	0.069	0.592	0.015	0.011	0.115
4 years	-0.236	0.047	0.634	0.011	0.009	0.105

$2 \log B = 10$

	Excess Returns			Real Dividend Growth		
Horizon	$\hat{\beta}$	$SE(\hat{\beta})$	R^2	$\hat{\beta}$	$SE(\hat{\beta})$	R^2
1 year	-0.680	0.297	0.410	0.030	0.016	0.185
2 years	-0.474	0.124	0.572	0.019	0.012	0.142
3 years	-0.354	0.072	0.637	0.013	0.010	0.115
4 years	-0.278	0.050	0.664	0.009	0.080	0.102

Table 12: Predictive Regressions Based on 1919-1933 Training Sample

Beta Prior

	Excess Returns			Real Dividend Growth		
Horizon	$\hat{\beta}$	$SE(\hat{\beta})$	R^2	$\hat{\beta}$	$SE(\hat{\beta})$	R^2
1 year	-0.175	0.103	0.165	0.010	0.011	0.068
2 years	-0.144	0.058	0.284	0.004	0.009	0.054
3 years	-0.122	0.041	0.368	0.000	0.008	0.054
4 years	-0.107	0.032	0.432	-0.002	0.007	0.062

$2 \log B = 5$

	Excess Returns			Real Dividend Growth		
Horizon	$\hat{\beta}$	$SE(\hat{\beta})$	R^2	$\hat{\beta}$	$SE(\hat{\beta})$	R^2
1 year	-0.246	0.135	0.211	0.024	0.014	0.120
2 years	-0.191	0.068	0.345	0.016	0.011	0.098
3 years	-0.153	0.043	0.429	0.011	0.010	0.088
4 years	-0.128	0.031	0.487	0.008	0.009	0.083

$2 \log B = 10$

	Excess Returns			Real Dividend Growth		
Horizon	$\hat{\beta}$	$SE(\hat{\beta})$	R^2	$\hat{\beta}$	$SE(\hat{\beta})$	R^2
1 year	**-0.285**	**0.138**	**0.227**	**0.028**	**0.015**	**0.139**
2 years	**-0.215**	**0.071**	**0.358**	**0.019**	**0.013**	**0.112**
3 years	**-0.172**	**0.045**	**0.440**	**0.013**	**0.011**	**0.096**
4 years	**-0.142**	**0.031**	**0.495**	**0.010**	**0.010**	**0.091**

Our model tells a different story. Return regressions run *ex post* on simulated data do not measure ex ante variation in expected returns. On the contrary, they represent the benefit of hindsight. Because the consumer has a constant discount factor, ex ante expected returns are constant and the risk premium is zero. Under rational expectations, the combination of a constant discount factor and predictable excess returns would signify the presence of arbitrage opportunities. But with learning, *ex post* 'predictability' of excess returns coexists with risk neutrality and absence of arbitrage because good deals emerge only in hindsight, too late for the consumer to exploit.

4 Concluding Remarks

This paper explores an asset pricing model in which Bayes' law gradually removes a representative consumer's initial pessimism about conditional distributions of consumption growth. We study an endowment economy with a representative consumer who learns about transition probabilities. We induce pessimistic prior beliefs via a short initial sample and/or a robustness calculation, then allow the consumer to update his beliefs via Bayes' theorem. Those evolving beliefs contribute a component to what a rational expectation econometrician would measure as the stochastic discount factor. We study how the market price of risk, Sharpe ratios, and the equity premium behave as the consumer learns.

At least in broad terms, the model can explain two asset pricing anomalies. One concerns a conflict between measures of risk aversion inferred from security market data and those derived from surveys or experiments in which individuals are presented with well-understood and well-controlled gambles. The former are usually very high, the latter quite low. Hansen and Jagannathan (1991) characterize this dissonance in terms of the market price of risk defined as the ratio of the standard deviation of a stochastic discount factor to its mean. A high market price of risk emerges from security market data, yet the degree of risk aversion found in surveys suggests a much lower price of risk.[25]

In our model, the conflict is explained by noting that there are two market prices of risk that correspond to the two probability measures in play. One uses the representative consumer's subjective transition probabilities to evaluate the mean and standard deviation, and the other uses the true transition probabilities. If one were to interview the representative consumer in our model, his answers would reflect subjective probabilities and convey a small price of risk. But if one calculated the market price of risk needed to reconcile equilibrium returns with the true transition probabilities, as in a rational-expectations model, the estimate would be substantially higher, reflecting the near-arbitrage[26] opportunities seemingly available *ex post*. A mild degree of risk aversion for investors thereby coexists with a high price of risk estimated from asset return data.

One version of our model in which priors are sufficiently diffuse and pessimistic accounts reasonably well for the equity premium. Equities are initially undervalued relative to the true transition probabilities because the representative consumer over-estimates the likelihood of another Depression. The representative consumer becomes less pessimistic as he learns, and equity prices rise, correcting the initial undervaluation. A high ex-post equity premium emerges because realized excess returns systematically exceed expected excess returns, reflecting the prevalence of unexpected capital gains along the transition path. Because some transitions occur rarely, the learning process takes a long time. The model equity premium falls as agents learn, in accordance with evidence reported by Blanchard (1993), Jagannathan et al. (2000), Fama and French (2002), DeSantis (2004), and others that the actual equity premium

[25] For example, see Barsky et al. (1997).

[26] By 'near arbitrage,' we mean an investment opportunity with a high Sharpe ratio. Cochrane and Saá-Requejo (2000) call this a 'good deal.' Under rational expectations, agents would have to be very risk averse to shy away from good deals.

is declining. These authors also emphasize that expected excess stock returns are now lower than historical averages, which is also an implication of our model.

We accomplish this within a model in which the discount factor β is slightly less than 1 and the coefficient of relative risk aversion α is 0. Our investors are patient and risk tolerant. The action is due to pessimism and learning. Model uncertainty is more important than risk.

References

Abel, Andrew B. 2002. An Exploration of the Effects of Pessimism and Doubt on Asset Returns. *Journal of Economic Dynamics and Control* 26:1075–1092.

Adam, K., A. Marcet, and J.P. Nicolini. 2006. Stock Market Volatility and Learning. Manuscript, European Central Bank.

Anderson, E., L.P. Hansen, and T.J. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1 (1):68–123.

Azeredo, Francisco and Viral Shah. 2007. Asset Pricing in an Exchange Economy with Bayesian Agents. Department of Economics, University of California, Santa Barbara.

Barsky, R. and J.B. DeLong. 1993. Why Does the Stock Market Fluctuate? *Quarterly Journal of Economics* 108:291–311.

Barsky, Robert B., Miles S. Kimball, F. Thomas Juster, and Matthew D. Shapiro. 1997. Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Survey. Nber working papers, National Bureau of Economic Research, Inc.

Blanchard, O.J. 1993. Movements in the Equity Premium. *Brookings Papers on Economic Activity* 2:75–118.

Bossaerts, Peter. 2002. *The Paradox of Asset Pricing*. Princeton, New Jersey: Princeton University Press.

———. 2004. Filtering Returns for Unspecified Biases in Priors when Testing Asset Pricing Theory. *Review of Economic Studies* 71:6386.

Brandt, M.W., Q. Zeng, and L. Zhang. 2004. Equilibrium Stock Return Dynamics Under Alternative Rules of Learning About Hidden States. *Journal of Economic Dynamics and Control* 28:1925–1954.

Cagetti, Marco, Lars Peter Hansen, Thomas J. Sargent, and Noah Williams. 2002. Robustness and Pricing with Uncertain Growth. *Review of Financial Studies* 15 (2):363–404.

Campbell, John Y. and John H. Cochrane. 1999. By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107:205–251.

Cecchetti, Stephen G., Pok-Sang Lam, and Nelson C. Mark. 2000. Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True? *American Economic Review* 90 (4):787–805.

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

———. 2001. *Asset Pricing*. Princeton, New Jersey: Princeton University Press.

Cochrane, John H. and Lars Peter Hansen. 1992. Asset Pricing Explorations for Macroeconomics. *NBER Macroeconomics Annual* 7:115–164.

Cochrane, John H. and J. Saá-Requejo. 2000. Beyond Arbitrage: Good Deal Asset Price Bounds in Incomplete Markets. *Journal of Political Economy* 108:79–119.

DeGroot, M.H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.

DeSantis, M. 2004. Movements in the Equity Premium: Evidence from a Bayesian Time-Varying VAR. Unpublished manuscript, University of California, Davis.

Fama, E.F. and K.R. French. 2002. The Equity Premium. *Journal of Finance* 57:637–659.

Fishburn, P.C. 1970. *Utility Theory for Decision Making*. New York: Wiley.

Friedman, M. and A.J. Schwartz. 1963. *A Monetary History of the United States, 1857-1960*. Princeton, New Jersey: Princeton University Press.

Geweke, J.F. 2001. A Note on Some Limitations of CRRA Utility. *Economics Letters* 71:341–46.

Hamilton, James D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57 (2):357–84.

Hansen, Lars Peter and Ravi Jagannathan. 1991. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99:225–262.

Hansen, Lars Peter and Thomas J. Sargent. 2005. Robust Estimation and Control under Commitment. *Journal of Economic Theory* 124 (2):258–301.

———. 2006. Fragile beliefs and the price of model uncertainty. Unpublished.

———. 2007a. Recursive Robust Estimation and Control without Commitment. *Journal of Economic Theory* In press.

———. 2007b. Robustness. Princeton University Press, forthcoming.

Hansen, Lars Peter, Thomas Sargent, and Thomas Tallarini. 1999. Robust Permanent Income and Pricing. *Review of Economic Studies* 66:873–907.

Hansen, Lars Peter, Thomas J. Sargent, and Neng E. Wang. 2002. Robust Permanent Income and Pricing with Filtering. *Macroeconomic Dynamics* 6:40–84.

Jagannathan, R., E.R. McGratten, and A. Scherbina. 2000. The Declining U.S. Equity Premium. *Federal Reserve Bank of Minneapolis Quarterly Review* 4:3–19.

Kandel, S. and R.F. Stambaugh. 1991. Asset Returns and Intertemporal Preferences. *Journal of Monetary Economics* 27:39–71.

Kim, C.J. and C.R. Nelson. 1999. *State-Space Models with Regime Switching*. Cambridge MA: MIT Press.

Kocherlakota, Narayana R. 1996. The Equity Premium: It's Still a Puzzle. *Journal of Economic Literature* 34 (1):42–71.

Kurz, Mordecai and A. Beltratti. 1997. The Equity Premium is No Puzzle. In *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs*, edited by M. Kurz. New York: Springer-Verlag.

Kurz, Mordecai, H. Jin, and M. Motolese. 2004. Determinants of Stock Market Volatility and Risk Premia. Department of Economics, Stanford University.

Ljungqvist, Lars and Thomas J. Sargent. 2004. *Recursive Macroeconomic Theory, Second edition.* Cambridge, Mass.: MIT press.

Mehra, Rajnish and Edward C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (2):145–161.

Newey, Whitney K. and Kenneth D. West. 1987. A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–708.

Raftery, A. 1996. Hypothesis Testing and Model Selection. In *Markov Chain Monte Carlo in Practice*, edited by W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. London: Chapman and Hall/CRC.

Rietz, T.A. 1988. The Equity Premium: A Solution. *Journal of Monetary Economics* 22:117–131.

Savage, L.J. 1954. *The Foundations of Statistics.* New York: John Wiley & Sons.

Tallarini, Thomas D. 2000. Risk-sensitive real business cycles. *Journal of Monetary Economics* 45 (3):507–532.

Timmermann, Allan G. 1993. How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices. *Quarterly Journal of Economics* 108 (4):1135–45.

———. 1996. Excess Volatility and Predictability of Stock Prices in Autoregressive Dividend Models with Learning. *Review of Economic Studies* 63 (4):523–57.

Weitzman, M. 2007. Subjective Expectations and Asset-Return Puzzles. *American Economic Review* 97(4):1102–1130.

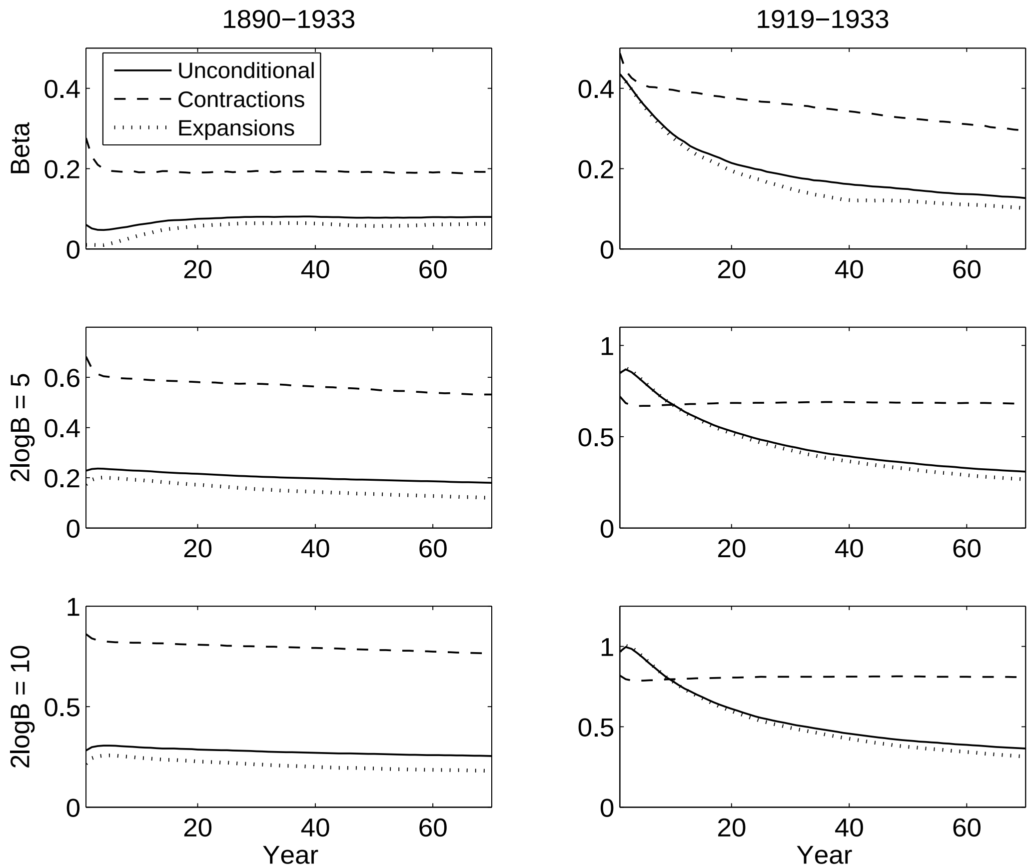

Figure 1: Rational Expectations Prices of Risk. Solid lines depict the simulation average of the unconditional MPR, and dashed and dotted lines portray its value in contractions and expansions, respectively. The columns correspond to the two training samples, and the rows refer to differing degrees of robustness.

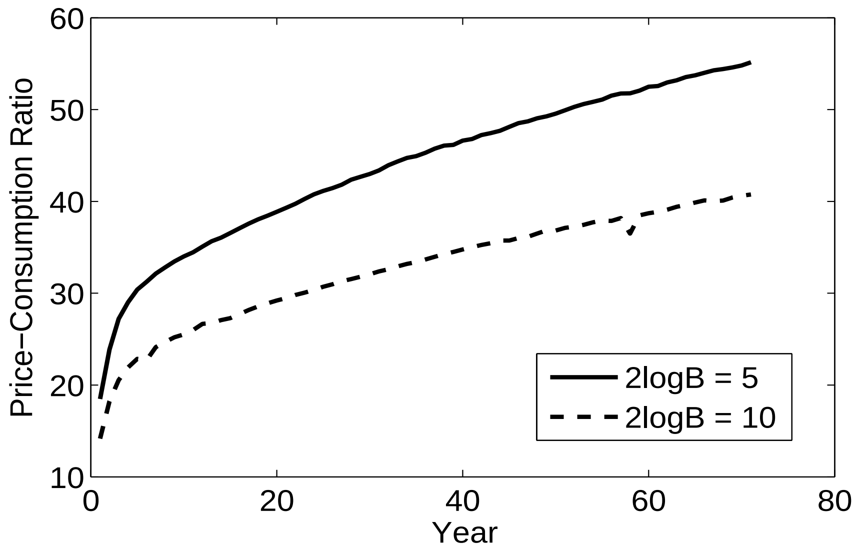

Figure 2: Price-Consumption Ratio. The two curves depict the simulation average of the price-consumption ratio for worst-case priors constructed from the 1919-1933 training sample. The dashed line represents our preferred scenario.