Learning about Consumption Dynamics

Michael Johannes, Lars Lochstoer, and Yiqun Mou*
Columbia Business School

First draft: March 2010

June 29, 2011

Abstract

This paper studies the asset pricing implications of Bayesian learning about the parameters, states, and models determining aggregate consumption dynamics. Our approach is empirical and focuses on the quantitative implications of learning in real-time using post World War II consumption data. We characterize this learning process and find that revisions in beliefs stemming from parameter and model uncertainty are significantly related to realized aggregate equity returns. This evidence is novel, providing strong support for a learning-based story. Further, we show that beliefs regarding the conditional moments of consumption growth are strongly time-varying and exhibit business cycle and/or long-run fluctuations. Much of the long-run behavior is unanticipated ex ante. We embed these subjective beliefs in a general equilibrium model to investigate further asset pricing implications. We find that learning significantly improves the model's ability to fit standard asset pricing moments, relative to benchmark model with fixed parameters. This provides additional evidence supporting the importance of learning.

---

*All authors are at Columbia Business School, Department of Finance and Economics. We would like to thank Francisco Barillas (CEPR Barcelona discussant), Pierre Collin-Dufresne, Kent Daniel, Lars Hansen (AFA discussant), Lubos Pastor (NBER discussant), Tano Santos, and seminar participants at AFA Denver 2011, CEPR conference in Barcelona May 2011, Columbia University, NBER Asset Pricing meeting in Chicago 2011, Stanford, UT Austin, and the University of Wisconsin (Madison) for helpful comments. Corresponding author: Lars Lochstoer, 405B Uris Hall, 3022 Broadway, New York, NY 10027. Email: LL2609@columbia.edu

1 Introduction

This paper studies the asset pricing implications of learning about aggregate consumption dynamics. We are motivated by practical difficulties generated from the use of complicated consumption-based asset pricing models with many difficult-to-estimate parameters and latent states. For example, parameters or states controlling long-run consumption growth are at once extremely important for asset pricing and particularly difficult to estimate. Thus, we are interested in studying an economic agent who is burdened with some of the same econometric problems faced by researchers, a problem suggested by Hansen (2007).[1]

A large existing literature studies asset pricing implications of statistical learning – the process of updating beliefs about uncertain parameters, state variables, or even model specifications. Pastor and Veronesi (2009) provide a recent survey. In theory, learning can generate a wide range of implications relating to stock valuation, levels and variation in expected returns and volatility, and time series predictability, with many of the results focussed on the implications of learning about dividend dynamics.

Our analysis differs from existing work along three key dimensions. First, we focus on the empirical implications of simultaneously learning about parameters, state variables, and even model specifications. Most existing work focuses on learning a single parameter or state variable. Learning about multiple unknowns is more difficult as additional unknowns often confounds inference, slowing the learning process. Second, we focus on the specific implications of real-time learning about consumption dynamics from macroeconomic data during the U.S. post World War II experience. Thus, we are not expressly interested in general asset pricing implications of learning in repeated sampling settings, but rather the *specific* implications generated by the historical macroeconomic shocks realized in the United States over the last 65 years. Third, we use a new and stringent test of learning that relates updates in investor beliefs to contemporaneous, realized equity returns.

In studying the implications of learning, we focus on the following types of questions. Could an agent who updates his beliefs rationally detect non-i.i.d. consumption growth dynamics in real time? How rapidly does the agent learn about parameters and models? Are the revisions in beliefs about consumption moments correlated with asset returns, as a

[1] Hansen (2007) states: *"In actual decision making, we may be required to learn about moving targets, to make parametric inferences, to compare model performance, or to gauge the importance of long-run components of uncertainty. As the statistical problem that agents confront in our model is made complex, rational expectations' presumed confidence in their knowledge of the probability specification becomes more tenuous. This leads me to ask: (a) how can we burden the investors with some of the specification problems that challenge the econometrician, and (b) when would doing so have important quantitative implications" (p.2).*

learning story would require? Is there evidence that learning effects can help us understand standard asset pricing puzzles, such as the high equity premium, return volatility, and degree of return predictability?

One of the key implications of learning is that the agent's beliefs are nonstationary. For example, the agent may gradually learn that one model fits the data better than an alternative model or that a parameter value is higher or lower than previously thought, both of which generate nonstationarity in beliefs. The easiest way to see this is to note that the posterior mean of a parameter, $\mathbb{E}[\theta|y^t]$, where y^t is data up to time t, is trivially a martingale. Thus revisions in beliefs represent permanent, nonstationary shocks, that can have important asset pricing implications. For instance, nonstationary dynamics can generate a quantitatively important wedge between *ex post* outcomes and *ex ante* beliefs, providing an alternative explanation for standard asset pricing quantities such as the observed equity premium or excess return predictability.[2]

We study learning in the context of three standard Markov switching models of consumption growth: unrestricted 2- and 3-state models and a restricted 2-state model that generates i.i.d. consumption growth. The states capture business cycle fluctuations and can be labeled as expansion and recession in 2-state models, with an additional 'disaster' state in 3-state models.[3] Our key assumption is that the agent views the parameters, states, and even models as unknowns, using Bayes rule to update beliefs using consumption data, as well as additional macroeconomic data such as GDP growth in extensions.

To focus on different aspects of learning, we consider three sets of initial parameter beliefs. The first, the 'historical prior,' trains the prior using Shiller's consumption data from 1889 until 1946, a common approach to generate 'objective' priors.[4] The second, the 'look-ahead prior,' sets prior parameter means to full-sample maximum likelihood point estimates using post World War II data. We embed substantial uncertainty around these estimates to study the effect of parameter uncertainty. This is often called an 'empirical Bayes' approach.

[2] See also Cogley and Sargent (2008), Timmermann (1993), and Lewellen and Shanken (2002).

[3] Markov switching models for consumption or dividends are a benchmark specification in the literature, see, e.g., Mehra and Prescott (1985), Rietz (1988), Cecchetti, Lam, and Mark (1990, 1993), Whitelaw (2000), Cagetti, Hansen, Sargent, and Williams (2002), Barro (2006), Barro and Ursua (2008), Chen (2008), Bhamra, Kuehn, and Strebulaev (2008), Barro, Nakamura, Steinsson and Ursua (2009), Backus, Chernov, and Martin (2009), and Gabaix (2009). Rietz (1988) and, more recently, Barro (2006, 2009) argue that consumption disaster risk can help explain some of the standard macro-finance asset pricing puzzles.

[4] We do account for measurement error, which likely increased reported macroeconomic volatility during the pre-war period, as argued in Romer (1989). Malmendier and Nagel (2011) present evidence that the experience of the Great Depression affected investors' subsequent beliefs about risk and return, broadly consistent with the Historical prior calibration approach.

The third, the fixed parameter prior, is a rational expectations benchmark with dogmatic beliefs that are fixed at the end-of-sample parameter estimates. Thus, there is no parameter uncertainty. There is state uncertainty, however, which allows us to separate the effects of parameter and state uncertainty.

Our first results characterize the beliefs about parameters, states, models, and future consumption dynamics (e.g., moments) through the sample. The perceived dynamic behavior of aggregate consumption is at the heart of consumption-based asset pricing as it, jointly with preferences, determines the dynamic properties of the pricing kernel. In terms of beliefs, we compute at each point in time the posterior distribution of parameters, states, and models. As new data arrives, we update beliefs using Bayes rule. In addition to usual summaries of parameters and states, we also compute model probabilities and perform 'model monitoring' in real time as new data arrives. We find that the posterior probability of the i.i.d. model falls dramatically over time, provided the prior weight is less than one. Thus our agent is able to learn in real-time that consumption growth is not i.i.d., but has persistent components.[5] The agent believes that expected consumption growth is low in recessions and high in expansions, with the opposite pattern for consumption growth volatility. The 2-state model quickly emerges as the most likely, but the 3-state model with a disaster state has $5-10\%$ probability at the end of the sample. At the onset of the financial crisis in 2008, the probability of the disaster model increases.[6]

There is significant learning about the expansion state parameters, slower learning about the recession state, and almost no learning about the disaster state, as it is rarely, if ever, visited. Thus, there is an observed differential in the speed of learning. Standard large sample theory implies that all parameters converge at the same rate, but the realized convergence rate depends on the actual observed sample path. There is also strong evidence for non-stationary time-variation in the conditional means and variances of consumption growth, as well as measures of non-normality such as skewness and kurtosis. For both the historical and the look-ahead priors, the agent's perception of the long-run mean (volatility) of consumption growth generally increases (decreases) over the sample. The perceived persistence of recessions (expansions) decreases (increases).[7] As the agent's beliefs about these parameters

[5] This result is robust to persistence induced by time-aggregation of the consumption data (see Working (1960)).

[6] The posterior probability of the three-state model would change dramatically, if visited. For example, if a -3% quarterly consumption growth shock were realized at the end of the sample, the posterior probability of the three-state model would increase to almost 50%.

[7] All of the results described in the current and previous paragraphs are robust to learning from additional GDP growth data.

and moments change, asset prices and risk premia will also change.

The first formal test of the importance of learning regresses contemporaneous excess stock market returns on revisions in beliefs about expected consumption growth. This test, which to our knowledge is new to the literature, is a fundamental implication of any learning-based explanation: for learning to matter, unexpected revisions in beliefs about expected consumption growth should be reflected in the unexpected aggregate equity returns.[8]. We find strong statistical evidence that this relationship is positive, and the results are similar for both the historical and the look-ahead prior. To disentangle parameter from state learning, we include revisions in beliefs generated by the fixed parameter prior as a control. Revisions in beliefs obtained using the historical and look-ahead priors remain statistically significant, but revisions in beliefs generated by models with known parameters are statistically insignificant.

These results imply that learning about parameters and models is a statistically significant determinant of asset returns in our sample, confirming our main hypothesis. This result is strengthened if the agent learns from both consumption and GDP growth. It is important to note that our agent only learns in real-time and from macroeconomic fundamentals, as no asset price data (such as the dividend-price ratio) is used when forming beliefs. Since the revisions in beliefs obtained from the models with fixed parameters are statistically insignificant, the evidence questions the standard full-information, rational expectations implementation of the standard consumption-based model, at least for the models of consumption dynamics that we consider.[9]

As mentioned earlier, parameter and model learning generate nonstationary dynamics and permanent shocks that could have important implications. To investigate these implications, we consider a formal asset pricing exercise assuming Epstein-Zin preferences. Because the specific time-path of beliefs is important, the usual calibration and simulation approach used in the literature is not applicable, and we consider the following alternative pricing procedure. At time t, given beliefs over parameters, models, and states, our agent prices a levered claim to a future consumption stream, computing quantities such as *ex-ante* expected returns and dividend-price ratios.[10] Then, at time $t+1$, our agent updates beliefs

[8] The sign of the effect would in a model depend on the elasticity of intertemporal substitution, and also on the other moments that change at the same time (volatility, skewness, kurtosis, etc.). In the model section, we show that this positive relation is consistent with a model with an elasticity of intertemporal substitution greater than 1.

[9] Parameter and model learning, on the one hand, and state learning on the other hand are distinct in our setting because the former generates a non-stationary path of beliefs, while the latter, after an initial burn-in period, is stationary.

[10] We do price a levered consumption claim and introduce idiosyncratic noise to break the perfect rela-

using new macro realizations at time $t+1$, recomputes prices, expected returns and dividend-price ratios. From this time series of prices, we compute realized equity returns, volatilities, etc. Thus, we feed historically realized macroeconomic data into the model and analyze the asset pricing implications for various models and prior specifications. This process is required when the time path matters and was previously used in, for example, Campbell and Cochrane (1999), where habit is a function of past consumption growth. We use standard preference parameters taken from Bansal and Yaron (2004).

Solving the full pricing problem with priced parameter uncertainty is computationally prohibitive, as the dimensionality of the problem is too large.[11] To price assets in a tractable way, while still incorporating learning, we follow Piazzesi and Schneider (2010) and Cogley and Sargent (2009) and use a version of Kreps' (1994) anticipated utility. This implies that our agent prices claims at each point in time using current posterior means for the parameters and model probabilities, assuming those values will persist into the indefinite future. We do account for state uncertainty when pricing.

This pricing experiment provides additional evidence, along multiple dimensions, for the importance of learning. Focussing on the 3-state model, we first note that the model with parameters fixed at the full-sample values has a difficult time with standard asset pricing moments: the realized equity premium and Sharpe ratio are less than half the values observed in the data. The volatility of the price-dividend ratio is eighty percent less than the observed value. Parameter learning uniformly improves *all* of these statistics, bringing them close to observed values. The results are, after a burn-in period, similar for the look-ahead and the historical prior as the agent quickly unlearns the mean parameter beliefs of the look-ahead prior early in the sample. It is important to note that this is not a calibration exercise – we did not choose the structural parameters to generate these returns.

The increase in the realized equity premium and return volatility is due to unexpected revisions in beliefs resulting from the parameter and model learning. In particular, the average annualized *ex ante* quarterly risk premium is similar across the models at about 1.7%, but the models with uncertain parameters generate a higher realized equity premium of about 3.8% to 4.2%, close to the 4.7% observed over the sample. This documents a dramatic

tionship between consumption and dividend growth. The dividends are calibrated to match the volatility of dividend growth and the correlation between dividend and consumption growth.

[11]As an example, for the 3-state model there are twelve parameters, each with two hyperparameters characterizing the posteriors. This implies that we would have to have to solve numerically for prices on a very high dimensional grid, which is infeasible. There are additional difficult technical issues associated with priced parameter uncertainty, as noted by Geweke (2001) and Weitzman (2007).

impact of the specific time path of beliefs about parameters and models for standard asset pricing statistics, at least relative to the fixed parameter, rational expectations benchmark. This also implies, looking forward, that the perceived equity premium is much smaller than the realized equity premium over the post World War II period. These points are consistent with the results in Cogley and Sargent (2008).[12]

In terms of predictability, the returns generated by learning over time closely match the data. For the historical and look-ahead priors and for forecasting excess market returns with the lagged log dividend-price ratio, the generated regression coefficients and R^2's are increasing with the forecasting horizon and similar to those found in the data. The fixed parameters case, however, does not deliver significant *ex post* predictability, although the *ex ante* risk premium is in fact time-varying in these models as well because the risk premium time-variation assuming fixed parameters is too small relative to the volatility of realized returns to result in significant t-statistics. The intuition for why *in-sample* predictability occurs when agents are uncertain about parameters and models is the same as in Timmermann (1993) and Lewellen and Shanken (2002) – unexpected updates in growth and discount rates impact the dividend-price ratio and returns in opposite directions leading to the observed positive in-sample relation. Thus, in-sample predictability can be expected with parameter and model learning. The quantitatively large degree of in-sample relative to out-of-sample predictability we find is consistent with the literature.[13]

We also note that the model exhibits volatile long maturity risk-free yields, consistent with the data. Learning about fixed quantities such as models or parameters generate permanent shocks that affect agents' expectations of the long-run (infinite-horizon) distribution of consumption growth. This is different from existing asset pricing models where only stationary variables affect marginal utility growth (see, e.g., Bansal and Yaron (2004), and Wachter's (2005) extension of Campbell and Cochrane (1999) model, as well as our fixed parameters benchmark model). In these models, long-run (infinite-horizon) risk-free yields are

[12] Cogley and Sargent (2008) assume negatively biased beliefs about the consumption dynamics to highlight the same mechanism and also consider the role of robustness. In their model, the subjective probability of recessions is higher than the 'objective' estimate from the data. The results we present here are consistent with their conclusions, but our models are estimated from fundamentals in real-time, which allows for an out-of-sample examination of the time-series of revisions in beliefs. Further, we allow for learning over different models of the data generating process, as well as *all* the parameters of each model.

[13] For example, Fama and French (1988) document a high degree of in-sample predictability of excess (long-horizon) stock market returns using the price-dividend ratio as the predictive variable. On the other hand, Goyal and Welch (2008) and Ang and Bekaert (2007) document poor *out-of-sample* performance of these regressions in the data, and the historical and look-ahead prior learning models presented here are consistent with this evidence.

constant as the transitory shocks to marginal utility growth die out in the long run. This is additional evidence supporting a learning-based explanation relative to the fixed parameters alternative.

In conclusion, our results strongly support the importance of parameter and model learning for understanding the joint behavior of consumption and asset prices in the U.S. post World War II sample. First, parameter and model learning leads to a time path of belief revisions that are correlated with realized equity returns, controlling for realized consumption growth. Second, the time series of beliefs help explain the time-series of the price level of the market (the time-series of the price-dividend ratio) in a general equilibrium model. Third, beliefs display strong nonstationarity over time, driving a wedge between *ex ante* beliefs and *ex post* realizations that is absent in rational expectations models. Fourth, permanent shocks to beliefs generate permanent shocks to marginal utility growth. These features help explain common asset pricing puzzles such as excess return volatility, the high sample equity premium, the high degree of in-sample return predictability, and the high volatility of long-run yields, all relative to a fixed parameter alternative. The results are generated by real-time learning from consumption (and GDP growth), using standard preference parameters without directly calibrating to asset returns. In this sense the results are entirely "out-of-sample."

2 The Environment

2.1 Model

We follow a large literature and assume an exogenous Markov or regime switching process for aggregate, real, per capita consumption growth dynamics. Log consumption growth, Δc_t, evolves via:

$$\Delta c_t = \mu_{s_t} + \sigma_{s_t} \varepsilon_t, \tag{1}$$

where ε_t are i.i.d. standard normal shocks, $s_t \in \{1, ..., N\}$ is a discretely-valued Markov state variable, and $\left(\mu_{s_t}, \sigma^2_{s_t}\right)$ are the Markov state-dependent mean and variance of consumption growth. The Markov chain evolves via a $N \times N$ transition matrix Π with elements π_{ij} such that $\text{Prob}[s_t = j | s_{t-1} = i] = \pi_{ij}$, with the restriction that $\sum_{j=1}^{N} \pi_{ij} = 1$. The fixed parameters of the N-state model contain the means and variances in each state, $\left\{\mu_n, \sigma^2_n\right\}_{n=1}^{N}$ as well as the elements of the transition matrix. The transition matrix controls the persistence of the

Markov state.

Markov switching models are flexible and tractable and have been widely used since Mehra and Prescott (1985) and Rietz (1988). By varying the number, persistence, and distribution of the states, the model can generate a wide range of economically interesting and statistically flexible distributions. Although the ε_t's are i.i.d. normal and the distribution of consumption growth, conditional on s_t and parameters, is normally distributed, the distribution of future consumption growth is neither i.i.d. nor normal due to the shifting Markov state. This time-variation induces very flexible marginal and predictive distributions for consumption growth. These models are also tractable, as it is possible to compute likelihood functions and filtering distributions, given parameters.

We consider two and three state models and also consider a restricted version of the two state model generating i.i.d consumption growth by imposing the restriction $\pi_{11} = \pi_{21}$ and $\pi_{22} = \pi_{12} = 1 - \pi_{11}$. Under this assumption, consumption growth is an i.i.d. mixture of two normal distributions, essentially a discrete-time version of Merton's (1976) mixture model. The general two and 3-state models have 6 and 12 parameters, respectively. The i.i.d. two state model has 5 parameters ($\mu_1, \mu_2, \sigma_1, \sigma_2$ and π_{11}).

It is common in these models to provide business cycle labels to the states. In a 2-state model, we interpret the two states as 'recession' and 'expansion,' while the three state model additionally allows for a 'disaster' state.[14] Although rare event models have been used for understanding equity valuation since Rietz (1988), there has been a recent resurgence in research using these models (see, e.g., Barro (2006, 2009), Barro and Ursua (2008), Barro, Nakamura, Steinsson and Ursua (2009), Backus, Chernov, and Martin (2009), and Gabaix (2009)).

2.2 Information and learning

To operationalize the model, additional assumptions are required regarding the economic agent's information set. Since we want to model learning similar to that faced by the econometrician, we assume agents observe aggregate consumption growth, but are uncertain about the Markov state, the parameters, and the total number of Markov states. We label these unknowns as state, parameter, and model uncertainty, respectively. We assume agents

[14] We do not consider, for instance, 1- or 4-state models as the Likelihood ratios of these relative to the 2- or 3-state model show that the 2- and 3-state models better describe the data. As we will show, however, there is some time-variation in whether a 2- or 3-state model matches the data better, which is one of the reasons we entertain both of these as alternative models.

are Bayesian, which means they update initial beliefs via Bayes' rule as data arrives. Later in the paper, we develop an extension to this model where agents can also learn from a vector of additional macro variables and consider the case of additional learning from GDP growth data.

The learning problem is as follows. We consider $k = 1, ..., K$ models, $\{\mathcal{M}_k\}_{k=1}^K$, and in model $\mathcal{M}_k$, the state variables and parameters are denoted as s_t and θ, respectively.[15] The distribution $p(\theta, s_t, \mathcal{M}_k|y^t)$ summarizes beliefs after observing data $y^t = (y_1, ...y_t)$. To understand the components of the learning problem, we can decompose the posterior as:

$$p\left(\theta, s_t, \mathcal{M}_k|y^t\right) = p\left(\theta, s_t|\mathcal{M}_k, y^t\right) p\left(\mathcal{M}_k|y^t\right). \tag{2}$$

$p(\theta, s_t|\mathcal{M}_k, y^t)$ solves the parameter and state "estimation" problem conditional on a model and $p(\mathcal{M}_k|y^t)$ provides model probabilities. It is important to note that this is a non-trivial, high-dimensional learning problem, as posterior beliefs depend in a complicated manner on past data and can vary substantially over time. The dimensionality of the posterior can be high, in our case more than 10 dimensions.

One of our primary goals is to characterize and understand the asset pricing implications of the transient process of learning about the parameters, states, and models.[16] Learning generates a form of nonstationarity, since parameter estimates and model probabilities are changing through the sample. When pricing assets, this can lead to large differences between *ex ante* beliefs and *ex post* outcomes, as shown in Cogley and Sargent (2008). Given this nonstationarity, we are concerned with understanding the implications of learning based on the specific experience of the U.S. post-war economy.[17]

To operationalize the learning problem, we need to specify the prior distribution, the data the agent uses to update beliefs, and develop an econometric method for sampling from the posterior distribution. In terms of data, we in a benchmark case assume that agents learn only from observing past and current consumption growth, a common assumption in the learning literature (see, e.g., Cogley and Sargent (2008) and Hansen and Sargent

[15] This is a notational abuse. In general, the state and dimension of the parameter vector should depend on the model, thus we should superscript the parameters and states by 'k', θ^k and s_t^k. For notational simplicity, we drop the model dependence and denote the parameters and states as θ and s_t, respectively.

[16] These type of problems received quite a bit of theoretical attention early in the rational expectations paradigm - see for example Bray and Savin (1986) for a discussion of model specification and convergence to rational expectations equilibria by learning from observed outcomes.

[17] This is different from the standard practice of looking at population or average small-sample unconditional asset price and consumption growth moments from a model calibrated to the U.S. postwar data – we are looking at a single outcome corresponding to the U.S. post-war economy.

(2009)). The primary data used is the 'standard' data set consisting of real, per capita quarterly consumption growth observations obtained from the Bureau of Economic Analysis (the National Income and Product Account tables) from 1947:$Q1$ until 2009:$Q1$.

2.3 Initial beliefs

The learning process begins with initial beliefs or the prior distribution. In terms of functional forms, we assume proper, conjugate prior distributions (Raiffa and Schlaifer (1956)). One alternative would be flat or 'uninformative' priors, but this is not possible in Markov switching models, as this creates identification issues (the label switching problem) and causes problems sampling from the posterior.[18] Conjugate priors imply that the functional form of beliefs is the same before and after sampling, are analytically tractable for econometric implementation, and are flexible enough to express a wide range initial beliefs.

For the mean and variance parameters in each state, (μ_i, σ_i^2), the conjugate prior is $p(\mu_i|\sigma_i^2)p(\sigma_i^2) \sim \mathcal{NIG}(a_i, A_i, b_i, B_i)$, where $\mathcal{NIG}$ is the normal/inverse gamma distribution. The transition probabilities are assumed to follow a Beta distribution in 2-state specification and its generalization, the Dirichlet distribution, in models with three states. Calibration of the hyperparameters completes the specification.

We endow our agent with economically motivated initial beliefs to study how learning proceeds from various starting points. We consider three prior distributions and use an 'objective' approach to calibrate the prior parameters. The first, the 'historical prior,' uses a training sample to calibrate the prior distribution. Training samples are the most common way of generating objective prior distributions (see, e.g., O'Hagan (1994)). In this case, an initial data set is used to provide information on the location and scale of the parameters. In our application, we use the annual consumption data from Shiller from 1889 until 1946. Given the prior generated from the training sample, learning proceeds on the second data set – in our case, the post World War II sample.[19]

[18] The label switching problem refers to the fact that the likelihood function is invariant to a relabeling of the components. For example, in a two-state model, it is possible to swap the definitions of the first and second states and the associated parameters without changing the value of the likelihood. The solution is to impose parameter constraints in optimization for MLE or to use informative prior distributions for Bayesian approaches. These constraints/information often take the form of an ordering of the means or variances of the parameters. For example in a two state model, it is common to impose that $\mu_1 < \mu_2$ and/or $\sigma_1 < \sigma_2$ to breaks the symmetry of the likelihood function.

[19] Romer (1989) presents evidence that a substantial fraction of the volatility of macro variables such as consumption growth pre-WW2 is due to measurement error. To alleviate this concern, we set the prior mean over the variance parameters to a quarter of the value estimated over the training sample. See the Appendix

The second is called the 'look-ahead prior.' This prior sets the prior mean for each parameter equal to full-sample maximum likelihood estimates using the post World War II sample, similar to the procedure employed in an 'Empirical Bayes' approach. The prior variances are chosen to be relatively flat around these full-sample estimates, in order to allow for meaningful learning about the parameters as new data arrives, without running into label-switching identification problems. This approach violates the central idea of the Bayesian approach, as the prior contains information from the sample, but it is useful for analyzing the evolution of parameter uncertainty through the post World War II sample. The main differences between the historical and the look-ahead priors are that the historical priors have on average higher consumption growth volatility, shorter expansions, and longer recessions. For the 3-state model, the disaster state is also more severe in the historical prior, reflecting the Great Depression.

The third is called the 'fixed parameter' prior. This is a point-mass prior located at the end-of-sample estimates. In this case, the agent only learns about the latent Markov state. This prior mimics the typical rational expectations approach and allows us to separately identify the role of state and parameter learning, since the other priors have both state and parameter learning.

The details of the priors, the specific prior parameters chosen, as well as a description of the econometric technique we apply to solve this high-dimensional learning problem (particle filtering) are given in the Appendix.

3 The time-series of subjective beliefs

This section characterizes the learning process. We first discuss state, parameter, and model learning and their implications for the time series of conditional consumption moments, as perceived by the Bayesian agent. Next, we empirically investigate how revisions in the agent's beliefs are related to stock market returns. We also consider the case of learning from GDP data, in addition to consumption data. In the following section, we embed these beliefs in a general equilibrium model and discuss the asset pricing implications in more detail.

---

for further details.

3.1 State and parameter learning

Conditional on a model specification, our agent learns about the Markov state and the parameters, with revisions in beliefs generated by a combination of data, model specification, and initial beliefs. To start, consider the agent's beliefs about the current state of the economy, s_t, where state 1 is an 'expansion' state, state 2 the 'contraction' state and, if a 3-state model, state 3 the 'disaster' state. Estimates are given by

$$E\left[s_t|\mathcal{M}_k, y^t\right] = \int s_t p\left(\theta, s_t|\mathcal{M}_k, y^t\right) d\theta ds_t.$$

Note that these are marginal *mean* state beliefs, as parameter uncertainty is integrated out. Although s_t is discrete, the mean estimates need not be integer valued. Figure 1 displays the posterior state beliefs over time, for each model and for different priors.

There are a number of notable features of these beliefs. NBER recessions (shaded yellow) and expansions are clearly identified in the models. The only exceptions are the recessions in the late 1960s and 2001, which were not associated with substantial consumption declines. Comparing the panels, one area in which the models generate strong differences is persistence of the states. The i.i.d. model identifies recessions as a one-off negative shock, but since shocks are i.i.d., the agent does not forecast that the recession state will persist with high likelihood. In contrast, the 2- and 3-state models clearly show the persistence of the recession states. Disaster states are rare – after the initial transient post war period, there are only really two observations that place even modest probability on the disaster state – the recession in 1981 and the financial crisis at the end of 2008. This implies that disaster states are nearly 'Peso' events in the post WW2 sample.

The agent's beliefs are quite volatile early in the sample in all of the models. This is not surprising. Since initial parameter beliefs are highly uncertain, the agent has a difficult time discerning the current state as parameter uncertainty exacerbates state uncertainty. As the agent learns, parameter uncertainty decreases and state identification is easier. It is important to note that even with full knowledge of the parameters, the agent will never be able to perfectly identify the state.[20] The results also show that the priors do not have a large impact on the mean state beliefs, at least for the unrestricted 2- and 3-state models, as the posterior beliefs are roughly similar for the historical and look-ahead priors.

[20] The posterior variance of the state, $var\left[s_t|\mathcal{M}_k, y^t\right]$, does decline over time due to decreasing parameter uncertainty. This will be discussed further when we use GDP growth as an additional observation to help identify the state.

Figure 1 - Evolution of Mean State Beliefs

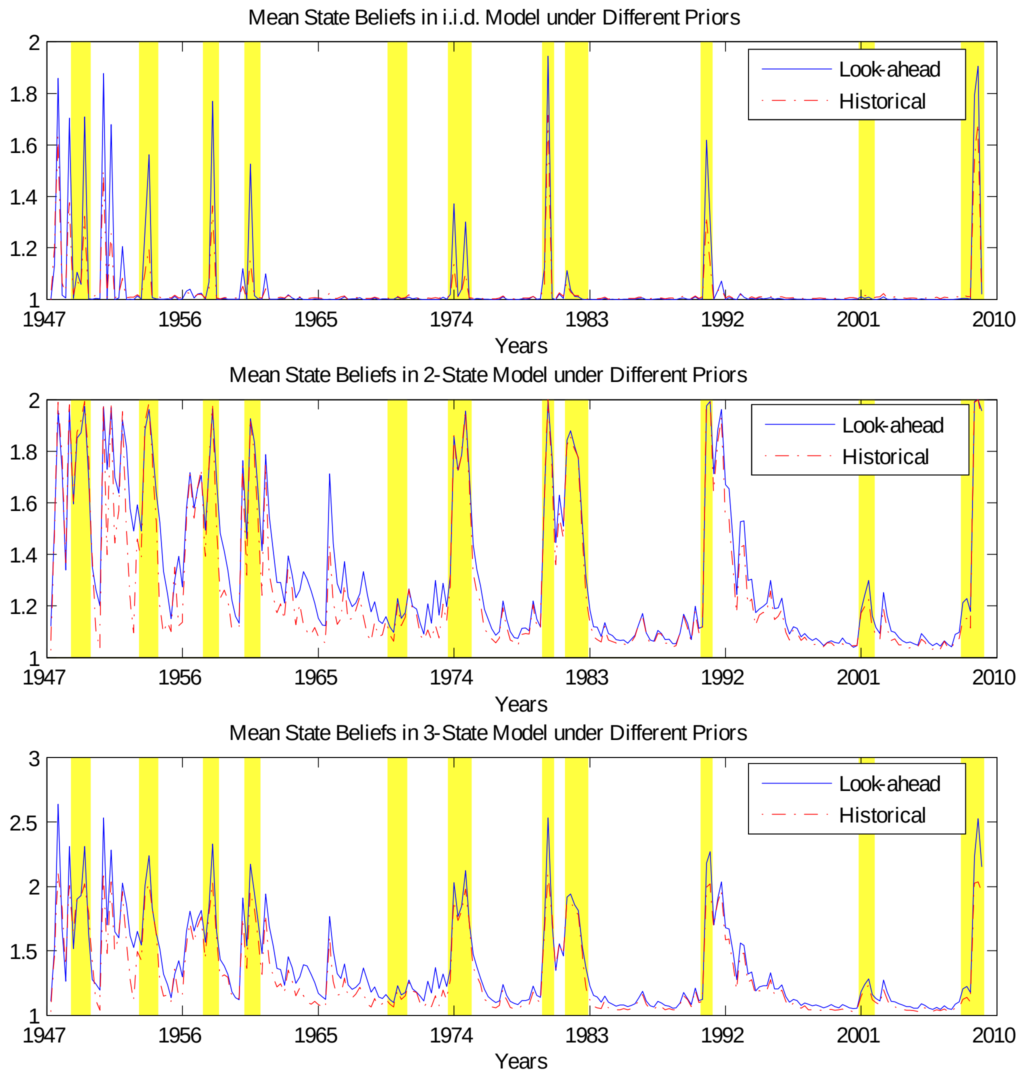


Figure 1: The plots show the means of agents' beliefs about the state of the economy at each point in time. "1" is an expansion good state, "2" is a recession state, and "3" is a disaster state. The models have either 2 or 3 states as indicated on each plot, and the time t state beliefs are formed using the history of consumption only up until and including time t. The "i.i.d. Model" is a model with i.i.d. consumption growth but that allows for jumps ("2" is a jump state). The sample is from 1947:Q2 until 2009:Q1.

Next, consider beliefs over parameters. Due to the large number of parameters and in the interests of parsimony, we focus on a few of the more economically interesting and important parameters. For the 2-state models, the top panels of Figure 2 display posterior means of the beliefs over σ_1 and σ_2. Notice that for the Historical prior the conditional volatilities slowly decrease, after a short (about 5 year) burn-in period, essentially throughout the sample. This is a combination of the Great Moderation (realized consumption volatility did decrease over the post-war sample) and the initial beliefs, which based on the historical experience expected higher consumption growth volatility. Interestingly, for the look-ahead prior, which is centered at the end of sample posterior values, the agents quickly unlearns the low full sample consumption growth volatility, and after about 5-year burn-in, the volatility is close to that observed for the historical prior. This occur because volatility was higher in the first portion of the sample. The subsequent decline in the volatility in the good state is quantitatively large (about a 30% drop).

The lower panels in Figure 2 display the transition probabilities, π_{11} and π_{22}. After the burn-in period, the first is essentially increasing over the sample, while the latter is decreasing. That is, 50 years of, on average, long expansions and high consumption growth leads to revisions in beliefs that are manifested in higher probabilities of staying in the good state and lower probabilities of staying in recession state. The probability of staying in a recession, conditional on being in a recession, goes down from about 0.85 to 0.75. Clearly, such positive shocks to the agents' perception of the data generating process will lead to higher ex post equity returns than compared to ex ante expectations.

The first three panels of Figure 3 displays estimates of the mean parameters, $E\left[\mu_i|\mathcal{M}_k, y^t\right]$ for $i = 1, 2, 3$, as well as a posterior two standard deviation band for the 3-state model using the historical prior. Learning is most apparent in the good state and least apparent in the disaster state. This is intuitive, since the economy spends most of its time in the good state and little, if any, time in the disaster state. This provides empirical evidence supporting the argument that a high level of parameter uncertainty is a likely feature of a model with a rarely observed state and is an important feature for disaster risk models (see also Chen, Joslin, and Tran, 2010).

The fourth, lower right panel shows how the speed of learning differs in the three models we consider. We use the conditional variance over the infinite horizon mean of quarterly consumption growth, $Var\left(E\left[\Delta c_{t+\infty}\right]|y^t\right)$, as a measure of the amount of parameter uncertainty (with no parameter uncertainty, the long-run mean of consumption growth is constant in all

Figure 2 - Evolution of Mean Parameter Beliefs

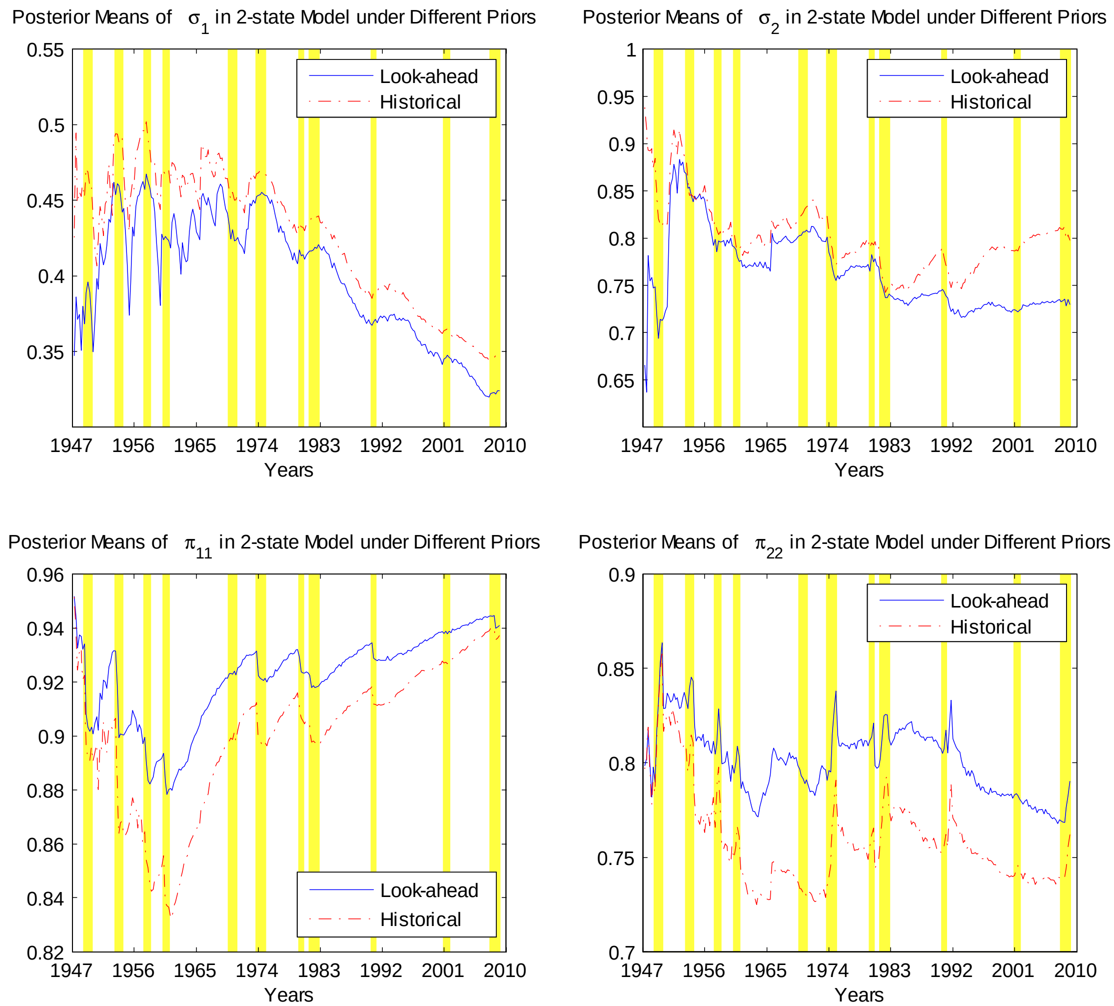


Figure 2: The two top plots in this figure show the mean beliefs of the volatility parameters within each state for the 2-state model, based on historical consumption data only. The two lower plots show the mean beliefs of the probabilities of remaining in the current state. The sample is from 1947:Q2 until 2009:Q1.

Figure 3 - Speed of Learning

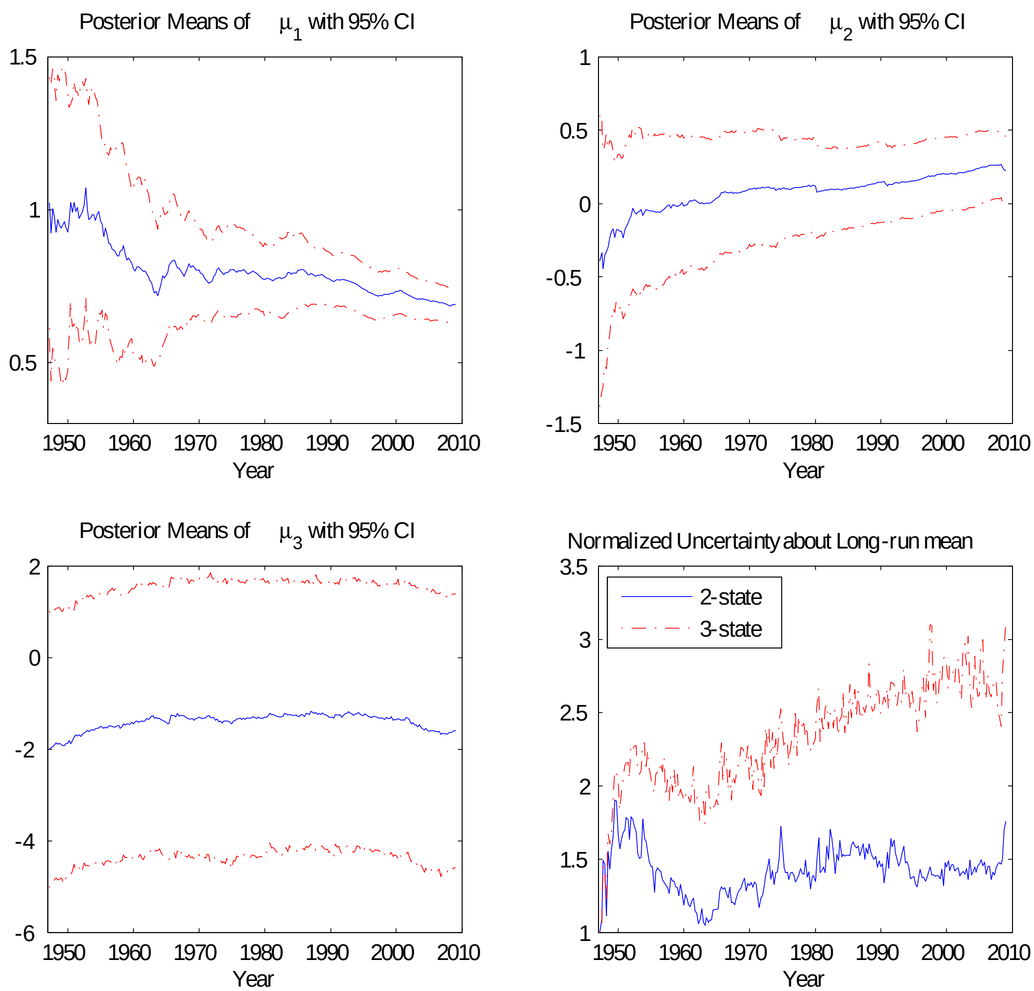


Figure 3: The first three Panels of this Figure shows the average and 2 standard deviation bounds of agents' time t beliefs about the mean parameters in the 3-state model, where only consumption data is used. The sample is from 1947:Q2 until 2009:Q1. The fourth, lower right panel shows the variance of the infinite-horizon mean of consumption growth, $Var(E[\Delta c_{t+\infty}]|y^t)$, for the 2- and 3-state models, normalized by the same for the i.i.d. model. Thus, the graph shows a measure of the relative speed of parameter learning in the models.

models), and show this variance for the unrestricted 2- and 3-state models normalized by the variance from the simpler i.i.d. model. The plot shows that learning happens faster in the simpler i.i.d. model in that both the variance ratios quickly increases. The unrestricted 2-state model settles at a variance about 50% higher than for the i.i.d. 2-state model, while the 3-state model increases its relative amount of parameter uncertainty to about 3 times that of the i.i.d. model at the end of the sample. This is due to the very slow learning about the disaster state and the difficulty present in learning the transition probabilities.

There is additional interesting time-variation in beliefs about the parameters, but this time-variation is best summarized via the total impact across all parameters, which is measured via predictive moments and discussed in the next section.

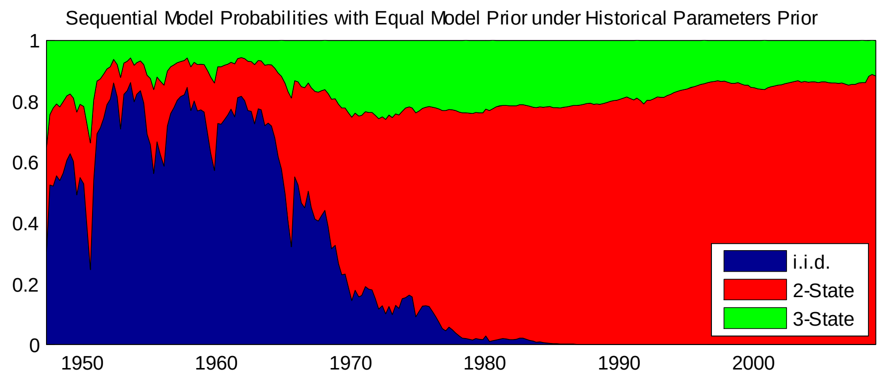
3.2 Beliefs about models and consumption dynamics

Figure 4 shows the marginal model probabilities, $p(\mathcal{M}_k|y^t)$, for each of the models we consider for the Historical and the Look-ahead priors, respectively.[21] For simplicity, the prior probability of each model was set to 1/3. Note first that the posterior probability of the i.i.d. model decreases towards zero for both priors. Thus, i.i.d. consumption growth is rejected by a Bayesian agent that updates by observing past realized consumption growth. Although not reported for brevity, this conclusion is robust even if the prior probability of the i.i.d. model is set to 0.95 - in this case it takes somewhat longer (but still just a little over half the sample) for the probability of the i.i.d. model to drop very close to zero. The 3-state model also sees a reduction in its likelihood and ends at about 10% and 20% probability levels at the end of the sample for the Historical and Look-ahead priors, respectively. The Look-ahead prior has a less severe disaster state, as it does not reflect the Great Depression, and this is why the probability of the 3-state model is higher in this case. As mentioned in the introduction, a single large negative consumption shock would quickly change these probabilities. In sum, we observe large changes in the model uncertainty over the sample.

The fact that the agent can learn that consumption growth is not i.i.d. is important. Many asset pricing models specify i.i.d. consumption growth with the implicit assumption that it is not possible or difficult to detect non-i.i.d. dynamics in consumption. Our results show that agents, using only consumption growth data, can detect non-i.i.d. dynamics, and

[21] Note that marginal model probabilities (i.e., where parameter uncertainty is integrated out) penalizes extra parameters as more sources of parameter uncertainty tends to flatten the likelihood function. Thus, it is not the case, as we see an example of here, that a 3-state model always dominates a 2-state model in Bayesian model selection.

Figure 4 - Marginal Model Probabilities



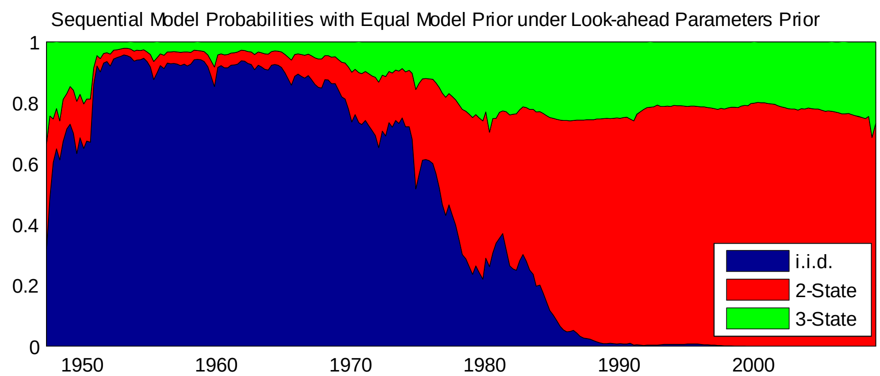


Figure 4: The top panel shows for the case of the Historical prior the evolution of the probability of each model being the true model, where the models at the beginning of the sample are set to have an equal probability, and where state and parameter uncertainty have been integrated out. The lower plot shows the same for the Look-ahead prior. The sample period is 1947:Q2 - 2009:Q1.

can do so in real time, which is an even stronger result. The agent does not need to wait until the end of the sample. This result holds for various prior specifications and is robust to time-aggregation.[22]

The results of the previous section indicate that beliefs about the parameters vary through the sample, even for the look-ahead prior, but it is not clear from this how much variation in conditional moments is present.[23] To provide asset-pricing relevant measures, we report the agent's beliefs regarding the first four moments of conditional consumption growth and model probabilities. All of these quantities are marginal, integrating out parameter, state, and/or model uncertainty. For example, the predictive mean for a given model, $\mathcal{M}_k$, is

$$E\left[\Delta c_{t+1}|\mathcal{M}_k, y^t\right] = \int \Delta c_{t+1} p\left(\Delta c_{t+1}|\theta, s_t, \mathcal{M}_k, y^t\right) p\left(\theta, s_t|\mathcal{M}_k, y^t\right) d\theta ds_t.$$

In describing these moments, we generally abstract from the first ten years and treat it is a 'burn-in' period, in order to allow the prior some time to adjust to the data, as there is some transient volatility over these first few years.

The top two panels in Figure 5 (for historical and look-ahead priors, respectively) display the conditional expected quarterly consumption growth for each model. The two and 3-state models generate relatively modest differences in this moment – both pick up business cycle fluctuations in expected consumption growth, with the 3-state model identifying the recessions in the early 80's and the financial crisis in '08 as severe. Persistent recessions are missing from the i.i.d. model, as expected. All three models exhibit a low frequency increase in expected consumption growth over the first half of the sample, due to parameter learning.

The bottom panel of Figures 5 shows *model averaged* expected quarterly consumption growth for the two priors. In the first third of the sample, the presence of the i.i.d. model smooths business cycle fluctuations in expected consumption growth. Thereafter, only the 2- and 3-state models are relevant and model uncertainty has a minor impact as the conditional expected growth is similar in these models. Overall, recessions are associated with a mean quarterly consumption growth of about 0.3%, while the mean consumption growth in

[22] In the Appendix, we show that taking out an autocorrelation of 0.25 from the consumption growth data, which is what time-aggregation of i.i.d. data predicts (see Working (1960)), does not qualitatively change these results - if anything it makes the rejection of the i.i.d. model occur sooner. The same is true if we purge the data of its full sample first order autocorrelation.

[23] As an example, consider the conditional volatility of consumption growth. A decrease in the probability of the bad state, which has higher consumption growth volatility, could be offset by an increase in the consumption volatility in the good state, σ_1, keeping the total conditional volatility of consumption growth constant.

Figure 5 - Quarterly Expected Consumption Growth

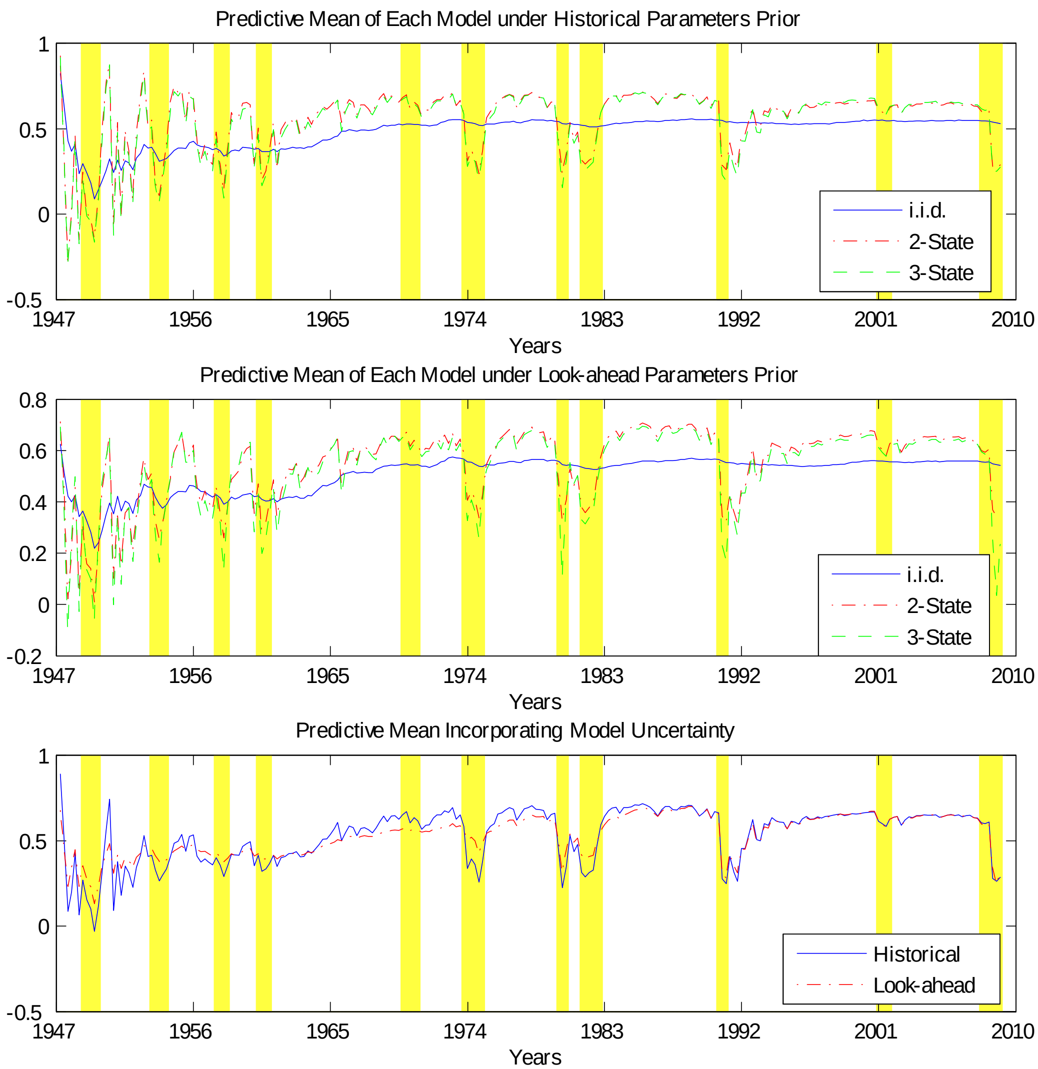


Figure 5: The top panel shows the quarterly conditional expected consumption growth, computed using the Historical Prior, from the three benchmark models: the "i.i.d. model", and hidden 2- and 3-state switching regime models. The middle plot shows the same for the Look-ahead prior. The lower plot shows the expected quarterly conditional consumption growth for both priors after model uncertainty has been integrated out. The sample period is 1947:Q2 - 2009:Q1.

expansions is about 0.6%. Since business cycles are relatively persistent, these fluctuations in conditional consumption growth are a source of long-run consumption risk, akin to that of Bansal and Yaron (2004). However, the lower frequency fluctuations we observe in expected consumption growth, which is due to parameter learning, constitute "truly" long-run risk, as shocks to parameter beliefs are permanent.

Turning to the conditional volatility of quarterly consumption growth, Figure 6 shows that for both priors there is a downward trend in consumption growth volatility through the sample, with marked increases during recessions for the non-i.i.d. models. Again, the bottom panel shows the belief about conditional standard deviation for each prior when model uncertainty is integrated out. Model probabilities could be driven by unexpected volatility, but this does not appear to be a primary determinant. Conditional consumption growth volatility is not particularly affected by model uncertainty, since both the two and the 3-state models have similar volatility patterns, and since the i.i.d. model is essentially phased out in the first third of the sample.

The secular decline is largely driven by downward revisions in estimates of the volatility parameters as realized consumption growth was less volatile in the second half of this century. This is particularly strong for the historical prior, as the conditional volatility of consumption growth decreases from about 1% per quarter to about 0.5%. Interestingly, the look-ahead prior has a similar trend, after a short burn-in period, as the prior's low consumption growth volatility is quickly unlearned, though the size of the effect is about half as large. This is the Great Moderation - the fact that consumption volatility has decreased also over the post-war sample. In the models considered here, the agent learning in real-time perceives this decrease to happen gradually, in contrast to studies that find *ex post* evidence of structural breaks or regime shifts at certain dates.

Every recession is associated with higher consumption growth volatility, although the size of the increase varies. The largest increase, on a percentage basis, occurs with the financial crisis of 2008. The increase is largest in the 3-state model, as the mean state belief at this time approaches the third state, which has a very high volatility. There is little updating about the volatility of the disaster state through the sample, since there have been no prolonged visits to this state. Thus, this reflects the fear that prevailed in the fall of 2008 that the economy was potentially headed into a depression not seen since the 1930s. This econometric result squares nicely with anecdotes from the crisis.

Figure 7 shows the time-series of conditional consumption growth skewness for the both

Figure 6 - Quarterly Consumption Growth Standard Deviation

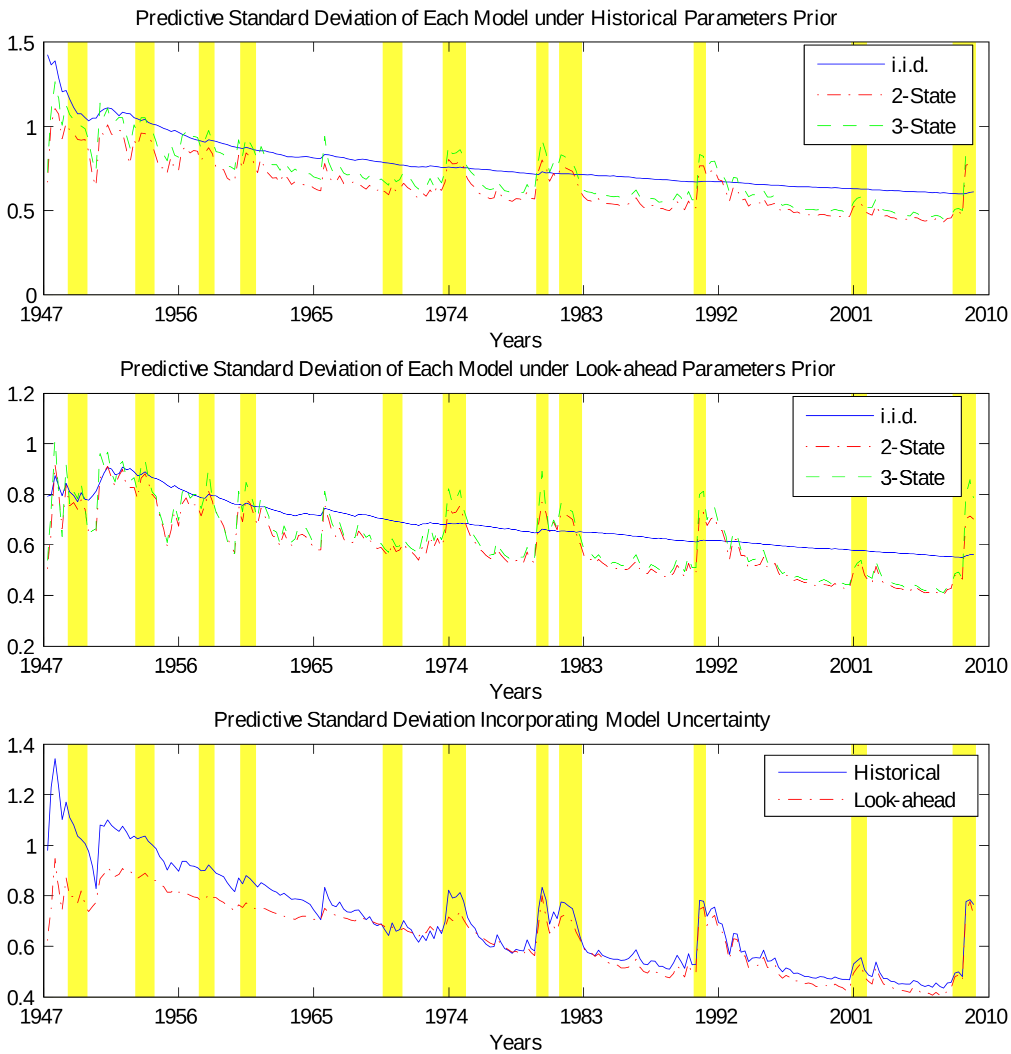


Figure 6: The top panel shows the quarterly conditional standard deviation of consumption growth, computed using the Historical Prior, from the three benchmark models: the "i.i.d. model", and hidden 2- and 3-state switching regime models. The middle plot shows the same for the Look-ahead prior. The lower plot shows the expected quarterly conditiona standard deviation of consumption growth for both priors after model uncertainty has been integrated out. The sample period is 1947:Q2 - 2009:Q1.

Figure 7 - Quarterly Consumption Growth Skewness

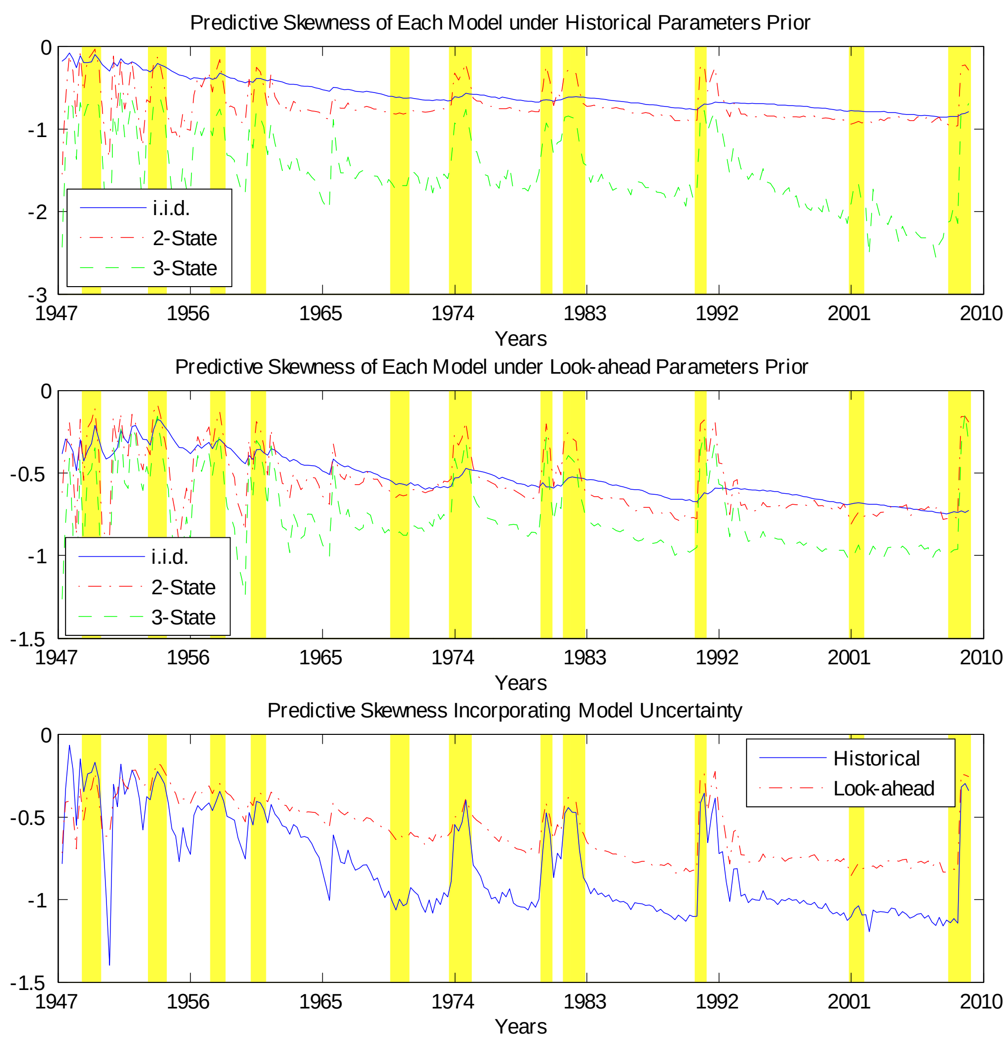


Figure 7: The top panel shows the quarterly conditional skewness of consumption growth, computed using the Historical Prior, from the three benchmark models: the "i.i.d. model", and hidden 2- and 3-state switching regime models. The middle plot shows the same for the Look-ahead prior. The lower plot shows the expected quarterly conditional skewness of consumption growth for both priors after model uncertainty has been integrated out. The sample period is 1947:Q2 - 2009:Q1.

priors, again with the model averaged estimates in the bottom panel. The time-variation in the conditional skewness is dominated by business cycle variation for the two and 3-state models, and there is a slight downward trend, as the probability of a disaster and recession decrease. When the economy is in a recession, consumption growth is naturally less negatively skewed for two reasons: (1) there is a high probability that the economy jumps to a higher (i.e. better) state and (2) expected consumption volatility is high, which tends to decrease skewness. Note that in terms of skewness, the 3-state model, with its severe recession (disaster) state, is quite different from the 2-state model. Thus model uncertainty plays a larger role for the agent's overall consumption beliefs in terms of the skewness. The 3-state model, especially for the Historical prior, strongly impacts the total perception of conditional consumption growth skewness as given in the bottom panel.

Figure 8 shows the time-series of conditional consumption growth kurtosis for the both priors. Conditional kurtosis is lower in bad states as these states are the least persistent and volatility is highest. Large, rare, outcomes are more likely when the economy is in the good state. This has potentially interesting option pricing implications (see, e.g., Backus, Chernov, and Martin (2009)), as the skewness and kurtosis will be related to volatility smiles. It is worth noting that parameter uncertainty gives an extra 'kick' to conditional skewness and kurtosis measures relative to the case of fixed parameters, where the skewness and kurtosis both move little over time (the fixed parameter case is not reported here for brevity). Both for skewness and kurtosis, there is clear evidence of parameter learning over the business cycle: the skewness becomes more negative and the kurtosis higher the longer an expansion last, reflecting updating of the transition probabilities, which reflect business cycle dynamics. Similar to skewness, there are now relatively large differences between the 2- and 3-state models. The 3-state model has significantly higher conditional kurtosis than the 2-state model, due to the presence of the disaster-state. Interestingly, the differences are greater in expansions than in recessions, again due to the 'rare' nature of recessions and, especially, disasters. In terms of the conditional kurtosis after model uncertainty is integrated out (bottom panel), the 3-state model has large impact on kurtosis even at the end of the sample where the probability of this model being the right model is low. Thus, among the models considered here, model uncertainty and its dynamic behavior is likely to have the strongest implications for assets such as out-of-the-money options that are more sensitive to the tail behavior of consumption growth.

Figure 8 - Quarterly Consumption Growth Kurtosis

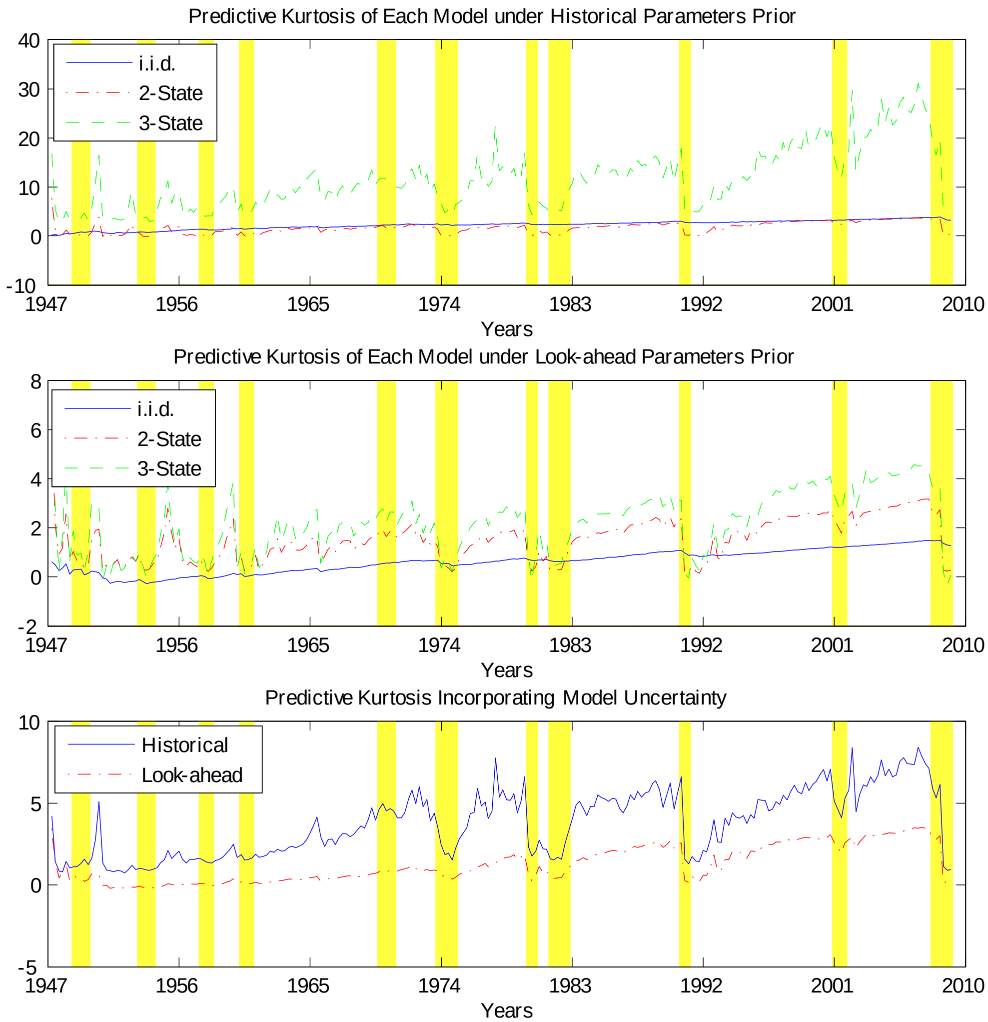


Figure 8: The top panel shows the quarterly conditional expected consumption growth, computed using the Historical Prior, from the three benchmark models: the "i.i.d. model", and hidden 2- and 3-state switching regime models. The middle plot shows the same for the Look-ahead prior. The lower plot shows the expected quarterly conditional skewness of consumption growth for both priors after model uncertainty has been integrated out. The sample period is 1947:Q2 - 2009:Q1.

4 Does learning matter for asset prices?

4.1 A new test for the importance of learning

The previous results indicate that the agent's beliefs – about parameters, moments, and models – vary substantially at both very low frequencies and over the business cycle. If learning is an important determinant of asset prices, changes in beliefs should be a significant determinant of asset returns. This is a fundamental test of the importance of learning about the consumption dynamics. For example, if agents learn that expected consumption growth is higher than previously thought, this revision in beliefs will be reflected in the aggregate wealth-consumption ratio (if the elasticity of intertemporal substitution is different from one). In particular, if the substitution effect dominates, the wealth-consumption ratio will increase when agents revise their beliefs about the expected consumption growth rate upwards (see, e.g., Bansal and Yaron (2004)). As another example, if agents learn that aggregate risk (consumption growth volatility) is lower than previously thought, this will generally lead to a change in asset prices as both the risk premium and the risk-free rate are affected. In the Bansal and Yaron (2004) model, an increase in the aggregate volatility leads to a decrease in the stock market's price-dividend ratio.

To test this, we regress excess quarterly stock market returns (obtained from Kenneth French's web site) on changes in beliefs about expected consumption growth and expected consumption growth variance. This is a particularly stringent test of learning, which to our knowledge has not been done in the previous literature. We use the beginning of period timing for the consumption data here and elsewhere in the paper.[24] The regressors are the shocks, $E_t(\Delta c_{t+1}) - E_{t-1}(\Delta c_{t+1})$ and $\sigma_t(\Delta c_{t+1}) - \sigma_{t-1}(\Delta c_{t+1})$. Notice that the only thing that is changing is the conditioning information set as we go from time $t-1$ to time t; the regressors are revisions in beliefs. We calculate these conditional moments for each prior integrating out state, model and parameter uncertainty. The first 10 years of the sample are used as a burn-in period to alleviate any prior misspecification (there is some excess volatility in state and parameter beliefs in these first years).

Separate regressions are run for the historical and look-ahead priors, and we control for

[24] Due to time-averaging (see Working, 1960), Campbell (1999) notes that one can use either beginning of period or end of period consumption in a given quarter as the consumption for that quarter. The beginning of period timing yields stronger results than using the end of period convention (although the signs are the same in the regressions). In principle, the results should be the same, so this is consistent with some information being impounded in stocks before the consumption data is revealed to the Bureau of Economic Analysis.

contemporaneous consumption growth and lagged consumption growth (the direct cash flow effect). By controlling for realized consumption growth, we ensure that the results are driven by model-based revisions in beliefs, and not just the fact that realized consumption growth (a direct cash flow effect) was, for example, unexpectedly high. To separate out the effects of parameter from state learning, we use revisions in expected consumption growth beliefs computed from the 3-state model with fixed parameters (set to their full-sample values) as an additional control.[25]

Specifications 1 and 2 in Panel A (historical prior) and Panel B (look-ahead prior) in Table 1 show that increases in expected conditional consumption growth are positively and strongly significantly associated with excess contemporaneous stock returns for both priors. This result holds controlling for contemporaneous and lagged consumption growth (the direct cash flow effect), and so we can conclude that revisions in beliefs are significantly related to shocks to the price-dividend ratio. This is a very strong result, pointing to the importance of a learning-based explanation for realized stock returns. These results could be driven by parameter or state learning.

Specification 3 shows that the updates in expected consumption growth derived from the model with fixed parameters (that is, a case with state learning only) are also significantly related to realized stock returns. The R^2, however, is lower than for the case of the full learning model, and when we include the revisions in beliefs about expected consumption growth from both the full learning model and the fixed parameters benchmark model in the regression (specification 4), the updates in expected consumption growth that arise in a model with fixed parameters are insignificant, while the belief revisions from the full learning model remain significant. That is, updates in expectations when learning about about parameters, states, and models are more closely related to realized stock market returns than the corresponding updates in expectations based on a single model with known parameters but hidden states estimated on the full sample. To our knowledge, this is the first direct comparison of learning about models and parameters versus the traditional implementation of the rational expectation explanations in terms of explaining the time-series of realized stock returns using the actual sequence of realized macro shocks.

This result is driven by the nonlinear process of jointly learning about parameters and states. In particular, specification 5 shows that updates in beliefs from the i.i.d. model cannot be distinguished from the direct cash flow effect. The i.i.d. model captures parameter uncertainty about the long-run mean and variance, but not the state dynamics. The fixed

[25] Using the fixed parameter 2-state model as the control instead does not change the results.

Table 1 - Updates in Beliefs versus Realized Stock Returns

Table 1: The table shows the results from regressions of innovations in agents' expectations of future consumption growth ($E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$) and conditional consumption growth variance ($\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$) versus contemporaneous excess stock market returns. In calculating the expectations, the parameter and model uncertainty is integrated out. The controls are lagged and contemporaneous realized log consumption growth, as well as the innovation in expected consumption growth derived from the 3-state model with fixed parameters (i.e., no model or parameter uncertainty), as well as the i.i.d. model with uncertain parameters. Panel A shows the results for the Historical priors, while Panel B shows the results for the Look-ahead priors. Heteroskedasticity and autocorrelation adjusted (Newey-West; 3 lags) standard errors are used. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. The sample is from 1947:Q2 until 2009:Q1. In the below regressions, we have removed the first 40 observations (10 years), as a burn-in period to alleviate misspecification of the priors.

Dependent variable: $r_{m,t+1} - r_{f,t+1}$ (excess market returns)

Panel A: Historical prior	1	2	3	4	5	6	7
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	40.43***	26.00**		42.41**			
	(9.36)	(11.44)		(18.97)			
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$						−36.34***	−13.83
						(10.90)	(10.14)
Controls:							
Δc_{t+1}		2.02			7.94**		3.76***
		(1.51)			(3.12)		(1.37)
Δc_t		2.31*			1.73		2.05
		(1.41)			(1.43)		(1.43)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta\text{ known}}$			24.98***	−1.76			
			(9.36)	(12.82)			
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{i.i.d. model}}_{\theta\text{ unknown}}$					−449.17		
					(392.94)		
R^2_{adj}	8.8%	10.9%	5.9%	8.4%	9.5%	5.0%	9.7%

Panel B: Look-ahead prior	1	2	3	4	5	6	7
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	52.44***	29.05**		39.63**			
	(11.71)	(13.62)		(18.42)			
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$						−46.10***	−23.26*
						(12.53)	(13.43)
Controls:							
Δc_{t+1}		2.77*			7.78**		3.51**
		(1.56)			(3.18)		(1.42)
Δc_t		2.25*			1.74		2.16
		(1.38)			(1.43)		(1.39)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta\text{ known}}$			24.98***	8.77			
			(9.36)	(10.62)			
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{i.i.d. model}}_{\theta\text{ unknown}}$					−427.05		
					(400.75)		
R^2_{adj}	7.3%	10.5%	5.9%	7.2%	9.5%	5.3%	10.2%

parameter model (specification 4) captures the transitory state learning, but not the parameter dynamics.[26] Thus, it is the updates in beliefs stemming from the more complicated, non-i.i.d. models' learning problem that drives the increased correlation with stock returns, relative to the direct cash flow effect. Recall also that our agent quickly learned that the i.i.d. model is not likely, relative to the other specifications.

For the variance (regression specifications 6 and 7 in Table 1) we get the opposite result, as one would expect (at least with a high elasticity of intertemporal substitution, as we will use later in the paper): unexpected increases in conditional consumption growth variance are associated with negative contemporaneous stock returns. This result is not significant at the 5% level when including contemporaneous and lagged consumption growth in the regressions (specification 7). This does not mean there is no effect; we just cannot distinguish it from the direct cash flow effect when learning from consumption data alone.

To summarize, we find strong evidence that the updates in beliefs elicited from our model/prior combinations are associated with actual updates in agent beliefs at the time, as proxied by stock market returns. Again, it is important to recall that no asset price data was used to generate these belief revisions.

4.2 Learning from additional macro variables

Agents have access to more than just aggregate consumption growth data when forming beliefs. Here we provide one approach for incorporating this additional information and apply this methodology to learning from quarterly GDP growth, in addition to consumption. Suppose x_t represents the common growth factor in the economy and evolves via:

$$x_t = \mu_{s_t} + \sigma_{s_t}\varepsilon_t, \tag{3}$$

where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and s_t is the state of the economy, which follows the same Markov chains specified earlier. Consumption growth Δc and J additional variables $Y_t = [y_t^1, y_t^2, ..., y_t^J]'$ are assumed to follow:

$$\Delta c_t = x_t + \sigma_c \varepsilon_t^c, \tag{4}$$

where

[26] One can show analytically that in a simple i.i.d. model, updates in expectations of consumption growth are very close to linear in the realized consumption growth.

$$y_t^j = \alpha_j + \beta_j x_t + \sigma_j \varepsilon_t^j, \quad for \quad j = 1, 2, ..., J \tag{5}$$

and $\varepsilon_t^c \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and $\varepsilon_t^j \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ for any j. Note that the coefficients in equation (5) are not state dependent, which implies that the additional variables will primarily aid in state identification. The specification allows for the additional observation variables to be stronger or weaker signals of the underlying state of the economy than consumption growth. For the case of GDP growth, this captures the idea that investment is more cyclical than consumption, which makes GDP growth a better business cycle indicator. The linearity of the relationship is an assumption that is needed for conjugate priors.

The similar conjugate priors for the parameters are applied. For each state $s_t = i$, $p(\mu_i|\sigma_i^2)p(\sigma_i^2) \sim \mathcal{NIG}(a_i, A_i, b_i, B_i)$, where $\mathcal{NIG}$ is the normal/inverse gamma distribution. σ_c is assumed to follow an inverse gamma distribution $\mathcal{IG}(b_c, B_c)$, and for each $j = 1, 2, ..., J$, $p([\alpha_j, \beta_j]'|\sigma_j^2)p(\sigma_j^2) \sim \mathcal{NIG}(a_j, A_j, b_j, B_j)$, where $p([\alpha_j, \beta_j]'|\sigma_j^2)$ is a bivariate normal distribution $\mathcal{N}(a_j, A_j\sigma_j^2)$, a_j is a 2×1 vector and A_j is a 2×2 matrix. Particle filtering is straightforward to implement in this specification by modifying the algorithm described in the Appendix.

To analyze the implications of additional information, we consider learning using real, per capita U.S. GDP growth as an additional source of information. This exercise generates a battery of results: time series of parameter beliefs, conditional moments, and model probabilities. We report only a few interesting statistics in the interests of parsimony. Figure 9 shows that the state beliefs do not change dramatically, although GDP growth is typically thought of as more informative about business cycle fluctuations than consumption growth. To characterize how the additional data aids in state identification, we compute posterior standard deviations for the states, $std\,[s_t|\mathcal{M}_k, y^t]$, again integrating out parameter uncertainty. The top Panel of Figure 10, shows that indeed the uncertainty about the state is much lower (about half) than what was the case when using consumption growth as the only source of information. Thus, adding GDP growth to the agent's information set increases the precision of the state identification.[27] The increased certainty about the state improves parameter identification also, which is confirmed in the two lower Panels in Figure 10. Here the uncertainty about the good and bad states mean consumption growth rates is lower, after a 10-year burn-in, than in the case using consumption as the only source of information.

[27] It is technically feasible to impose cointegration between consumption and GDP by including the log consumption to GDP ratio on the right hand side of Equation (5). We thank Lars Hansen for pointing this out.

Figure 9 - Mean State Beliefs (GDP)

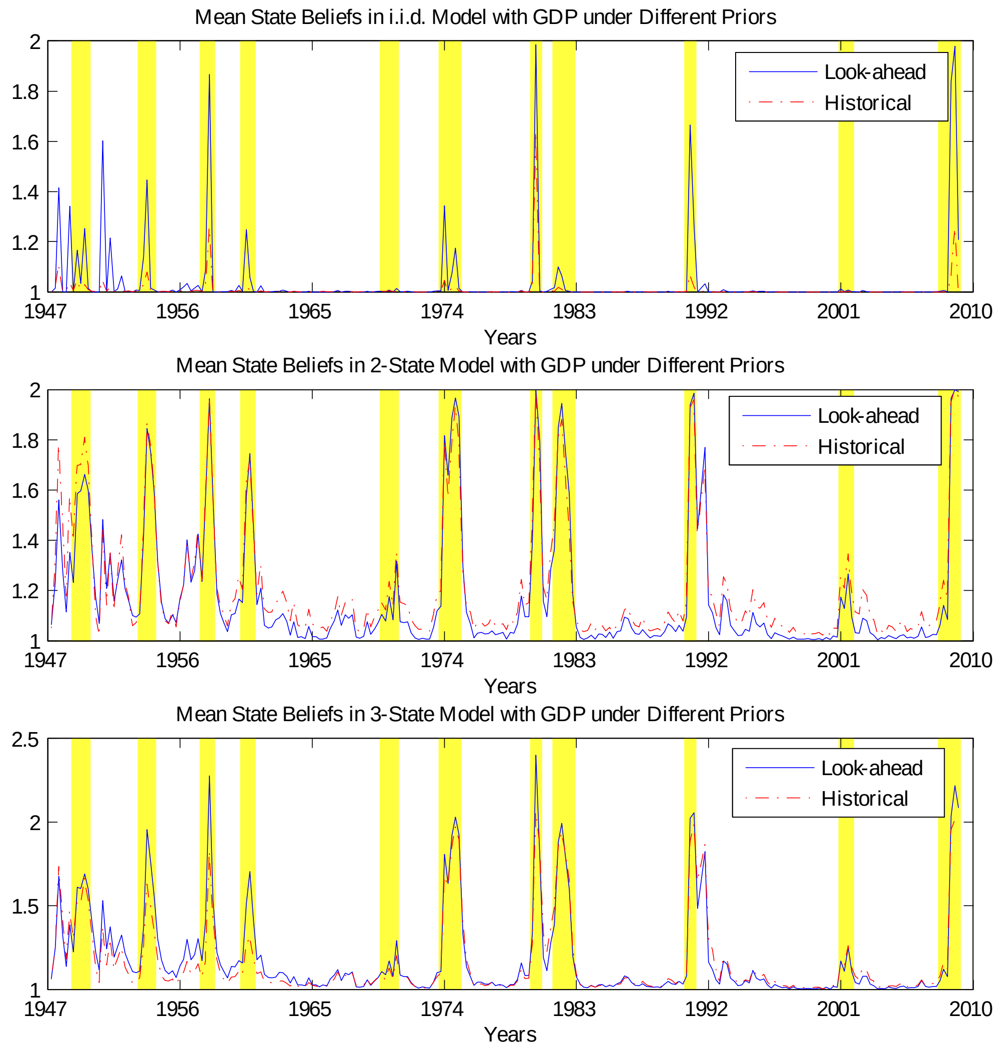


Figure 9: The figures show the means of agents' beliefs about the state of the economy at each point in time. "1" is an expansion good state, "2" is a recession state, and "3" is a disaster state. The models have either 2 or 3 states as indicated on each plot, and the time t state beliefs are formed using the history of both consumption and GDP up until and including time t. The sample is from 1947:Q2 until 2009:Q1.

Figure 11 shows that the model specification results are similar, as the data again favors the 2-state model, leaving the 3-state model with a very low probability at the end of the sample. It is noteworthy, however, that the probability of the 3-state (disaster) model again increases at the onset of the financial crisis in 2008.

Adding GDP growth also results in a greater difference in expected consumption growth across the states. Figure 12 shows that the difference in the expected consumption growth rate in recessions versus expansions is about 0.6% per quarter, versus about 0.3% in the case of consumption information only (see Figure 5). The dynamic behavior of the conditional standard deviation of consumption growth is not significantly changed (not reported for brevity).

Table 2 shows the regressions of contemporaneous stock returns and updates in agent beliefs about conditional expected consumption growth and consumption growth variance, as calculated from this extended model. The results are similar, but in fact overall *stronger* than the results using only consumption growth. Updates in agent expectations about these moments from the full learning model are significantly related to stock returns, also after controlling for contemporaneous and lagged consumption growth and updates in expected consumption growth derived from a model with fixed parameters. Again, this evidence indicates that learning about parameters and models is an important feature of the data.

4.3 Additional asset pricing implications

We now embed the beliefs of our learning agent in a general equilibrium asset pricing model. There are considerable computational and technical issues that need to be dealt with when considering such an exercise. First, the state space is prohibitively large. The 3-state model, as an example, have 12 parameters governing the exogenous consumption process, and the beliefs over each parameter are governed by 2 hyper-parameters. Thus, there are 24 state variables, in addition to beliefs over the state of the economy and the corresponding parameter and state beliefs for the i.i.d. and the general 2-state models. Second, as pointed out by Geweke (2001) and Weitzmann (2007), some parameter distributions must be truncated in order for utility to be finite. This introduces additional nuisance parameters.

Given the computational impediments, we follow Sargent and Cogley (2008) and Piazzesi and Schneider (2010) and apply the principle of "anticipated utility" to the pricing exercise (originally suggested by Kreps (1998)). Under this assumption, the agents maximize utility

Figure 10 - Uncertainty about state (GDP)

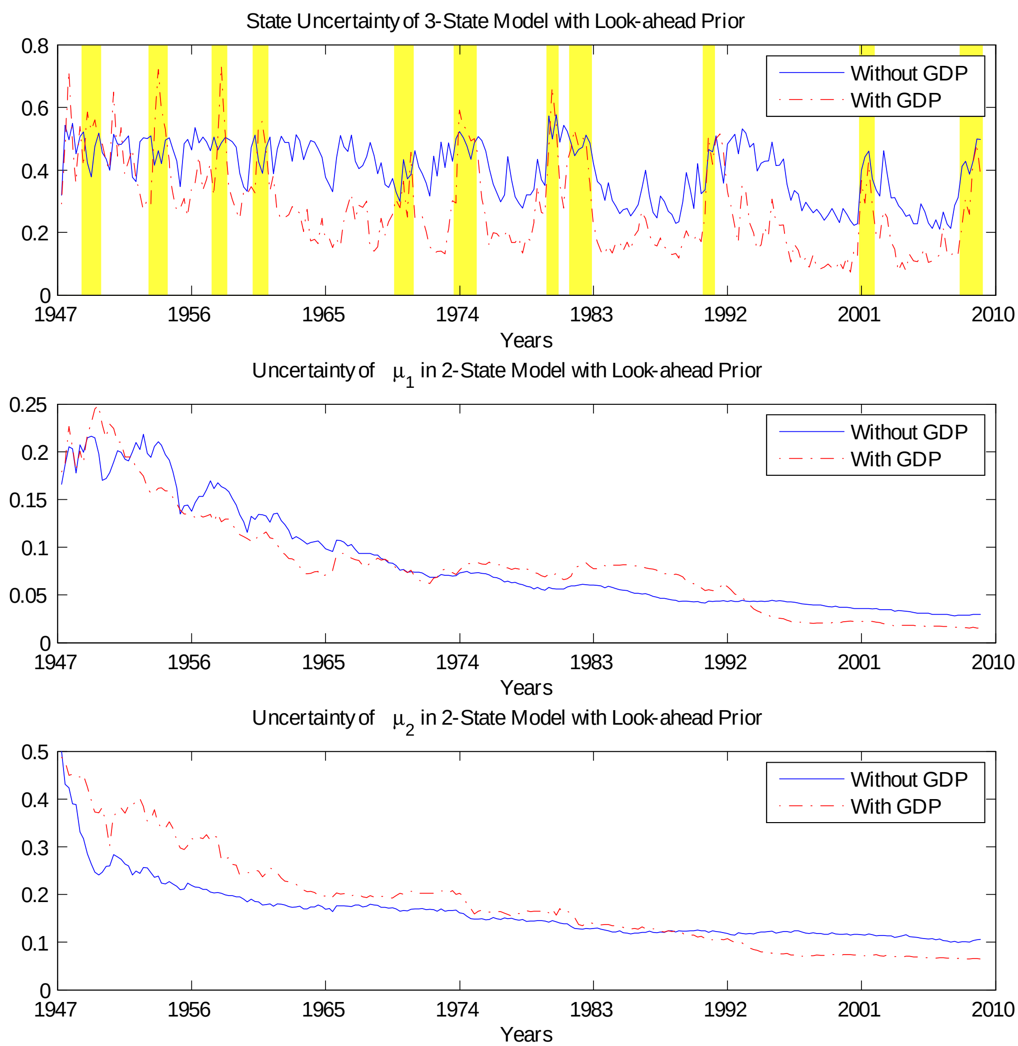


Figure 10: The top Palen shows the standard deviation of the posterior belief about the states for the case of Look-ahead priors when the consumption dynamics are estimated using consumption data only versus the consumption and GDP data. The two lower Panels show the standard deviation of posterior beliefs about the mean in the expansion and the recession states, respectively, for the 2-state model, Look-ahead prior. The solid line gives the case where agents learn from consumption growth only, while the dashed line shows the case of learning from both consumption and GDP growth. The sample is from 1947:Q2 until 2009:Q1.

Figure 11 - Model Probabilities(GDP)

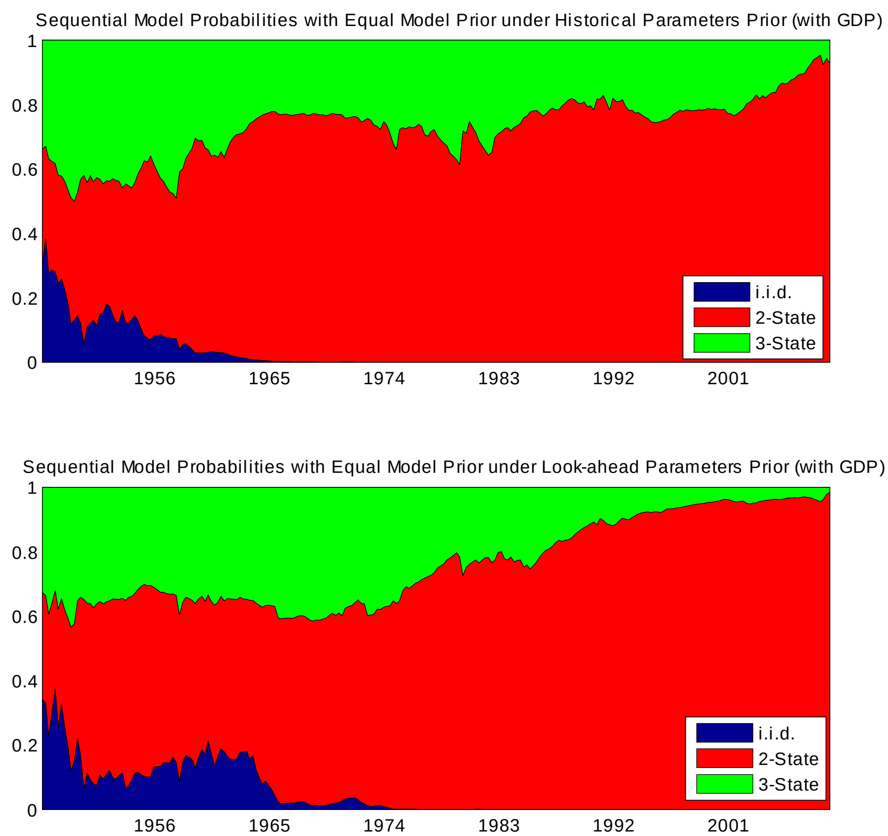


Figure 11: The top panel shows for the case of the Historical prior the evolution of the probability of each model being the true model, where the models at the beginning of the sample are set to have an equal probability, and where state and parameter uncertainty have been integrated out. In this case, agents also use GDP growth to learn about the state of the economy. The lower plot shows the same for the Look-ahead prior. The sample period is 1947:Q2 - 2009:Q1.

Figure 12 - Conditional expected consumption growth (GDP)

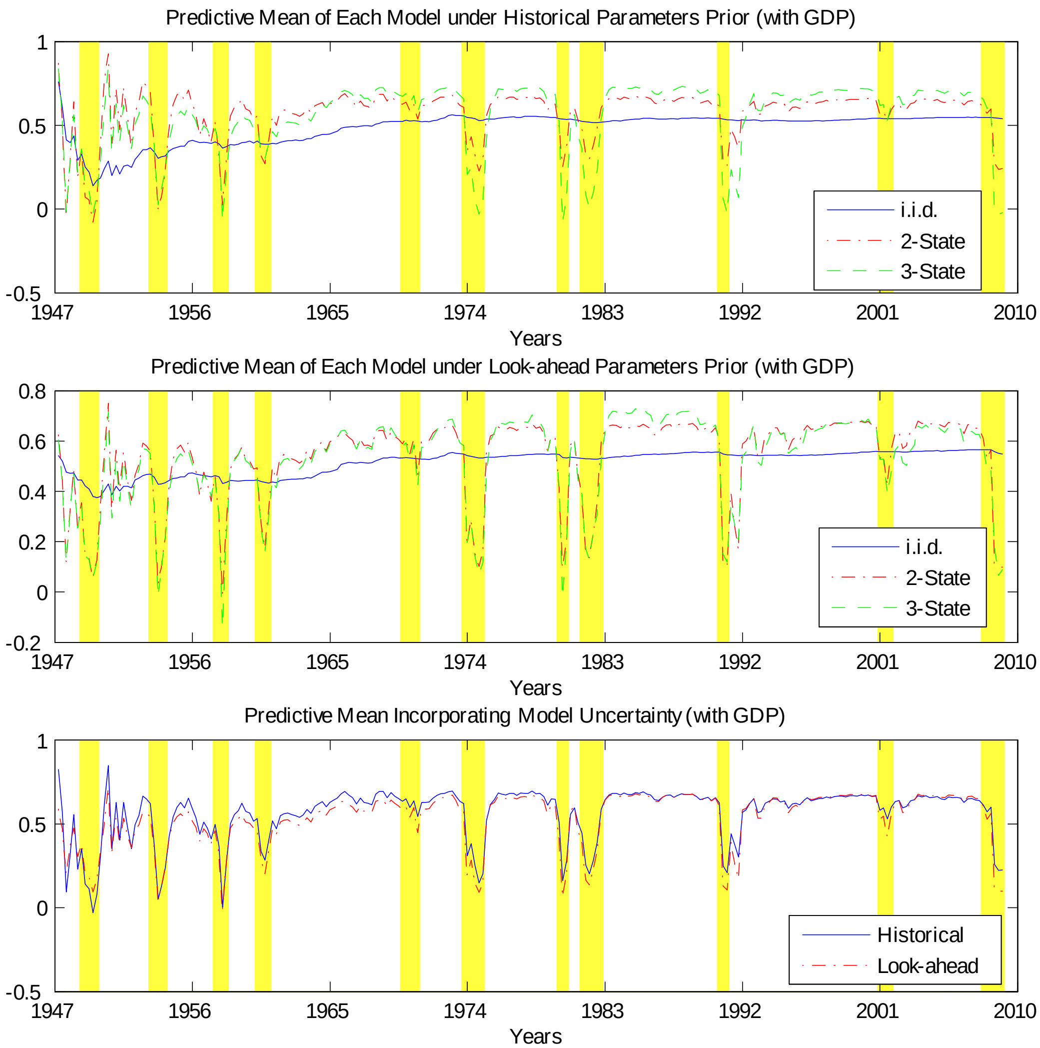


Figure 12: The top panel shows the quarterly conditional expected consumption growth, computed using the Historical Prior, from the three benchmark models: the "i.i.d. model", and hidden 2- and 3-state switching regime models. In this case, agents also use GDP growth to learn about the state of the economy. The middle plot shows the same for the Look-ahead prior. The lower plot shows the expected quarterly conditional consumption growth for both priors after model uncertainty has been integrated out. The sample period is 1947:Q2 - 2009:Q1.

Table 2 - Updates in Beliefs versus Realized Stock Returns (GDP)

Table 2: The table shows the results from regressions of innovations in agents' expectations of future consumption growth ($E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$) and conditional consumption growth variance ($\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$) versus contemporaneous excess stock market returns. In calculating the expectations, the parameter and model uncertainty is integrated out. The controls are lagged and contemporaneous realized log consumption growth, as well as the innovation in expected consumption growth derived from the 3-state model with fixed parameters (i.e., no model or parameter uncertainty). Both consumption and GDP data is used to estimate the models, as described in the main text. Panel A shows the results for the Historical priors, while Panel B shows the results for the Look-ahead priors. Heteroskedasticity and autocorrelation adjusted (Newey-West; 3 lags) standard errors are used. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. The sample is from 1947:Q2 until 2009:Q1. In the below regressions, we have removed the first 40 observations (10 years), as a burn-in period to alleviate misspecification of the priors.

Dependent variable: $r_{m,t+1} - r_{f,t+1}$ (excess market returns)

Panel A: Historical Prior	1	2	3	4	5
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	40.68^{***}	40.52^{***}	39.77^{**}		
	(6.62)	(8.99)	(19.13)		
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$				-56.94^{***}	-46.25^{***}
				(10.79)	(22.37)
Controls:					
Δc_{t+1}		-0.70			1.02
		(1.56)			(1.52)
Δc_t		1.93			2.36^{*}
		(1.33)			(1.41)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta\text{ known}}$			0.60		
			(10.47)		
R^2_{adj}	15.4%	15.6%	15.0%	11.9%	13.3%

Panel B: Look-ahead Prior	1	2	3	4	5
$E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]$	33.48^{***}	30.84^{***}	28.41^{**}		
	(5.56)	(7.24)	(14.03)		
$\sigma^2_{t+1}[\Delta c_{t+2}] - \sigma^2_t[\Delta c_{t+2}]$				-67.93^{***}	-48.59^{**}
				(13.76)	(17.94)
Controls:					
Δc_{t+1}		0.01			1.88
		(1.40)			(1.69)
Δc_t		2.11^{*}			2.50^{*}
		(1.33)			(1.41)
$[E_{t+1}[\Delta c_{t+2}] - E_t[\Delta c_{t+2}]]^{\text{3-state model}}_{\theta\text{ known}}$			3.89		
			(9.20)		
R^2_{adj}	14.5%	15.0%	14.1%	9.3%	11.9%

at each point in time assuming that the parameters and model probabilities are equal to the agents' current mean beliefs and will remain constant forever. Of course, at time $t+1$ the mean parameter beliefs will in general be different due to learning. While parameter and model uncertainty are not *priced* risk factors in this framework, they are nonetheless important for the time-series of asset prices as updates in mean parameter and model beliefs lead to changes in prices. We do integrate out state uncertainty in the pricing exercise, so state uncertainty is a priced risk factor (as in, e.g., Lettau, Ludvigson, and Wachter (2008)). The anticipated utility approach reduces the number of state variables to three (the belief about the state in the general 2-state model, and the 2-dimensional belief about the state in the 3-state model).[28]

The purpose of the pricing exercise is to examine what features of the post-WW2 U.S. aggregate consumption and asset price data a realistic, general learning problem can help explain. Since we do not integrate out the parameter and model uncertainty in the pricing exercise, we focus on two aspects of the model that are likely to be robust to the introduction of *priced* parameter and model uncertainty.

1. *Ex-ante* versus *ex post*

   With learning *ex ante* expectations need not in general equal average *ex post* outcomes, which is the assumption in the typical rational expectations implementation. In the following, we argue that substantial components of the observed equity premium, excess return volatility, the degree of in-sample excess return predictability, and the time-series of the aggregate price-dividend ratio can be explained by the (nonstationary) time-path of mean parameter beliefs.

2. Permanent versus transitory shocks

   The shocks to mean parameter beliefs are permanent shocks to investor information sets. This has implications for, for instance, the volatility of long-run bond yields, and is different from a model with transitory shocks to state variables (such as our state beliefs, the long-run risk variable in Bansal and Yaron (2004), or the surplus consumption ratio in Campbell and Cochrane (1999)).

[28] It would be computationally feasible to account for model uncertainty or to focus on parameter uncertainty over one of the parameters, but we leave such considerations for future research.

4.3.1 The model

The model is solved at the quarterly frequency, and the representative agent is assumed to have Epstein and Zin (1989) preferences, which are defined recursively as:

$$U_t = \left\{ (1-\beta) C_t^{1-1/\psi} + \beta \left(E_t \left[U_{t+1}^{1-\gamma} \right] \right)^{\frac{1-1/\psi}{1-\gamma}} \right\}^{\frac{1}{1-1/\psi}}, \tag{6}$$

where C_t is the consumption, $\psi \neq 1$ is the intertemporal elasticity of substitution (IES) in consumption, and $\gamma \neq 1$ is the coefficient of relative risk aversion. These preferences imply the stochastic discount factor:

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \left(\beta \frac{PC_{t+1} + 1}{PC_t} \right)^{\frac{1/\psi - \gamma}{1-1/\psi}}, \tag{7}$$

where PC_t is the wealth-consumption ratio – that is, the price-dividend ratio for the claim to the stream of future aggregate consumption. The first component of the pricing kernel is that which obtains under standard power utility, while the second component is present if the agent has a preference for the timing of the resolution of uncertainty (i.e., if $\gamma \neq 1/\psi$). As mentioned earlier, we consider an anticipated utility approach to the pricing problem in terms of parameter and model uncertainty, while state uncertainty is priced.[29] This corresponds to a world where investors understand and account for business cycle fluctuations, but where they simply use their best guess for the parameters governing these dynamics.

Our goal in this section is to, for reasonable preference parameters, understand how learning affects pricing relative to the benchmark case of fixed parameters. Given that the consumption dynamics are not *ex post* calibrated (in particular in the historical prior case) but estimated in real-time, we also do not calibrate preference parameters to match any particular moment(s). Instead, we simply use the preference parameters of Bansal and Yaron (2004). Thus, $\gamma = 10$, $\psi = 1.5$, and $\beta = 0.998\hat{}3$.

Following both Bansal and Yaron (2004) and Lettau, Ludvigson, and Wachter (2008), we

[29] The model is solved numerically through value function iteration *at each time t* in the sample, conditional on the mean parameter beliefs at time t, which gives the time t asset prices. The state variables when solving this model are the beliefs about the hidden states of the economy for each model under consideration. For a detailed description of the model solution algorithm, please refer to the Appendix.

Cogley and Sargent (2009) argue that anticipated utility approach is a close approximation to the true Bayesian approach, although their analysis is with respect to time-separable preferences. Piazzesi and Schneider (2010) is an example of a recent application of an anticipated utility pricing framework with Epstein-Zin preferences.

price a levered claim to the consumption stream with a leverage factor λ of 4.5. The annual consumption volatility over the post-war sample is only 1.34%, and so the systematic annual dividend volatility is therefore about 6%. Quarterly log dividend growth is defined as:

$$\Delta d_t = \lambda \Delta c_t + \varepsilon_{d,t}, \tag{8}$$

where $\varepsilon_{d,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(-\frac{1}{2}\sigma_d^2, \sigma_d^2\right)$ is the idiosyncratic component of dividend growth. σ_d is chosen to match the observed annual 11.5% volatility of dividend growth reported in Bansal and Yaron (2004). With these choices of λ and σ_d we also in fact closely match the sample correlation they report between annual consumption and dividend growth (0.55).[30]

Unconditional Moments Table 3 reports realized asset pricing moments in the data, and also those generated by our learning models over the same sample period. The first 10 years are removed as a burn-in period to reduce concerns with regards to prior misspecification. We consider cases with and without parameter learning.

The models with parameter uncertainty match the observed equity premium reasonably well: 4.7% in the data versus 3.8% and 3.4% for the consumption only historical and look-ahead priors, respectively. The models where GDP is used as an additional signal, which as reported earlier have a more severe recession state, have average sample excess equity returns of 4.2% and 4.0% for the historical and the look-ahead priors, respectively. This compares favorably to the benchmark fixed parameters two and 3-state models which sample equity premiums are 1.5% and 1.8%, respectively. Thus, allowing for parameter uncertainty more than doubles the sample risk premiums, despite the fact that parameter and model uncertainty are not priced risk factors in the anticipated utility pricing framework. The high sample equity premium arises because of the specific time path of beliefs, which we discuss next.

The table also reports the average *ex ante* equity risk premium ($E_T\left[E\left(R^{excess}_{m,t+1}|I_t\right)\right]$, where I_t denotes the information set (beliefs) of agents at time t and $E_T[\cdot]$ denotes the sample average). The cases with parameter and model learning have about the same *ex ante* risk premium. This implies that more than half of the excess returns achieved in these

[30] The dividend dynamics imply that consumption and dividends are not cointegrated, which is a common assumption (e.g., Campbell and Cochrane (1999), and Bansal and Yaron (2004)). One could impose cointegration between consumption and dividends, but at the cost of an additional state variable. Further, it is possible to also learn about λ and σ_d^2. However, quarterly dividends are highly seasonal, which would severely complicate such an analysis. Further, data on stock repurchases is mainly annual. We leave a rigorous treatment of these issues to future research.

Table 3 - Asset Price Moments

Table 3: The table reports the asset pricing implications of the models with an anticipated utility version of the Epstein-Zin preferences under different priors, as well as the fixed parameters cases. For all the models, $\gamma = 10$, $\beta = 0.994$, $\psi = 1.5$, $\lambda = 4.5$. The volatility of the idiosyncratic component of dividend growth ($\epsilon_{d,t}$) is calibrated to match the historical standard deviation of dividend growth, as reported in Bansal and Yaron (2004). The statistics are annualized. The expectation operator with a T subscript, E_T, denotes the sample average, while the volatility operator, σ_T denotes the sample standard deviation. 'Cons. only' denotes the model results in the case where only consumption growth is used to update beliefs, while 'Cons. + GDP' denotes the model results in the case where both consumption and GDP growth are used to update beliefs. The full sample period is from 1947:Q2 until 2009:Q1. However, we have removed the first 40 observations (10 years), as a burn-in period to alleviate misspecification of the priors. Similar results are obtained with no burn-in or with 80 quarter burn-in.

	Data	Historical prior		Look-ahead prior		Fixed parameters	
Moments	1957:Q2-2009:Q1	Cons. only	Cons. + GDP	Cons. only	Cons. + GDP	2-state model	3-state model
The real risk-free rate:							
$E_T(r_t^f)$	1.6%	3.8%	3.7%	3.7%	3.7%	3.7%	3.7%
$\sigma_T(r_t^f)$	1.6%	0.8%	0.9%	0.6%	0.8%	0.7%	0.8%
The dividend claim: $d_t = \lambda c_t + \varepsilon_{d,t}$							
ex post:							
$E_T(r_t - r_t^f)$	4.7%	3.8%	4.2%	3.4%	4.0%	1.5%	1.8%
$\sigma_T(r_t - r_t^f)$	17.1%	15.6%	15.7%	15.5%	15.4%	12.2%	12.4%
Sharpe ratio	0.27	0.24	0.27	0.22	0.26	0.12	0.14
$\sigma_T(pd_t)$	0.38	0.26	0.28	0.26	0.29	0.06	0.07
$Corr_T(pd_t^{Model}, pd_t^{Data})$	n/a	0.37	0.53	0.31	0.52	0.24	0.25
ex ante:							
$E_T[E_t(r_{t+1} - r_{t+1}^f)]$	n/a	1.5%	1.7%	1.4%	1.6%	1.5%	1.8%

models occur due to *ex post* positive surprises in updates of beliefs. This is one of the primary implications of learning for this sample. Interestingly, after the burn-in period, this effect is also strong in the look-ahead prior. With parameter and model uncertainty, agents beliefs quickly deviate from their full sample estimates, highlighting the difficulty of learning in real-time, similar to the problem faced by an econometrician. In particular, the sequence of shocks realized over the post-war sample generate a times series of beliefs that have a systematic time series pattern: the initial low mean and high volatility of consumption growth causes an upward revision in the mean growth rates and a negative revision in the volatility parameters, as described in Section 3. Fama and French (2002) reach a similar conclusion in terms of the *ex post* versus the *ex ante* risk premium when looking at the time-series of the aggregate price-earnings and price-dividend ratios. Sargent and Cogley (2008) assume negatively biased beliefs in their model to highlight the same mechanism. The results we present here are consistent with their conclusions, but our models are estimated from fundamentals alone.

The equity return volatility is, in all the cases permitting parameter and model uncertainty, close to the 17.1% annual return volatility in the data (from 15.4% to 15.7%). In contrast, the equity return volatility in the models with fixed parameters is about 12%, which is almost all cash flow volatility as the annual dividend growth volatility is 11.5%. Thus, the sample variation in discount and growth rates arising from updates in agents' beliefs cause excess return volatility (Shiller, 1980). This is reflected in the sample volatility of the log price-dividend ratio, which is 0.38 in the data. In the cases with parameter and model uncertainty the volatility of the log price-dividend ratio lies between 0.26 and 0.29.[31] While this is only about three quarters of its volatility in the data, it is 4 to 5 times the volatility of the log price-dividend ratio in the benchmark fixed parameters models (here the volatility of the log price-dividend ratio is 0.06 for the 2-state model and 0.07 for the 3-state model).

The sample correlation between the log price-dividend ratios from the model versus the data, is 0.53 and 0.52 for the models using both GDP and consumption to estimate beliefs and 0.31 and 0.37 for the models using consumption only to estimate beliefs. The models with fixed parameters have lower correlations, 0.24 for the 2-state model and 0.25 for the 3-state model. As an alternative measure of the fit between the time-series of the sample price-level in the data versus those in the models considered here, the highest covariance between the price-dividend ratio in the data and the models with parameter and model uncertainty is 0.0573, whereas the highest covariance between the price-dividend ratio in the data and the models with fixed parameters is 0.0067 – a difference close to an order of magnitude. Thus, with parameter and model learning the model tracks the aggregate stock market price level (normalized by dividends) much more closely than either of the models we consider with fixed parameters. The price-level, a first order moment, is arguably even more important than matching the second order moments that usually are the focus in asset pricing.

As a formal test of the learning model's match of the aggregate stock price level (the log

[31] The price-dividend ratio in each model is calculated as the corresponding in the data by summing the last four quarters of payouts to get annual payout. The price-dividend ratio from the data includes share repurchases in its definition of total dividends.

D/P ratio) relative to the fixed parameter benchmark model, we run the following regression:

$$dp_t^{data} = \alpha + \beta_1 dp_t^{ParModUnc} + \beta_2 dp_t^{FP3} + \varepsilon_t, \tag{9}$$

where dp_t^{data} refers to the historical quarterly log dividend price ratio of the market portfolio, $dp_t^{ParModUnc}$ refers to the log dividend price ratio from the model with parameter and model uncertainty, and dp_t^{FP3} refers to the log dividend price ratio from the fixed parameters, 3-state model. The first four columns of Table 4 shows that the regression coefficient on the model with parameter and model uncertainty (β_1) is significant at the 1% level for both the historical and look-ahead priors, as well as whether learning is from realized consumption growth only or also including realized GDP growth. The R^2 ranges from 12% to 26% and is the lowest for the look-ahead prior with learning from consumption only, and the highest for the historical prior with learning from both consumption and GDP growth. As before, the results are shown after a 10-year burn-in period, from 1957 to 2009. The coefficient on the dividend yield from the fixed parameters model is insignificant in all of these cases. The fifth column of Table 4 shows the regression with only the dividend yield from the fixed parameters model. It is significant in this case, but the R^2 is only 6%. Finally, the last column of the table shows the regression with both the dividend yield from the fixed parameter model and the dividend yield from the historical prior with learning from both GDP and consumption growth, but where the dividend yield from the model with parameter and model learning has been orthogonalized with respect to the dividend yield from the fixed parameter model. The coefficient on the orthogonalized dividend yield (β_1) is still significant at the 1% level which implies that including the dividend yield from the model with parameter and model learning leads to a statistically significant (at the 1% level) increase in the R^2, relative to the fixed parameters benchmark case. The increase in fit from the full learning models stems from a better match of the business cycle fluctuations in the dividend yield, as well as low-frequency fluctuations. In particular, with parameter learning the dividend yield displays a downward trend over the sample, similar to that found in the data as documented by, for instance, Fama and French (2002).

In sum, including parameter and model uncertainty leads to not only better fit of the unconditional asset pricing moments, but a significantly better fit of the realized aggregate stock price level in the post-WW2 era.

Table 4 - Dividend Yield Regression

Table 4: The table reports the results of regressions where the log aggregate stock market dividend price ratio is the independent variable and contemporaneous log dividend price ratios from the model with parameter and model uncertainty ($dp^{ParModUnc}$), and the benchmark 3-state model with fixed parameters (dp^{FP3}). The standard errors are corrected for heteroskedasticity and given in parantheses under the coefficient estimates. Each column corresponds to a different prior and learning information set (Consumption only, or both consumption and GDP), The final column shows a regression where the log dividend price ratio from the model with parameter and model uncertainty has been orthogonalized with respect to the log dividend price ratio from the model with fixed parameters. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. The full sample period is from 1947:Q2 until 2009:Q1. However, we have removed the first 40 observations (10 years), as a burn-in period to alleviate misspecification of the priors. Similar results are obtained with no burn-in or with 80 quarter burn-in.

	Historical Prior		Look-ahead Prior		Fixed parameters	Historical Prior
	Cons.	Cons.	Cons.	Cons.	3-state model	Cons. + GDP
Variables	only	+ GDP	Only	+ GDP	only	(orthogonal)
constant	0.82	0.19	1.25	0.16	1.86	1.86
	(1.82)	(1.72)	(1.91)	(1.71)	(2.06)	(1.86)
$pd^{ParModUnc}$	0.47^{***}	0.79^{***}	0.37^{***}	0.61^{***}		0.79^{***}
	(0.13)	(0.14)	(0.12)	(0.11)		(0.14)
pd^{FP3}	0.73	0.28	0.92^{*}	0.40	1.45^{***}	1.45^{***}
	(0.46)	(0.43)	(0.49)	(0.43)	(0.56)	(0.51)
R^2	15.0%	25.8%	11.7%	20.0%	6.2%	25.8%

Permanent shocks and the volatility of long-run yields. With parameter and model uncertainty, the updates in mean beliefs constitute permanent shocks to expectations about consumption growth rates, consumption growth volatility, and higher order moments. This is a distinguishing feature of models with learning about constant quantities relative to learning about or observing a stationary underlying process (such as our state of the Markov chain, long-run risk in Bansal and Yaron (2004), or the surplus consumption ratio in Campbell and Cochrane (1999)). The latter models have transitory variables only in marginal utility growth. Shocks to a transitory state variable eventually die out, and so (very) long-run expectations are constant. Shocks to, for instance, the mean belief about the unconditional growth rate of consumption are, on the other hand, permanent, leading to permanent shocks to marginal utility growth. This has implications for all asset prices, but can be most clearly seen when considering the volatility of long-run default-free real yields, which can be readily calculated from our model. Table 5 shows the volatility of annualized yields for default-free real, zero-coupon bonds at different maturities. The data column gives the volatility of yields on U.S. TIPS, calculated from monthly data for the longest available sample, 2003 to 2011, from the Federal Reserve Board, along with the standard error of the volatility estimates. In the remaining columns, the corresponding model-implied yield volatilities, calculated from each of the models considered in this paper over the post-WW2 sample, are given.

First, the yield volatilities for the models with parameter and model uncertainty are substantially higher than the yield volatilities from the models with fixed parameters. The 2-year yields are twice as volatile, while the 10-year yields are an order of magnitude more volatile. This is a direct consequence of the permanent shocks to expectations resulting from parameter learning, whereas the models with fixed parameters have constant long-run consumption growth mean and volatility. Notably, the long maturity yields in the data have about the same yield volatility as in the models with parameter uncertainty, and so this is another dimension along which learning about parameters and models can help explain historical asset pricing behavior.

Table 5 - Real risk-free yield volatilities

Table 5: The table reports the sample standard deviation of annualized real risk-free yields at different maturities as computed from each of the models considered in the paper over the post-WW2 sample (1957 – 2009). The data column reports the standard deviation of annualized yields from the available data on TIPS from the Federal Reserve, which is monthly from January 2003 to February 2011.

TIPS	*Data*	*Consumption*		*Consumption, GDP*		*Fixed Parameters*	
(2003 – 2011)	(*s.e.*)	*Historical*	*Lookahead*	*Historical*	*Lookahead*	2-*state*	3-*state*
5-yr yield	0.75% (0.18%)	0.35%	0.30%	0.44%	0.39%	0.17%	0.19%
10-yr yield	0.45% (0.11%)	0.31%	0.27%	0.42%	0.36%	0.09%	0.10%
20-yr yield	0.30% (0.06%)	0.30%	0.26%	0.42%	0.35%	0.05%	0.06%
30-yr yield	*n/a*	0.30%	0.25%	0.42%	0.35%	0.03%	0.03%

Return Predictability Lastly, we consider excess market return forecasting regression using the dividend yield as the predictive variable. These regressions have a long history in asset pricing and remain a feature of the data that asset pricing models typically aim to explain (e.g., Campbell and Cochrane (1999), Bansal and Yaron (2004)). However, the strength of the empirical evidence is under debate (see, e.g., Stambaugh (1999), Ang and Bekaert (2007), Boudoukh, Richardson and Whitelaw (2008), and Goyal and Welch (2008) for critical analyses). Here we run standard forecasting regressions overlapping at the quarterly frequency using the sample of market returns and dividend yields as implied by each of the models. Note that, as before, we are not looking at population moments or average small-sample moments, but the single sample generated by feeding the models the actual sample of realized consumption growth.

Table 6 shows the forecasting regressions over different return forecasting horizons from the data. We use both the market dividend yield and the approximation to the consumption-wealth ratio, *cay*, of Lettau and Ludvigson (2001) to show the amount of predictability implied by these regressions in the data. We then run the same regressions using model implied returns and dividend yields. The benchmark models with fixed parameters (bottom right in the table) show no evidence of return predictability at the 5% significance level and the R^2's are very small. These models do, in fact, feature time-variation in the equity risk premium, but the standard deviation of the risk premiums are only about 0.5% per year and so the signal-to-noise ratio in these regressions is too small to result in significant predictability in a sample of the length we consider here. The models with parameter uncertainty, however, display significant in-sample return predictability and the regression coefficients and the R^2's are large and increasing in the forecasting horizon similar to those in the data. The *ex ante* predictability in these models is in fact similar to that in the fixed parameters cases, but since the parameters are updated at each point in time, there is significant *ex post* predictability. For instance, an increase in the mean parameters of consumption growth leads to high returns and lower dividend yield. Thus, a high dividend yield *in sample* forecasts high excess returns *in sample.* This is the same effect of learning as that pointed out in Timmermann (1993) and Lewellen and Shanken (2002). The models here show that the significant regression coefficients in the classical forecasting regressions show up in the sample only in the model where there is parameter learning which generates a significant difference between ex ante expected returns and ex post realizations. Thus, the model predicts that the amount of predictability is much smaller out-of-sample, consistent with the empirical evidence in Goyal and Welch (2008) and Ang and Bekaert (2007).

Table 6 - Return Forecasting Regressions

Table 6: This table presents quarterly excess market return forecasting regressions over various forecasting horizons (q quarters; 1 to 16). The top right shows the results when using market data and a measure of the log aggregate dividend yield; the *cay*-variable of Lettau and Ludvigsson (2001) and the CRSP aggregate log dividend yield ($\ln \frac{D_t}{P_t}$ where dividends are measured as the sum of the last four quarters' dividends. The rest of the table shows the results using the returns and dividend yield generated within the models based on the Historical priors, the Look-ahead priors, and the fixed parameter case. "Cons. only" denotes the model results in the case where only consumption growth is used to update beliefs, while "Cons. and GDP" denotes the model results in the case where both consumption and GDP growth are used to update beliefs. Newey-West autocorrelation and heteroskedasticity adjusted standard errors are given in parentheses (the number of lags is equal to the number of overlapping observations). * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. The full sample period is from 1947:Q2 until 2009:Q1. However, we have removed the first 40 observations (10 years), as a burn-in period to alleviate misspecification of the priors.

$$r_{t,t+q} - r_{f,t,t+q} = \alpha_q + \beta_{q,dp} \ln (D_t/P_t) + \varepsilon_{t,t+q}$$

	Data				Historical prior			
	$\ln (D_t/P_t) := cay_t$		$\ln (D_t/P_t) := \ln \frac{\Sigma_{j=0}^{3} D_{t-j}^{Mkt.}}{P_t^{Mkt.}}$		Cons. only		Cons. and GDP	
q	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}
1	1.19***	4.67%	0.03*	1.6%	0.04	1.4%	0.03	1.3%
	(0.31)		(0.02)		(0.03)		(0.02)	
4	4.29***	15.65%	0.11**	6.6%	0.18**	8.3%	0.14**	6.8%
	(1.18)		(0.05)		(0.07)		(0.06)	
8	7.60***	28.1%	0.17*	8.5%	0.38***	19.2%	0.28***	13.7%
	(1.72)		(0.10)		(0.09)		(0.08)	
16	12.31***	41.6%	0.22**	9.5%	0.61***	28.4%	0.44***	17.9%
	(1.82)		(0.11)		(0.15)		(0.13)	

	Look-ahead prior				Fixed parameters			
	Cons. only		Cons. and GDP		2-state model		3-state model	
q	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}	β_{dp} $(s.e.)$	R^2_{adj}
1	0.03	1.3%	0.03	0.9%	-0.01	0.0%	0.004	0.0%
	(0.02)		(0.03)		(0.062)		(0.062)	
4	0.18**	7.7%	0.15**	5.4%	0.19	1.0%	0.20	1.2%
	(0.07)		(0.07)		(0.17)		(0.16)	
8	0.38***	18.3%	0.29**	10.8%	0.37*	2.3%	0.41*	2.7%
	(0.12)		(0.09)		(0.24)		(0.23)	
16	0.64***	28.9%	0.42**	13.0%	0.26	0.7%	0.28	0.8%
	(0.17)		(0.17)		(0.31)		(0.30)	

5 Conclusion

This paper studies the statistical problem and asset pricing implications of learning about parameters, states, and models in a standard class of models for consumption dynamics. Our approach is empirical, focuses on the specific implications generated by learning about U.S. consumption dynamics during the post World War II period, and contributes to a growing empirical literature documenting the importance of learning for asset prices (e.g., Malmendier and Nagel (2011), and Pastor and Veronesi (2003)).

We find broad support for the importance of learning about parameters and models. Agents' beliefs about consumption growth dynamics are strongly time-varying, nonstationary, and help explain the realized time-series of equity returns and price-dividend ratio. In particular, the new and significant relationship we document between contemporaneous realized returns and revisions in beliefs is strong support for the importance of learning. Incorporating learning and our estimated time-series of beliefs in a general equilibrium model uniformly improves the model fit with respect to the standard asset pricing moments.

Taken together, this evidence questions the typical implementations of rational expectations consumption-based exchange economy models, in which agents know with certainty the data generating process for aggregate consumption growth. Further, the nonstationary dynamics induced by learning about fixed quantities such as parameters and models translates to nonstationary dynamics in marginal utility growth and asset valuation ratios. This, in turn, implies that standard econometric approaches to model tests and parameter estimation should be used with caution (see also Cogley and Sargent (2008)).

The procedure implemented in this paper can in a straightforward way be implemented for other countries or markets, or extended to multi-country or multi-asset settings. For instance, learning about the joint dynamics of dividends and consumption is an interesting exercise abstracted away from in this paper. In terms of other countries, it is clear that the post World War II experience of Japan would lead to a very different path of beliefs. Learning about the joint dynamics of, say, the U.S. and Japan's economies would have interesting implications, not only for their respective equity markets, but also for the real exchange rate dynamics. It will in future research be interesting to consider priced parameter uncertainty with Epstein-Zin preferences. Parameter and model uncertainty will be major sources of anxiety for agents with preferences for early resolution of uncertainty as these risks are nonstationary and thus truly "long-run." As in Bansal and Yaron (2004), these sources of uncertainty will likely command high risk prices.

References

[1] Ai, H. (2010), "Information about Long-Run Risk: Asset Pricing Implications," forthcoming, *Journal of Finance.*

[2] Ang, A. and G. Bekaert (2007), "Return predictability: Is it there?", *Review of Financial Studies*, 20, 3, 651-707.

[3] Backus, D., M. Chernov and I. Martin (2009), "Disasters Implied by Equity Index Options", *NYU Working Paper*

[4] Bakshi, G. and G. Skoulakis (2010), "Do Subjective Expectations Explain Asset Pricing Puzzles?", *Journal of Financial Economics,* December 2010, 117 - 140.

[5] Bansal, R. and A. Yaron (2004), "Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles", *Journal of Finance* 59(4), 1481 - 1509

[6] Bansal, R. and I. Shaliestovich (2010), "Confidence Risk and Asset Prices," *American Economic Review P&P*, 100, 537 – 541.

[7] Barberis, N. (2000), "Investing for the Long Run When Returns Are Predictable", *Journal of Finance* 55(1), 225 - 264

[8] Barillas, F., Hansen, L. and T. Sargent, "Doubts or Variability?", *Journal of Economic Theory*, forthcoming.

[9] Barro, R. (2006), "Rare Disasters and Asset Markets in the Twentieth Century", *Quarterly Journal of Economics* 121, 823 - 866

[10] Barro, R., Nakamura, E., Steinsson, J., and J. Ursua (2009), "Crises and Recoveries in an Empirical Model of Consumption Disasters," Columbia Business School working paper

[11] Barro, R. and J. Ursua (2008), "Consumption Disasters since 1870", *Brookings Papers on Economic Activity*

[12] Bhamra, H., L. Kuehn and I. Strebulaev (2009), "The Levered Equity Risk Premium and Credit Spreads: A Unified Framework", *Review of Financial Studies* (forthcoming)

[13] Boguth, O. and L. Kuehn (2009), "Consumption Volatility Risk," Working Paper Carnegie Mellon University

[14] Boudoukh, J., M. Richardson, and R. Whitelaw (2008), "The Myth of Long Horizon Predictability," *Review of Financial Studies* 21, 1577 - 1605.

[15] Brandt, M. , 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.

[16] Brennan, M. and Y. Xia (2002), "Dynamic Asset Allocation Under Inflation," *Journal of Finance* 57, 1201 – 1238.

[17] Bray, M. M. and N. E. Savin (1986), "Rational Expectations Equilibria, Learning, and Model Specification," *Econometrica* 54, 1129 - 1160.

[18] Cagetti, M., L. Hansen, T. Sargent and N. Williams (2002), "Robustness and Pricing with Uncertain Growth", *Review of Financial Studies* 15, 363 - 404

[19] Carvalho, Johannes, Lopes, and Polson (2010a), "Particle Learning and Smoothing," *Statistical Science*, 25, 88-106.

[20] Carvalho, Johannes, Lopes, and Polson (2010b – forthcoming), "Particle learning: Simulation-based Bayesian inference," *Bayesian Statistics* 9.

[21] Cecchetti, S., P. Lam and N. Mark (1990), "Mean Reversion in Equilibrium Asset Prices", *American Economic Review* 80, 398 - 418

[22] Cecchetti, S., P. Lam and N. Mark (1993), "The Equity Premium and the Risk Free Rate: Matching the Moments", *Journal of Monetary Economics* 31, 21 - 46

[23] Chen, H. (2009), "Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure", *Journal of Finance* (forthcoming)

[24] Chen, H., S. Joslin, and N. Tran (2010), "Rare Disasters and Risk Sharing with Heterogeneous Beliefs," MIT Working paper

[25] Cogley, T. and T. Sargent (2008), "The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression?", *Journal of Monetary Economics* 55, 454 - 476

[26] Cogley, T. and T. Sargent (2009), "Anticipated Utility and Rational Expectations as Approximations of Bayesian Decision Making", *International Economic Review* 49, 185 - 221

[27] David, A. and P. Veronesi (2009), "What ties return volatilities to price valuations and fundamentals?" University of Calgary Working Paper.

[28] Detemple, J. (1986), "Asset pricing in a production economy with incomplete information." *Journal of Finance*, 41, 383–390.

[29] Dothan, M. U. and D. Feldman (1986), "Equilibrium interest rates and multiperiod bonds in a partially observable economy." *Journal of Finance*, 41, 369 – 382.

[30] Drechsler, I. and A. Yaron (2008), "What's Vol Got to Do with It?", Working Paper, Wharton School of Business, University of Pennsylvania

[31] Epstein, L. and S. Zin (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework", *Econometrica* 57, 937 - 969

[32] Fama, E. and K. French (1988), "Dividend yields and expected stock returns," *Journal of Financial Economics* 22, 3 - 25.

[33] Fama, E. and K. French (2002), "The Equity Premium," *Journal of Finance* 57, 637 – 659.

[34] Froot, K. and S. Posner (2002), "The Pricing of Event Risks with Parameter Uncertainty", *GENEVA Papers on Risk and Insurance - Theory* 27, 153 - 165

[35] Hall, R. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, December 1978, 86(6), pp. 971-987.

[36] Gabaix, X. (2009), "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance", *NYU Working Paper*

[37] Gennotte, G. (1986), "Optimal Portfolio Choice under Incomplete Information," *Journal of Finance*, 61, 733-749.

[38] Geweke, J. (2001), "A note on some limitations of CRRA utility," *Economics Letters* 71, 341 - 345.

[39] Goyal A. and I. Welch, “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction,” July 2008, *Review of Financial Studies* 21(4) 1455-1508.

[40] Hansen, L. (2007), “Beliefs, Doubts and Learning: Valuing Macroeconomic Risk,” Richard T. Ely Lecture, *The American Economic Review* 97, No. 2., 1 - 30

[41] Hansen, L. and T. Sargent (2009), "Robustness, Estimation and Detection," Working paper.

[42] Hansen, L. and T. Sargent (2010), "Fragile Beliefs and the Price of Uncertainty," *Quantitative Economics*, Vol. 1, Issue 1, pp. 129-162

[43] Johannes, M. and N. Polson (2006), "Particle Filtering," *Springer Verlag Handbook of Financial Time Series*, edited by Torben G. Andersen Richard A. Davis, Jens-Peter Kreiss, and Thomas Mikosch, September 2006.

[44] Kandel, S. and R. Stambaugh (1990), "Expectations and Volatility of Consumption and Asset Returns", *Review of Financial Studies* 2, 207 - 232

[45] Kandel, S. and R. Stambaugh (1996), "On the Predictability of Stock Returns: An Asset-Allocation Perspective", *Journal of Finance* 51, 66 - 74

[46] Kreps, D. (1998), "Anticipated Utility and Dynamic Choice", *Frontiers of Research in Economic Theory*, (Cambridge: Cambridge University Press) 242 - 274

[47] Lettau, M. and S. Ludvigson (2001), "Consumption, Aggregate Wealth and Stock Returns," *Journal of Finance* 56, 815 - 849

[48] Lettau, M., S. Ludvigson and J. Wachter (2008), "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?", *Review of Financial Studies* 21(4), 1653 - 1687

[49] Lewellen, J. and J. Shanken (2002), "Learning, Asset-Pricing Tests, and Market Efficiency," *Journal of Finance* (57(3), 1113 - 1145.

[50] Lucas, R. (1978), “Asset Prices in an Exchange Economy,” *Econometrica* 46, 1429-1446

[51] Lucas, R. and T. Sargent (1979), "After Keynesian Macroeconomics," The Federal Reserve Bank of Minneapolis, *Quarterly Review* 321.

[52] Malmendier, U. and S. Nagel (2011), "Depression babies: Do Macroeconomic Experience Affect Risk-Taking?", *Quarterly Journal of Economics* 126, 373 – 416.

[53] Mehra, R. and E. Prescott (1985), "The Equity Premium: A Puzzle", *Journal of Monetary Economics* 15, 145 - 161

[54] Moore, B. and H. Schaller (1996), "Learning, Regime Switches, and Equilibrium Asset Pricing Dynamics", *Journal of Economic Dynamics and Control* 20, 979 - 1006

[55] Pastor, L. (2000), "Portfolio Selection and Asset Pricing Models," *Journal of Finance* 55, 179 - 223

[56] Pastor, L. and P. Veronesi (2003), "Stock valuation and learning about profitability," *Journal of Finance*, 58, 1749 – 1789.

[57] Pastor, L. and P. Veronesi (2006), "Was there a NASDAQ bubble in the late 1990s?," *Journal of Financial Economics*, 81, 61 – 100.

[58] Pastor, L. and P. Veronesi (2009), "Learning in Financial Markets," *Annual Review of Financial Economics.*

[59] Piazzesi, M. and M. Schneider (2010), "Trend and Cycle in Bond Risk Premia," Working Paper Stanford University

[60] Rietz, T. (1988), "The equity risk premium: A solution?", *Journal of Monetary Economics*, Volume 22, Issue 1, July 1988, Pages 117-131

[61] Romer, C. (1989), "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869-1908", *Journal of Political Economy* 97, 1 - 37

[62] Shaliastovich, I. (2008), "Learning, Confidence and Option Prices," Working Paper, Duke University

[63] Stambaugh, R. (1999), "Predictive Regressions." *Journal of Financial Economics,* 1999, 54, pp. 375–421.

[64] Timmermann, A. (1993), "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices." *Quarterly Journal of Economics*, 1993, 108, 1135-1145.

[65] Veronesi, P. (1999), "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model," *Review of Financial Studies*, 12, 5, Winter 1999

[66] Veronesi, P. (2000), "How does information quality affect stock returns?" *Journal of Finance*, 55, .

[67] Weitzman, M. (2007), "Subjective Expectations and Asset-Return Puzzles," *American Economic Review*, 97(4), 1102-1130.

[68] Whitelaw, R. (2000), "Stock Market Risk and Return: An Equilibrium Approach", *Review of Financial Studies* 13, 521 - 547.

[69] Working, H. (1960), "Note on the Correlation of First Differences of Averages in a Random Chain," *Econometrica* 28(4), 916 - 918.

[70] Xia, Y. (2001), "Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation", *Journal of Finance* 56(1), 205 - 246

6 Appendix

6.1 Existing literature and alternative approaches for parameter, state, and model uncertainty.

Our paper is related to a large literature studying the asset pricing implications of parameter or state learning. Most of this literature focuses on learning about a single unknown parameter or state variable (assuming the other parameters and/or states are known) that determines dividend dynamics and power utility. For example, Timmerman (1993) considers the effect of uncertainty on the average level of dividend growth, assuming other parameters are known, and shows in simple discounted cash-flow setting that parameter learning generates excess volatility and patterns consistent with the predictability evidence (see also Timmerman 1996). Lewellen and Shanken (2002) study the impact of learning about mean cash-flow parameters with exponential utility with a particular focus on return predictability.

Veronesi (2000) considers the case of learning about mean-dividend growth rates in a model with underlying dividend dynamics with power utility and focuses on the role of signal precision or information quality. Pastor and Veronesi (2003, 2006) study uncertainty about a fixed dividend-growth rate or profitability levels with an exogenously specified pricing kernel, in part motivated in order to derive cross-sectional implications. Weitzman (2007) and Bakshi and Skoulakis (2009) consider uncertainty over volatility.

Cogley and Sargent (2008) consider a 2-state Markov-switching model, parameter uncertainty over one of the transition probabilities, tilt beliefs to generate robustness via pessimistic beliefs, and use power utility. After calibrating the priors to the 1930s experience, they simulate data from a true model calibrated to the post War experience to show how priced parameter uncertainty and concerns for robustness impact asset prices, in terms of the finite sample distribution over various moments.

A number of papers consider state uncertainty, where the state evolves discretely via a Markov switching model or smoothing via a Gaussian process. Moore and Shaller (1996) consider consumption/dividend based Markov switching models with state learning and power utility. Brennen and Xia (2001) consider the problem of learning about dividend growth which is not a fixed parameter but a mean-reverting stochastic process, with power utility. Veronesi (2004) studies the implications of learning about a peso state in a Markov switching model with power utility. David and Veronesi (2010) consider a Markov switching model with learning about states.

In the case of Epstein-Zin utility, Brandt, Zeng, and Zhang (2004) consider alternative rules for learning about an unknown Markov state, assuming all parameters and the model is known. Lettau, Ludvigson, and Wachter (2008) consider information structures where the economic agents observe the parameters but learn about states in Markov switching consumption based asset pricing model. Chen and Pakos (2008) consider learning about the mean of consumption growth which is a Markov switching process. Ai (2010) studies learning in a production-based long-run risks model with Kalman learning about a persistent latent state variable. Bansal and Shaliastovich (2008) and Shaliastovich (2010) consider learning about the persistent component in a Bansal and Yaron (2004) style model with sub-optimal Kalman learning.

Additionally, some papers consider combinations of parameter or model uncertainty and robustness, see, e.g., Hansen and Sargent (2000,2009) and Hansen (2008).

6.2 Econometrics

This section briefly reviews the mechanics of sequential Bayesian learning and introduces the econometric methods needed to solve the high-dimensional learning problem. For ease of exposition, we abstract here from the problem of model uncertainty and drop the dependence on the model specification. Model uncertainty can be dealt with easily in a fashion analogous to the problem considered here.

The agent begins with initial beliefs over the parameters and states, $p(\theta, s_t) = p(s_t|\theta)\, p(\theta)$, and then updates via Bayes' rule. If at time t the agent holds beliefs $p(\theta, s_t|y^t)$, then updating occurs in a two step process by first computing the predictive distribution, $p(\theta, s_{t+1}|y^t)$, and then updating via the likelihood function, $p(y_{t+1}|s_{t+1}, \theta)$:

$$p\left(\theta, s_{t+1}|y^{t+1}\right) \propto p\left(y_{t+1}|\theta, s_{t+1}\right) p\left(\theta, s_{t+1}|y^t\right).$$

The predictive distribution is

$$p\left(\theta, s_{t+1}|y^t\right) = \int p\left(s_{t+1}|s_t, \theta\right) p\left(\theta, s_t|y^t\right) ds_t,$$

which shows the recursive nature of Bayesian updating, as $p(\theta, s_{t+1}|y^{t+1})$ is functionally dependent on $p(\theta, s_t|y^t)$.

The main difficulty is characterizing $p(\theta, s_t|y^t)$ for each t, which is needed for sequential learning. Unfortunately, even though s_t is discretely valued, there is no analytical form for $p(\theta, s_t|y^t)$, as it is high-dimensional and the dependence on the data is complicated and nonlinear. We use Monte Carlo methods called particle filters to generate approximate samples from $p(\theta, s_t|y^t)$. Johannes and Polson (2008) developed the general approach we use, and it was extended and applied to Markov switching models by Carvalho, Johannes, Lopes, and Polson (2010a, 2010b) and Carvalho, Lopes and Polson (2009). Details of the algorithms are given in those papers.

The first step of the approach, data augmentation, introduces a conditional sufficient statistics, T_t, for the parameters. Sufficient statistics imply that the full posterior distribution of the parameters conditional on entire history of latent states and data takes a known functional form conditional on a vector of sufficient statistics: $p(\theta|s^t, y^t) = p(\theta|T_t)$, where $p(\theta|T_t)$ is a known distribution. The conditional sufficient statistics are given by $T_{t+1} = \mathcal{T}(T_t, s_{t+1}, y_{t+1})$, where the function $\mathcal{T}(\cdot)$ is analytically known, which implies the sufficient statistics can be recursively updated. For Markov switching models, the sufficient statistics contain random variables such as the number of times and duration of each state visit, the mean and variance of y_t in those visits, etc. This step requires conjugate priors.

The key is that it is easier to sample from $p(\theta, T_t, s_t|y^t)$ than $p(\theta, s_t|y^t)$, where

$$p\left(\theta, T_t, s_t|y^t\right) = p\left(\theta|T_t\right) p\left(T_t, s_t|y^t\right). \tag{10}$$

By the definition of sufficient statistics and the use of conjugate priors, $p(\theta|T_t)$ is a known distribution (e.g., normal). This transforms the problem of sequential learning of parameters and states into one of sequential learning of states and sufficient statistics, and then standard updating by drawing from $p(\theta|T_t)$. The dimensionality of the target distribution, $p(\theta, T_t, s_t|y^t)$, is fixed as the sample size increases.

An $N-$ particle approximation, $p^N(\theta, T_t, s_t|y^t)$, approximates $p(\theta, T_t, s_t|y^t)$ via 'particles' $\left\{(\theta, T_t, s_t)^{(i)}\right\}_{i=1}^N$ so that:

$$p^N(\theta, T_t, s_t|y^t) = \frac{1}{N}\sum_{i=1}^{N}\delta_{(\theta,T_t,s_t)^{(i)}},$$

where δ is a Dirac mass. A particle filtering algorithm merely consists of a recursive algorithm for generating new particles, $(\theta, T_{t+1}, s_{t+1})^{(i)}$, given existing particles and a new observation, y_{t+1}. The approach developed in Johannes and Polson (2008) and Carvalho, Johannes, Lopes, and Polson (2009a, 2009b) generates a direct or exact sample from $p^N(\theta, T_t, s_t|y^t)$, without resorting to importance sampling or other approximate methods. The algorithm is straightforward to code and runs extremely quickly so that it is possible to run for large values N, which is required to keep the Monte Carlo error low. These draws can be used to estimate parameters and states variables.

In addition to sequential parameter estimation, particle filters can also be used for Bayesian model comparison. Bayesian model comparison and hypothesis testing utilizes the Bayes factor, essentially a likelihood ratio between competing specifications. Formally, given a number of competing model specifications, generically labeled as model $\mathcal{M}_k$ and $\mathcal{M}_j$, the Bayesian approach computes the probability of model k as:

$$p\left(\mathcal{M}_k|y^t\right) = \frac{p(y^t|\mathcal{M}_k)p\left(\mathcal{M}_k\right)}{\sum_{j=1}^{N}p(y^t|\mathcal{M}_j)p\left(\mathcal{M}_j\right)},$$

where $p(\mathcal{M}_k)$ is the prior probability of model k,

$$p(y^{t+1}|\mathcal{M}_k) = p(y_{t+1}|y^t, \mathcal{M}_k)p\left(y^{t-1}|\mathcal{M}_k\right),$$

and

$$p(y_{t+1}|y^t, \mathcal{M}_i) = \int p\left(y_{t+1}|\theta, s_t, \mathcal{M}_i\right) p\left(\theta, s_t|y^t, \mathcal{M}_i\right) d\left(\theta, s_t\right)$$

is the marginal likelihood of observation y_{t+1}, given data up to time t in model k. Marginal likelihoods are not known analytically and are difficult to compute even using MCMC methods. Since our algorithm provides approximate samples from $p(s_t, \theta|y^t)$, it is straightforward to estimate marginal likelihoods via

$$p^N(y_{t+1}|y^t, \mathcal{M}_k) = \frac{1}{N}\sum_{i=1}^{N} p\left(y_{t+1}|\left(\theta, s_t\right)^{(i)}, \mathcal{M}_k\right).$$

For all of our empirical results, we ran particle filtering algorithms with $N = 10K$ particles. We performed extensive simulations to insure that this number of particles insured a low and negligible Monte Carlo error.

6.3 Priors

Table 7 shows the prior parameters for the three different models we consider. The historical and look-ahead priors are different along some important dimensions. In particular, pre-WW2 consumption data is a lot more volatile than the post-war data (annual standard deviation of 4.8% in the pre-WW2 data versus 1.36% in post-WW2 data). This has been, in part, attributed to inferior pre-war data that is more noisy and sample that contains a more cyclical component of the economy (Romer, 1989). What is true, nevertheless, is that recessions were more frequent and lasted longer in the pre-WW2 data, and that the Great Depression was a worse recession than ever experienced afterwards, current crisis included. This is reflected in the disaster state in the 3-state models, in particular for the historical prior, akin to the disaster risk considered in Barro (2008).

For the historical prior, we have estimated, respectively, the 2- and 3-state models starting with very flat priors on the annual Shiller data. The posterior obtained at the end of the pre-war sample is transformed into a prior for the quarterly post-WW2 sample by dividing the average expected means and standard deviations within each regime by 4, and the average transition probability matrix, Π, is taken to the power of 1/4. This is of necessity somewhat ad hoc - first, a 2-state model on annual data does not imply a 2-state model on quarterly data; second, one would usually divide standard deviations by 2 to go from annual to quarterly. However, a large fraction of the pre-WW2 excess volatility is likely due to noisy data, which is not what we intend to capture with our prior. What is more, applying priors where the mean belief of the standard deviation of consumption growth within each regime is counter-factually high, leads to a state identification issue: the difference in the average beliefs of the mean within each state is too small relative to the volatilities and so the procedure cannot identify the separate states.

The look-ahead priors have mean values equal to the posterior from the corresponding historical priors in 2009:$Q1$. These are very close to what would be the maximum likelihood estimates obtained from estimating the 2- and 3-state models using the post-WW2 quarterly sample. The look-ahead priors have lower consumption growth volatility and higher persistence of the good state relative to the historical priors. Thus, the look-ahead prior reflects an expectation in 1947:$Q1$ of the world having higher growth and lower volatility than in the period before WW2. In terms of the tightness of the priors, the expansion state (always state 1), which has occurred the most, has the tightest priors, the recession state (state 2) has flatter priors as this state is visited less often, while the disaster state (state 3), for the 3-state models, has the flattest priors. This state is the one agents has the least information about, as it is a rare event.

For the extended model with both consumption and GDP growth, the priors are set to match the consumption-only model as much as possible to minimize the priors' effect on the comparison of the models. Since the means of the hidden state variable are equal to the means of the consumption growth in each state, the priors of these means are the same as in the consumption-only model. We also match the prior means of the total variance of consumption growth with similar flatness. However, since the specification allows for idiosyncratic noise in consumption growth ($\sigma_c \varepsilon_t^c$), we set both the mean of the variance of the hidden state variable in each state and the mean of the variance of the noise component to half of the prior mean of the total variance of consumption growth, with similar flatness. This way, the total prior mean variance of consumption growth, is the same as in the consumption only case. The priors for the transition probabilities are the same as in the consumption only case. For α and β in the GDP growth equation, the prior mean is -0.2 for α and 1.2 for β, and prior standard deviation is 0.45 for both. Finally, the prior mean of the idiosyncratic component of the variance of GDP growth is set by matching the variance of the GDP growth

APPENDIX: Table 7 - Priors

Table 7: The table shows the historical and look-ahead priors for the different models considered in the paper. The parameters within a state (mean and variance) have Normal/Inverse Gamma distributed priors, while the transition probabilities have Beta distributed priors. Note that $\hat{\pi}_{ij} \equiv \frac{\pi_{ij}}{1-\pi_{ij}}$.

Historical priors

Priors for i.i.d. model			Priors for 2-state model			Priors for 3-state model		
Par.	Mean	St.Dev	Par.	Mean	St.Dev	Par.	Mean	St.Dev
μ	0.9%	0.5%	μ_1	1.0%	0.25%	μ_1	1.0%	0.25%
μ_J	−2.0%	0.5%	μ_2	−0.5%	0.5%	μ_2	−0.4%	0.5%
						μ_3	−2.0%	1.5%
σ^2	$(0.7\%)^2$	$(0.7\%)^2$	σ_1^2	$(0.5\%)^2$	$(0.5\%)^2$	σ_1^2	$(0.5\%)^2$	$(0.5\%)^2$
σ_J^2	$(1.0\%)^2$	$(1.0\%)^2$	σ_2^2	$(1.0\%)^2$	$(1.0\%)^2$	σ_2^2	$(1.0\%)^2$	$(1.0\%)^2$
						σ_3^2	$(1.5\%)^2$	$(1.5\%)^2$
λ	0.05	0.05	π_{11}	0.95	0.034	π_{11}	0.95	0.034
						$\hat{\pi}_{12}$	0.80	0.16
						$\hat{\pi}_{21}$	0.80	0.16
			π_{22}	0.80	0.16	π_{22}	0.75	0.19
						$\hat{\pi}_{31}$	0.33	0.24
						π_{33}	0.40	0.20

Look-ahead priors

Priors for i.i.d. model			Priors for 2-state model			Priors for 3-state model		
Par.	Mean	St.Dev	Par.	Mean	St.Dev	Par.	Mean	St.Dev
μ	0.63%	0.22%	μ_1	0.68%	0.18%	μ_1	0.68%	0.18%
μ_J	−1.2%	0.25%	μ_2	0.2%	0.5%	μ_2	0.3%	0.5%
						μ_3	−1.14%	0.5%
σ^2	$(0.45\%)^2$	$(0.45\%)^2$	σ_1^2	$(0.36\%)^2$	$(0.36\%)^2$	σ_1^2	$(0.35\%)^2$	$(0.35\%)^2$
σ_J^2	$(0.55\%)^2$	$(0.55\%)^2$	σ_2^2	$(0.7\%)^2$	$(0.7\%)^2$	σ_2^2	$(0.7\%)^2$	$(0.7\%)^2$
						σ_3^2	$(0.7\%)^2$	$(0.7\%)^2$
λ	0.05	0.05	π_{11}	0.95	0.034	π_{11}	0.95	0.034
						$\hat{\pi}_{12}$	0.83	0.14
						$\hat{\pi}_{21}$	0.67	0.24
			π_{22}	0.80	0.16	π_{22}	0.75	0.19
						$\hat{\pi}_{31}$	0.50	0.29
						π_{33}	0.33	0.24

in the post-war data.

6.4 Time-Averaging of Consumption Data and Model Probabilities

The aggregate consumption data is time-averaged, which has implications for the volatility and autocorrelation structure of measured consumption growth. In particular, Working (1960) shows that time-averaging of i.i.d. data leads to lower variance (the variance is decreased by a factor of 1.5) and an autocorrelation of 0.25. Time-averaging can therefore artificially lead us to conclude that consumption growth follows a non-i.i.d. process (e.g., as we would get in the 2-state model with persistent states). Further, Hall (1978) argues theoretically and empirically that consumption growth is close to i.i.d. To ensure the rejection of the i.i.d. model we document in the paper is not an artifact of the time-averaging, we here assume the null hypothesis that consumption growth is in fact i.i.d., and remove the autocorrelation induced by time-averaging by creating the following residuals:

$$\nu_{c,t} = \Delta c_t - 0.25 * \Delta c_{t-1}. \tag{11}$$

We then redo the filtration exercise (parameters and models) and assign a prior probability of the i.i.d. model of 0.95. Figure 13 shows that also in this case, even with the strong model prior imposed, the i.i.d. model is rejected by the Bayesian agent about half-way through the sample.

6.5 Model solution and pricing

Here we give the details for how the prices of the consumption and aggregate equity claim in Section 4 are computed. At each point in time t, we price the equity claim given a set of model parameters, which are set equal to the mean beliefs at the time. The i.i.d. 2-state model, and the general 2- and 3-state models have parameters:

$$\begin{aligned}
\theta^{(1)} &= \{\mu_1, \mu_2, \sigma_1, \sigma_2, \pi_{11}\} \\
\theta^{(2)} &= \{\mu_1, \mu_2, \sigma_1, \sigma_2, \pi_{11}, \pi_{22}\}, \\
\theta^{(3)} &= \{\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3, \pi_{11}, \pi_{12}, \pi_{22}, \pi_{23}, \pi_{13}, \pi_{33}\},
\end{aligned}$$

respectively. In addition, there is the probability that the i.i.d 2-state model is the correct model, the probability that the general 2-state model is the correct model versus the residual probability of the 3-state model being the correct model. We also set these probabilities as

Figure 13 - Model Probabilities and Time-Averaging of Consumption Data

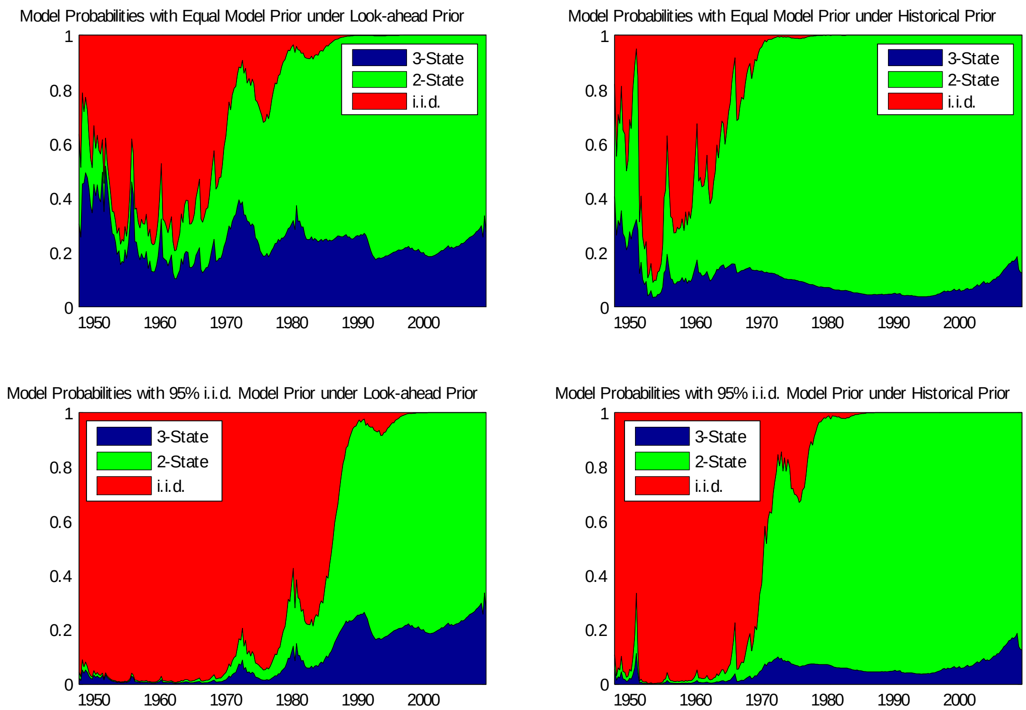


Figure 13: Model probabilities when assuming consumption growth is truly i.i.d. and removing the effect of time-averaging, as calculated by Working (1960).

constants when the agent prices the equity claim. Denote these probabilities p_1, p_2, and $p_3 = 1 - p_1 - p_2$. Thus, there is a total of 25 parameters that all are estimated using the particle filter and realized consumption (and GDP) data in real time. These mean parameter estimates will change at each time t, but we do not give the parameters time-subscripts to highlight that they are assumed to be constant following the anticipated utility framework in the pricing problem at each time t. In addition, there are the preference parameters γ, ψ, β, which are set to the values used in Bansal and Yaron (2004), and the leverage factor λ and the idiosyncratic dividend growth volatility σ_d. These parameters remain constant over the sample. When solving for the price-dividend ratio, we can and do ignore the idiosyncratic component of dividend growth.

First, we have to solve for the wealth-consumption ratio, PC. At each time t, the wealth-consumption ratio is solved using the recursion:

$$PC\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right) = \beta E\left[e^{(1-\gamma)\Delta c_{t+1}}\left(1 + PC\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)\right)^{\theta} |I_t\right]^{1/\theta}, \tag{12}$$

where the wealth-consumption ratio at time t is a function of the state-variables $s_t^{(2)}$ and $\tilde{s}_t^{(3)}$, and where I_t is the agent's information set which includes the mean parameter values used as constant parameters, as well as the mean state beliefs. The state-variable $s_t^{(2)}$ is the belief that the economy is in state 1 in the 2-state model. Remember that the states are still hidden, even though all the parameters are set to constants, so this belief will have a support of $(0, 1)$. Similarly, $\tilde{s}_t^{(3)}$ is the 2×1 vector of state belief probabilities from the 3-state model – the probability of being in state 1 and the probability of being in state 2.

In the model solution, the agent updates beliefs about $s^{(2)}$ and $\tilde{s}^{(3)}$ only by observing realized consumption growth – he does not know which model is the true model, or which state is the current state, so this uncertainty must be integrated out in the model solution. Below is a conceptual algorithm for the model solution.[32]

1. Given a set of parameters, start with an initial guess of the function $PC\left(s^{(2)}, \tilde{s}^{(3)}\right)$ on a grid for the 3 state variables, which all have support $(0, 1)$.

2. For each value of $s^{(2)}$, $\tilde{s}^{(3)}$ on the grid, do points 3. – 8. below:

[32]In actually solving the model, we employ numerical integration and not Monte Carlo simulation to find the wealth-consumption ratio. We compute the price-dividend ratio by summing over zero-coupon dividend claims. While we implement the model solution in this way for faster and more accurate model solution, this additional level of detail is not necessary for conceptually understanding how prices are computed.

3. Draw a model (the i.i.d. 2-state mode, or the general 2-state or 3-state model) according to the model probabilities p_1, p_2, and p_3.

4. Draw the current state of this model (state 1, state 2 (or state 3)), using the state belief for the current values in the grid for $s_t^{(2)}$ or $\tilde{s}_t^{(3)}$. Note: this step is irrelevant for the i.i.d. 2-state model.

5. Given the model and the state, draw a random standard normal shock ε_{t+1}, and compute consumption growth as

$$\Delta c_{t+1} = \mu_{M,j} + \sigma_{M,j}\varepsilon_{t+1}, \tag{13}$$

   where the subscript M refers to the model and the subscript j refers to the state in the same model. The parameters are assumed known and constant as discussed above.

6. Given observed log consumption growth (Δc_{t+1}) (the agent does not observe the shock ε), update the agent's belief using Bayes' rule. When finding $s_{t+1}^{(2)}$, condition on the 2-state model being the correct model, and when finding $\tilde{s}_{t+1}^{(3)}$, condition on the 3-state model being the correct model. See, e.g., Hamilton (1994) for how to update beliefs in switching regime models such as the ones considered here. Note that one has to update the belief for both models ($s^{(2)}$ and $\tilde{s}^{(3)}$), even though in the simulation of consumption growth we conditioned on one of the models, as the agent does not know the model.

7. Given $s_{t+1}^{(2)}$ and $\tilde{s}_{t+1}^{(3)}$ and the initial guess for PC, we have all we need to evaluate the expression inside the expectation of Equation (12).

8. Repeat 3. – 7. *many* times and take the average of the different values calculated for the expression inside the expectation of Equation (12). Use this average as an estimate of the expectation in Equation (12). Store the resulting value for $PC\left(s^{(2)}, \tilde{s}^{(3)}\right)$ found for the current place in the grid for $s^{(2)}$ and $\tilde{s}^{(3)}$.

9. Once 3. – 8. has been implemented for all values of $s^{(2)}$ and $\tilde{s}^{(3)}$ on the grid, update the function $PC\left(s^{(2)}, \tilde{s}^{(3)}\right)$.

10. Iterate on 2. – 9. until a suitable convergence criterion for the PC function has been achieved.

Points 1. – 10. gives the wealth consumption ratio at time t. The pricing functional $PC\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)$ must be computed in this way *for each* t, as the parameters will change at each time t. This is the anticipated utility component of the pricing. Denote the price-consumption ratio as a function of time t parameters as $PC_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)$.

The price-dividend ratio can be found similarly, by iterating on the below expression in the same manner as above for each time t in the sample with its corresponding time t set of parameter values:

$$PD_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right) = E\left[\beta^\theta e^{(\lambda-\gamma)\Delta c_{t+1}}\left(\frac{PC_t\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)+1}{PC_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)}\right)^{\theta-1}\left(1+PD_t\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)\right)|I_t\right]. \tag{14}$$

Finally, the returns to the equity claim are calculated as follows. For the return from time t to time $t+1$:

1. Set $s_t^{(2)}$ and $\tilde{s}_t^{(3)}$ equal to the mean state beliefs at time t (after parameter uncertainty is integrated out).

2. This gives the price dividend ratio at time t as $\frac{P_t}{D_t} = PD_t\left(s_t^{(2)}, \tilde{s}_t^{(3)}\right)$.

3. Set $s_{t+1}^{(2)}$ and $\tilde{s}_{t+1}^{(3)}$ equal to the mean state beliefs at time $t+1$ (after parameter uncertainty is integrated out).

4. This gives the price dividend ratio at time $t+1$ as $\frac{P_{t+1}}{D_{t+1}} = PD_{t+1}\left(s_{t+1}^{(2)}, \tilde{s}_{t+1}^{(3)}\right)$.

5. Next, using realized (in the data) consumption growth, obtain dividend growth as:

$$\frac{D_{t+1}}{D_t} = \left(\frac{C_{t+1}}{C_t}\right)^\lambda e^{-\frac{1}{2}\sigma_d+\sigma_d\varepsilon_{t+1}}, \tag{15}$$

   where ε_{t+1} is a draw from a standard normal distribution independent of everything else. These simulated shocks are constrained to have mean zero and variance one over the sample, such that $E_T\left[e^{-\frac{1}{2}\sigma_{d.t+1}^2+\sigma_d\varepsilon_{t+1}}\right] = 1$ (in practice, extremely close to 1). This is done to ensure that the level of the in-sample average equity return and equity return volatility are not affected by the (by chance) high or low draw of the idiosyncratic component of dividends, or (by chance) high or low volatility of idiosyncratic dividend growth.

6. Given this, the return is calculated as:

$$R_{t,t+1} = \frac{D_{t+1}}{D_t} \left(\frac{P_t}{D_t} \right)^{-1} \left(1 + \frac{P_{t+1}}{D_{t+1}} \right). \tag{16}$$