Measuring Ambiguity Aversion
A. Ronald Gallant
Penn State University*

Mohammad R. Jahan-Parvar
Federal Reserve Board†

Hening Liu‡§
University of Manchester

Original Draft: October 2014
This Draft: September 2015

Abstract

We confront the generalized dynamic intertemporal “smooth ambiguity aversion” preferences of Klibanoff, Marinacci, and Mukerji (2005, 2009) with data using Bayesian methods introduced by Gallant and McCulloch (2009) to close two existing gaps in the literature. First, we estimate the size of ambiguity aversion implied by financial data for the representative agent in a consumption-based equilibrium asset pricing model. Second, we investigate the contribution of ambiguity aversion in explaining variations in equity premium and consumption growth. Our estimates are comparable with those from existing empirical research, demonstrate sensitivity to sampling frequency, and suggest ample scope for ambiguity aversion.

JEL Classification: C61; D81; G11; G12.

Keywords: Ambiguity aversion, Bayesian estimation, Equity premium puzzle, Markov switching.

*Corresponding Author, Department of Economics, The Pennsylvania State University, 511 Kern Graduate Building, University Park, PA 16802 USA. e-mail: aronaldg@gmail.com.
†Office of Financial Stability Policy and Research, Board of Governors of the Federal Reserve System, 20th St. and Constitution Ave. NW, Washington, DC 20551 USA. e-mail: Mohammad.Jahan-Parvar@frb.gov.
‡Accounting and Finance Group, Manchester Business School, University of Manchester, Booth Street West, Manchester M15 6PB, UK. e-mail: Hening.Liu@mbs.ac.uk.
§We thank two anonymous referees, Toni Whited (the Editor), Geert Bekaert, Dan Cao, Yoosoon Chang, Marco Del Negro, Ana Fostel, Luca Guerrieri, Nour Meddahi, Joon Y. Park, Eric Renault, Jay Shambaugh, Chao Wei, seminar participants at Federal Reserve Board, George Washington University, Georgetown University, Indiana University, 2014 Midwest Econometric Group, CFE 2014 (Pisa), SoFiE 2015 annual conference (Aarhus University), the Econometric Society World Congress 2015 (Montreal) for helpful comments and discussions. The analysis and the conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. Any remaining errors are ours.
1 Introduction

In this paper, we confront the “smooth ambiguity aversion” model of Klibanoff, Marinacci, and Mukerji (2005, 2009), (henceforth, KMM), in its generalized form advanced by Hayashi and Miao (2011) and Ju and Miao (2012), with data to close two existing gaps in the literature. First, we empirically estimate the size of ambiguity aversion implied by financial data for a representative agent endowed with smooth ambiguity aversion preferences in a consumption-based equilibrium asset pricing model. Second, we empirically investigate the contribution of smooth ambiguity aversion in explaining variations in equity premium and consumption growth. Given the rising popularity of smooth ambiguity preferences in economics and finance, it is important to characterize this model’s empirical strengths and contributions, as well as its shortcomings. One salient feature of smooth ambiguity aversion is the separation of ambiguity and ambiguity aversion, where the former is a characteristic of the representative agent’s subjective beliefs, while the latter derives from the agent’s tastes. This study provides a fully data-based estimation of this ambiguity aversion parameter in a dynamic asset pricing model. Our estimated ambiguity aversion parameter is higher than its calibrated counterparts in existing endowment or production-based asset pricing studies. Other estimated structural parameters are comparable with estimated values reported in the literature.

Ambiguity aversion matters. Jeong, Kim, and Park (2015) show that in an equilibrium asset pricing model where agents are endowed with a different but related class of ambiguity aversion preferences, ambiguity aversion accounts for 45% of average equity premium. They also find that it is economically and statistically significant. Our findings confirm theirs, and extend the literature in new directions. In our benchmark model, ambiguity arises due to a mixture of distributions for dividend growth. The state determining the distribution of dividend growth is unobservable. The agent can learn about the hidden state, but this ability does not eliminate the difficulties in forming forecasts. In turn, these difficulties generate the scope for ambiguity aversion. Ambiguity aversion gives rise to intertemporal choices that differ dramatically from those made by an ambiguity-neutral agent.

KMM preferences and “multiple-priors utility” of Chen and Epstein (2002) (henceforth, MPU) have drawn considerable attention in the literature. In practice, smooth ambiguity aversion of KMM has two important advantages over MPU. Critically, MPU does not admit a sharp separa-
tion between ambiguity and ambiguity aversion. In the MPU framework, the set of priors, which characterizes ambiguity, also determines the degree of ambiguity aversion. Thus, in empirical studies based on MPU such as Jeong et al. (2015), one only obtains the estimate of the size of ambiguity instead of the magnitude of ambiguity aversion. In the MPU framework, it is therefore infeasible to do comparative statics analysis by holding the family of alternative distributions constant while varying the degree of ambiguity aversion. The most important feature of the generalized recursive smooth ambiguity preferences is precisely a separation between ambiguity and ambiguity aversion, and moreover a three-way separation among risk aversion, ambiguity aversion, and the elasticity of intertemporal substitution. The second advantage of KMM over MPU is tractability. Asset pricing models with MPU are generally difficult to solve with refined processes of fundamentals because MPU features kinked preferences.¹


Our estimation of the level of ambiguity aversion facilitates using this class of preferences by linking it directly to the data. Most of existing applications of smooth ambiguity preferences rely on the methodology of calibration. The popular methods of calibrating the degree of ambiguity aversion include the “detection-error probability” method of Anderson, Hansen, and Sargent (2003) and Hansen (2007) (see Jahan-Parvar and Liu (2014) for an application) and “thought experiments” similar to Halevy (2007) (see Ju and Miao (2012) and Chen et al. (2014) for applications). Clearly, the contribution of our study is methodological in that we use both financial and macroeconomic data to estimate the degree of ambiguity aversion together with other structural parameters in a dynamic asset pricing model with learning.

Similar to other macro-finance applications, we face sparsity of data. As has become standard in

¹ Strzalecki (2013) provides a rigorous and comprehensive discussion of ambiguity-based preferences.
the macro-finance empirical literature, we use prior information and a Bayesian estimation methodology to overcome data sparsity. Specifically, we use the “General Scientific Models” (henceforth, GSM) Bayesian estimation methodology developed by Gallant and McCulloch (2009). GSM is the Bayesian counterpart to the classical “indirect inference” and “efficient method of moments” (hereafter, EMM) methods introduced by Gouriéroux, Monfort, and Renault (1993) and Gallant and Tauchen (1996, 1998, 2010). These are simulation-based inference methods that rely on an auxiliary model for implementation. GSM follows the logic of the EMM variant of indirect inference and relies on the theoretical results of Gallant and Long (1997) in its construction of a likelihood. A comparison of Aldrich and Gallant (2011) with Bansal, Gallant, and Tauchen (2007) displays the advantages of a Bayesian EMM approach relative to a frequentist EMM approach, particularly for the purpose of model comparison. An indirect inference approach is an appropriate estimation methodology in the context of this study since the estimated equilibrium model is highly nonlinear and does not admit of analytically tractable solutions thereby severely inhibiting accurate, numerical construction of a likelihood by means other than GSM. GSM uses a sieve (Section 4) specially tailored to macroeconomic and finance time-series applications as the auxiliary model. When a suitable sieve is used as the auxiliary model, as here, the GSM method synthesizes the exact likelihood implied by the model. In this instance, the synthesized likelihood model departs significantly from a normal-errors likelihood, which suggests that alternative econometric methods based on normal approximations will give biased results. In particular, in addition to GARCH and leverage effects, the three-dimensional error distribution implied by the smooth ambiguity aversion model is skewed in all three components and has fat-tails for consumption growth and stock returns and thin tails for bond returns.

Ahn, Choi, Gale, and Kariv (2014) estimate the level of ambiguity aversion for several static specifications of ambiguity aversion preferences based on experimental data. Their findings differ from ours since a) they use a static specification and ignore the intertemporal choice, and b) their empirical findings imply that ambiguity aversion parameter estimates for the smooth ambiguity specification are not significantly different from zero. In addition, the mapping between their static estimation results and our dynamic model estimates is not clear. Similar to our findings and

2 Gallant and McCulloch (2009) use the terms “scientific model” and “statistical model” instead of the terms “structural model” and “auxiliary model” used in the indirect inference econometric literature. We will follow the conventions of the econometric literature. The structural models here are the “smooth ambiguity aversion” model and a restricted version of that model.

3 They get much tighter estimates and hence statistically significant results for their kinked ambiguity specification.
in line with the theory, they report the estimates of ambiguity aversion parameter that are larger than the estimates of risk aversion parameter. However, the magnitudes of their estimates are very different from ours and those reported in calibration studies. Such discrepancies between market data-based and experimental estimates are common.

Two recent papers, Jeong, Kim, and Park (2015) and Viale, Garcia-Feijoo, and Giannetti (2014), provide time series and cross-sectional estimates for the MPU model. As mentioned above, these studies do not provide a direct measurement of ambiguity aversion because MPU does not admit a functional separation between ambiguity and ambiguity aversion. In a production based general equilibrium setting, Ilut and Schneider (2014) assume that ambiguity is an exogenously determined autoregressive process, while Bianchi, Ilut, and Schneider (2014) define ambiguity as parameter uncertainty or measurement error in volatility of marginal product of capital and in operating costs. Anderson, Ghysels, and Juergens (2009) estimate the magnitude of ambiguity aversion using forecasts of professional forecasters.

In a recent study, Thimme and Völkert (2015) attempt to estimate the ambiguity aversion parameter in KMM preferences. Their methodology differs from ours along these dimensions: First, they do not specify how ambiguity arises in their underlying asset pricing model. In our specification, similar to Ju and Miao (2012) and Jahan-Parvar and Liu (2014), ambiguity arises due to a mixture of distributions for economic fundamentals. Second, they linearize the stochastic discount factor in order to simplify the computation of the expectations formed on the state of the world – the source of ambiguity in KMM preferences. We do not linearize the SDF. When needed, we compute the SDF numerically. Third, given the linearized structure of their model, they use a combination of calibrated intertemporal elasticity of substitution (IES) parameter and generalized method of moments (GMM) to estimate the remaining preference parameters. Given the sparsity of the available data, their estimation methodology does not allow for direct model comparison. We, on the other hand, estimate all structural parameters, including the IES parameter and parameters pertaining to state variables. In addition, our Bayesian estimation model admits model comparisons. Given the calibrated IES parameters they use, their study yields several estimates for ambiguity aversion parameter that are broadly comparable to ours.

The rest of the paper proceeds as follows. Section 2 introduces the data used in this study. Section 3 presents the consumption-based asset pricing model with generalized recursive smooth ambiguity preferences developed by Ju and Miao (2012), and the numerical method used to solve
the model. Section 4 discusses the estimation methodology and presents our empirical findings. Section 5 presents model comparison results, forecasts, and asset pricing implications. Section 6 concludes.

2 Data

Throughout this paper, lower case denotes the logarithm of an upper case quantity; e.g., \( c_t = \ln(C_t) \), where \( C_t \) is the observed consumption in period \( t \), and \( d_t = \ln(D_t) \), where \( D_t \) is dividends paid in period \( t \). Similarly, we use logarithmic risk-free interest rate \( (r^f_t) \) and aggregate equity market return inclusive of dividends \( (r^e_t = \ln(P^e_t + D_t) - \ln P^e_{t-1}) \) in the analysis, where \( P^e_t \) is the stock price in period \( t \).

We use real annual data from 1929 to 2013 and real quarterly data from the second quarter of 1947 to the second quarter of 2014 for the purpose of inference, indexed by 2005 price levels. We use the data from 1929 to 1949 (1947:Q2 to 1955:Q2) to provide initial lags for the recursive parts of the model and 1950-2013 (1955:Q3-2014:Q2) data for estimation of parameters and for diagnostics. Our measure for the risk-free rate is one-year U.S. Treasury Bill rate for annual data and 3-months U.S. Treasury Bill rate for quarterly data. Our proxy for risky asset returns is the value weighted returns on CRSP-Compustat stock universe. We use the sum of nondurable and services consumption from Bureau of Economic Analysis (BEA) and deflate the series using the appropriate price deflator (also provided by the BEA). We use mid-year population data to obtain per capita consumption values. As noted in Garner, Janini, Passero, Paszkiewicz, and Vendemia (2006), there are notable discrepancies between measures of consumption released by different agencies. Thus, throughout the paper, we assume a 5\% measurement error in the level of real per capita consumption.\(^4\) We assume a linear error structure. That is, \( C_t = C^*_t + u_t \) where \( C_t \) is the observed value, \( C^*_t \) is the true value, and \( u_t \) is the measurement error term. Since logarithmic values yield the consumption growth series, we have \( c_t = \ln(C_t) = \ln(C^*_t + u_t) \) and \( \Delta c_t = \ln(C_t/C_{t-1}) \).

Table 1 presents the summary statistics of the data used in this study. Reported mean and standard deviations of risk-free rates \( (r^f_t) \), aggregate market returns \( (r^e_t) \), excess returns \( (r^e_t - r^f_t) \), and real, per capita, log consumption growth \( (\Delta c_t) \) are in percentages. The reported \( p \)-values of Jarque and Bera (1980) test of normality imply that the assumption of normality is rejected for risk-free rate and log consumption growth series, but it cannot be rejected for aggregate market

\(^4\) We also experimented with 1\% and 10\% error levels. Empirical results are robust to the level of measurement errors.
returns and excess returns at annual frequency. Annual data plots are shown in Figure 1.

3 The Model

The intuitive notions behind any consumption based asset pricing model are that agents receive income (wage, interest, and dividends) which they use to purchase consumption goods. Agents reallocate their consumption over time by trading shares of stock that pay a random dividend and bonds that pay interest with certainty. This is done for consumption smoothing over time (for example, insurance against unemployment, saving for retirement, · · · ). Trading activity enters the model via the agent’s budget constraint that implies an agent’s purchase of consumption, bonds, and stock cannot exceed income (in the form of aggregated wage, interest, and dividends) in any period. When applied to a national closed economy, consumption and dividends can be used as the driving processes instead of wages and dividends. Agents are endowed with a utility function that depends on the entire consumption process. The first order conditions of their utility maximization problem determine a map from present and past values of the driving processes to the present price of a stock and a bond. These models are simulated by first simulating the driving processes and then evaluating the map that determines stock and bond prices.

3.1 The Benchmark Structural Model

We consider the representative-agent pure exchange economy model of Ju and Miao (2012). Aggregate consumption follows the process

$$\Delta c_{t+1} \equiv \ln \left( \frac{C_{t+1}}{C_t} \right) = \kappa z_{t+1} + \sigma \epsilon_{t+1},$$

(1)

where $\epsilon_t$ is an $i.i.d.$ standard normal random variable, and $z_{t+1}$ follows a two-state Markov chain with state 1 being the good state and state 2 being the bad state ($\kappa_1 > \kappa_2$). The transition matrix is

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

where $p_{ij}$ denotes the probability of switching from state $i$ to state $j$, and $p_{12} = 1 - p_{11}$ and $p_{21} = 1 - p_{22}$.

Dividend growth is modeled as containing a component proportional to consumption growth.
and an idiosyncratic component,

\[ \Delta d_{t+1} \equiv \ln \left( \frac{D_{t+1}}{D_t} \right) = \lambda \Delta c_{t+1} + g_d + \sigma_d \varepsilon_d,t+1, \]  

(2)

where \( \varepsilon_d,t+1 \) is an i.i.d. standard normal random variable and is independent of all other shocks in the model. The parameter \( \lambda \) can be interpreted as the leverage ratio on expected consumption growth as in Abel (1999) and Bansal and Yaron (2004). The parameters \( g_d \) and \( \sigma_d \) can be pinned down by calibrating the model to the first and second moments of dividend growth.

Ju and Miao (2012) assume that economic regimes are not observable, but the agent can learn about the state \( (z_t) \) through observing the history of consumption and dividends. The agent also knows the parameters of the model, namely, \( \{\kappa_1, \kappa_2, \lambda, g_d, \sigma_d\} \). The agent updates beliefs \( \mu_t = \Pr (z_{t+1} | \Omega_t) \) according to Bayes’ rule:

\[ \mu_{t+1} = \frac{p_{11} f (\Delta c_{t+1}, 1) \mu_t + p_{21} f (\Delta c_{t+1}, 2) (1 - \mu_t)}{f (\Delta c_{t+1}, 1) \mu_t + f (\Delta c_{t+1}, 2) (1 - \mu_t)}, \]  

(3)

where \( f (\Delta c_{t+1}, i), i = 1, 2 \) is the normal density function of consumption growth conditional on state \( i \).

The agent’s preferences are represented by the generalized recursive smooth ambiguity utility function,

\[ V_t(C) = \left[ (1 - \beta) C_t^{1 - 1/\psi} + \beta \{ R_t(V_{t+1}(C)) \}^{1 - 1/\psi} \right]^{1 - 1/\psi}, \]  

(4)

\[ R_t(V_{t+1}(C)) = \left( \mathbb{E} \mu_t \left[ \left( \mathbb{E}_{z_{t+1}, t} \left[ V_{t+1}^{1 - \gamma}(C) \right] \right)^{\frac{1 - \eta}{1 - \gamma}} \right] \right)^{\frac{1}{1 - \eta}}, \]  

(5)

where \( \beta \in (0, 1) \) is the subjective discount factor, \( \psi \) is the IES parameter, \( \gamma \) is the coefficient of relative risk aversion, and \( \eta \) is the ambiguity aversion parameter and must satisfy \( \eta > \gamma \) to maintain ambiguity aversion in the utility function. Equation (5) characterizes the certainty equivalent of future continuation value, which is the key ingredient that distinguishes this utility function from Epstein-Zin’s recursive utility. In Equation (5), the expectation operator \( \mathbb{E}_{z_{t+1}, t} [\cdot] \) is with respect to the distribution of consumption conditioning on the next period’s state \( z_{t+1} \), and the expectation operator \( \mathbb{E}_{\mu_t} \) is with respect to the filtered probabilities about the unobservable state.
Under this utility function, the stochastic discount factor (SDF) is given by

$$M_{zt+1,t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-1/\psi} \left( \frac{V_{t+1}}{R_t(V_{t+1})} \right)^{1/\psi-\gamma} \left( \frac{E_{zt+1,t} \left[ V_{t+1}^{1-\gamma} \right]}{R_t(V_{t+1})} \right)^{\frac{1}{1-\gamma}}. \quad (6)$$

Stock returns, $R_{t+1}^e$, are defined by

$$R_{t+1}^e = \frac{P_{t+1}^e + D_{t+1}}{P_t^e} = 1 + \varphi(\mu_{t+1}) \frac{D_{t+1}}{D_t},$$

where $\varphi(\mu_t)$ denotes the price-dividend ratio. Stock returns satisfy the Euler equation

$$E_{\mu,t} \left[ M_{zt+1,t+1} R_{t+1}^e \right] = 1. \quad (7)$$

The risk-free rate, $R_{f,t}$, is the reciprocal of the expectation of the SDF:

$$R_f^t = \frac{1}{E_{\mu,t} \left[ M_{zt+1,t+1} \right]}.$$ 

We can rewrite the Euler equation as

$$0 = \tilde{\mu}_t \left[ M_{zt+1,t+1}^E (R_{e,t+1} - R_{f,t}) \right] + (1 - \tilde{\mu}_t) \left[ M_{zt+1,t+1}^E (R_{e,t+1} - R_{f,t}) \right],$$

where $M_{zt+1,t+1}^E$ can be interpreted as the SDF under Epstein-Zin recursive utility:

$$M_{zt+1,t+1}^E = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} \left( \frac{V_{t+1}}{R_t(V_{t+1})} \right)^{\frac{1}{\psi}-\gamma},$$

and $\tilde{\mu}_t$ can be interpreted as ambiguity distorted beliefs and represented by:

$$\tilde{\mu}_t = \frac{\mu_t \left[ E_{1,t} \left[ V_{t+1}^{1-\gamma} \right] \right]^{-\frac{\eta-\gamma}{1-\gamma}}}{\mu_t \left[ E_{1,t} \left[ V_{t+1}^{1-\gamma} \right] \right]^{-\frac{\eta-\gamma}{1-\gamma}} + (1 - \mu_t) \left[ E_{2,t} \left[ V_{t+1}^{1-\gamma} \right] \right]^{-\frac{\eta-\gamma}{1-\gamma}}}. \quad (8)$$

As long as $\eta > \gamma$, distorted beliefs are not equivalent to Bayesian beliefs. This distortion is an equilibrium outcome and is entirely driven by ambiguity aversion. Panel A, Figure 6 shows the Bayesian belief and the ambiguity-distorted belief filtered using the quarterly historical consumption growth data for the period 1947–2014. The parameter values in the model are set to the estimates
using the GSM Bayesian estimation method, which are shown below. It is obvious from Panel A, Figure 6 that ambiguity aversion distorts the Bayesian belief in a pessimistic way, and thus an ambiguity averse agent slants his beliefs towards the bad regime.

We follow Ju and Miao (2012) and use the projection method with Chebyshev polynomials to solve the model. Specifically, homogeneity in utility preferences implies \( V_t(C) = G(\mu_t)C_t \), and \( G(\mu_t) \) satisfies the following functional equation

\[
G(\mu_t) = \left[ (1 - \beta) + \beta \left( \mathbb{E}_{\mu_t} \left[ \left( \mathbb{E}_{z_{t+1},t} \left[ G(\mu_t)^{1-\gamma} \left( \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right] \right] \right) \right)^{\frac{1}{1-\gamma}} \right]^{\frac{1}{1-1/\psi}}.
\]

To solve for the value function, we approximate \( G(\mu_t) \) using Chebyshev polynomials in the state variable \( \mu_t \). The approximation takes the form

\[
G(\mu) \simeq \sum_{k=0}^{p} \phi_j T_j(y(\mu)),
\]

where \( p \) is the order of Chebyshev polynomials, \( T_j \) with \( j = 1, ..., p \) are Chebyshev polynomials, and \( y(\mu) \) maps the state variable \( \mu \) onto the interval \([-1, 1]\). We then choose a set of collocation points for \( \mu \) and solve for the coefficients \( \{\phi_j\}_{j=0,\ldots,p} \) using a nonlinear equations solver. The expectation \( \mathbb{E}_{z_{t+1},t}[\cdot] \) is approximated using Gauss-Hermite quadrature.

To solve for the equilibrium price-dividend ratio, we rewrite the Euler equation as

\[
\frac{P^e_t}{D_t} = \mathbb{E}_t \left[ M_{z_{t+1},t+1} \left( 1 + \frac{P^e_{t+1}}{D_{t+1}} \right) \frac{D_{t+1}}{D_t} \right].
\]

The price-dividend ratio can also be approximated using Chebyshev polynomials in the state variable \( \mu_t \). Since the SDF \( M_{z_{t+1},t+1} \) can be easily written as a functional of \( G(\mu_{t+1}) \) and consumption growth \( \Delta c_t = \ln \left( \frac{C_{t+1}}{C_t} \right) \), we can solve for the equilibrium price-dividend ratio in a similar way as we solve for the value function.

After solutions are found, we simulate logarithmic values of consumption growth, stock returns and risk-free rates:

\[
\left\{ \ln \left( \frac{C_{t+1}}{C_t} \right), \ln \left( R^e_{t+1} \right), \ln \left( R^f_{t+1} \right) \right\}_{t=1}^T.
\]
3.2 The Alternative Model with Ambiguity Neutrality

Our discussion of equations (4) and (5) conveys one important message: Ambiguity aversion has an impact on the intertemporal decisions of the representative economic agent if and only if \( \eta > \gamma \). Once \( \eta = \gamma \), then the agent is ambiguity neutral. As a result, the agent’s preferences collapse to the familiar Kreps and Porteus (1978) and Epstein and Zin (1989) preferences:

\[
V_t(C) = \left[ (1 - \beta)C_t^{1-1/\psi} + \beta \{ R_t(V_{t+1}(C)) \}^{1-1/\psi} \right]^{1/(1-\psi)},
\]

\[
R_t(V_{t+1}(C)) = \mathbb{E}_t \left[ V_{t+1}^{1-\gamma}(C) \right]^{1/(1-\gamma)}.
\]

It is immediately obvious that under ambiguity neutrality, the certainty equivalent in equation (5) collapses to the familiar temporal expectation. Thus, the representative agent simply is faced with a Markov switching structure in the aggregate consumption and dividend growth processes, following equations (1) and (2), and makes decisions based on Epstein-Zin preferences.

Given these preferences, the SDF is now:

\[
M_{t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-1/\psi} \left( \frac{V_{t+1}}{\mathbb{E}_t(V_{t+1})} \right)^{1/\psi-\gamma}.
\]

We use the same solution method that we use for the benchmark model to solve the problem for the ambiguity-neutral agent. After solving the model, we simulate logarithmic values of consumption growth, stock returns and risk-free rates.

4 Estimation of Model Parameters

To estimate model parameters we use a Bayesian method proposed by Gallant and McCulloch (2009), abbreviated GM hereafter, that they termed General Statistical Models (GSM). The GSM methodology was refined in Aldrich and Gallant (2011), abbreviated AG hereafter. The discussion here incorporates those refinements and is to a considerable extent a paraphrase of AG. The symbols \( \zeta, \theta, \) etc. that appear in this section are general vectors of statistical parameters and are not instances of the model parameters of Section 3.1 and Section 3.2.

Let the transition density of a structural model be denoted by

\[
p(y_t|x_{t-1}, \theta), \quad \theta \in \Theta,
\]
where $x_{t-1} = (y_{t-1}, \ldots, y_{t-L})$ if Markovian and $x_{t-1} = (y_{t-1}, \ldots, y_{1})$ if not. As a result, $x_{t-1}$ serves as a shorthand for lag-lengths that are generally greater than 1. Thus, transition densities may depend on $L$-lags of the data (if Markovian) or the entire history of observations (if non-Markovian).

There are two structural models under consideration in this application: the benchmark model, Subsection 3.1, and the alternative model, Subsection 3.2.

We presume that there is no straightforward algorithm for computing the likelihood but that we can simulate data from $p(\cdot | \cdot, \theta)$ for given $\theta$. We presume that simulations from the structural model are ergodic. We assume that there is a transition density

$$f(y_t|x_{t-1}, \zeta), \quad \zeta \in \mathbb{Z}$$

and a map

$$g: \theta \mapsto \zeta$$

such that

$$p(y_t|x_{t-1}, \theta) = f(y_t|x_{t-1}, g(\theta)) \quad \theta \in \Theta.$$ 

We assume that $f(y_t|x_{t-1}, \zeta)$ and its gradient $(\partial / \partial \zeta)f(y_t|x_{t-1}, \zeta)$ are easy to evaluate. $f$ is called the auxiliary model and $g$ is called the implied map. When (12) holds $f$ is said to nest $p$. Whenever we need the likelihood $\prod_{t=1}^{n} p(y_t|x_{t-1}, \theta)$, we use

$$L(\theta) = \prod_{t=1}^{n} f(y_t|x_{t-1}, g(\theta)),$$

where $\{y_t, x_{t-1}\}_{t=1}^{n}$ are the data and $n$ is the sample size. After substituting $L(\theta)$ for $\prod_{t=1}^{n} p(y_t|x_{t-1}, \theta)$, standard Bayesian MCMC methods become applicable. That is, we have a likelihood $L(\theta)$ from equation (13) and a prior $\pi(\theta)$ from Subsection 4.4 and need nothing beyond that to implement Bayesian methods by means of MCMC. A good introduction to these methods is Gamerman and Lopes (2006).

The difficulty is computing the implied map, $g$, accurately enough that the accept/reject decision in an MCMC chain (Step 5 in the algorithm below) is correct when $f$ is a nonlinear model. The algorithm proposed by AG is described next.
Given \( \theta, \zeta = g(\theta) \) is computed by minimizing Kullback-Leibler divergence

\[
d(f, p) = \int \int [\log p(y|x, \theta) - \log f(y|x, \zeta)] p(y|x, \theta) \, dy \, p(x|\theta) \, dx
\]

with respect to \( \zeta \). The advantage of Kullback-Leibler divergence over other distance measures is that the part that depends on the unknown \( p(\cdot|\cdot, \theta), \int \int \log p(y|x, \theta) \, p(y|x, \theta) \, dy \, p(x|\theta) \, dx \), does not have to be computed to solve the minimization problem. We approximate the integral that does have to be computed by

\[
\int \int \log f(y|x, \zeta) \, p(y|x, \theta) \, dy \, p(x|\theta) \, dx \approx \frac{1}{N} \sum_{t=1}^{N} \log f(\hat{y}_t|\hat{x}_{t-1}, \zeta),
\]

where \( \{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^{N} \) is a simulation of length \( N \) from \( p(\cdot|\cdot, \theta) \). Upon dropping the division by \( N \), the implied map is computed as

\[
g : \theta \mapsto \arg\max_{\zeta} \sum_{t=1}^{N} \log f(\hat{y}_t|\hat{x}_{t-1}, \zeta).
\]  \hspace{1cm} (14)

We use \( N = 1000 \) in the results reported below. Results (posterior mean, posterior standard deviation, etc.) are not sensitive to \( N \); doubling \( N \) makes no difference other than doubling computational time. It is essential that the same seed be used to start these simulations so that the same \( \theta \) always produces the same simulation.

GM run a Markov chain \( \{\zeta_t\}_{t=1}^{K} \) of length \( K \) to compute \( \hat{\zeta} \) that solves expression (14). There are two other Markov chains discussed below so, to help distinguish among them, this chain is called the \( \zeta \)-subchain. While the \( \zeta \)-subchain must be run to provide the scaling for the model assessment method that GM propose, the \( \hat{\zeta} \) that corresponds to the maximum of \( \sum_{t=1}^{N} \log f(\hat{y}_t|\hat{x}_{t-1}, \zeta) \) over the \( \zeta \)-subchain is not a sufficiently accurate evaluation of \( g(\theta) \) for our auxiliary model. This is mainly because our auxiliary model uses a multivariate specification of the generalized autoregressive conditional heteroscedasticity (GARCH) of Bollerslev (1986) that Engle and Kroner (1995) call BEKK. Likelihoods incorporating BEKK are notoriously difficult to optimize. AG use \( \hat{\zeta} \) as a starting value and maximize the expression (14) using the BFGS algorithm (Fletcher (1987)). This also is not a sufficiently accurate evaluation of \( g(\theta) \). A second refinement is necessary. The second refinement is embedded within the MCMC chain \( \{\theta_t\}_{t=1}^{R} \) of length \( R \) that is used to compute the posterior distribution of \( \theta \). It is called the \( \theta \)-chain. Its computation proceeds as follows.
The $\theta$-chain is generated using the Metropolis algorithm. The Metropolis algorithm is an iterative scheme that generates a Markov chain whose stationary distribution is the posterior of $\theta$. To implement it, we require a likelihood, a prior, and transition density in $\theta$ called the proposal density. The likelihood is equation (13) and the prior, $\pi(\theta)$, is described in Section 4.4.

The prior may require quantities computed from the simulation $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$ that are used in computing equation (13). In particular, quantities computed in this fashion can be viewed as the evaluation of a functional of the structural model of the form $p(\cdot|\cdot, \theta) \mapsto \varrho$, where $\varrho \in \mathbf{P}$. Thus, the prior is a function of the form $\pi(\theta, \varrho)$. But since the functional $\varrho$ is a composite function with $\theta \mapsto p(\cdot|\cdot, \theta) \mapsto \varrho$, $\pi(\theta, \varrho)$ is essentially a function of $\theta$ alone. Thus, we only use $\pi(\theta, \varrho)$ notation when attention to the subsidiary computation $p(\cdot|\cdot, \theta) \mapsto \varrho$ is required.

Let $q$ denote the proposal density. For a given $\theta$, $q(\theta, \theta^*)$ defines a distribution of potential new values $\theta^*$. We use a move-one-at-a-time, random-walk, proposal density that puts its mass on discrete, separated points, proportional to a normal. Two aspects of the proposal scheme are worth noting. The first is that the wider the separation between the points in the support of $q$ the less accurately $g(\theta)$ needs to be computed for $\alpha$ at step 5 of the algorithm below to be correct. A practical constraint is that the separation cannot be much more than a standard deviation of the proposal density or the chain will eventually stick at some value of $\theta$. Our separations are typically $1/2$ of a standard deviation of the proposal density. In turn, the standard deviations of the proposal density are typically no more than the standard deviations in Table 2 and no less than one order of magnitude smaller. The second aspect worth noting is that the prior is putting mass on these discrete points in proportion to $\pi(\theta)$. Because we never need to normalize $\pi(\theta)$ this does not matter. Similarly for the joint distribution $f(y|x, g(\theta))\pi(\theta)$ considered as a function of $\theta$. However, $f(y|x, \zeta)$ must be normalized such that $\int f(y|x, \zeta) dy = 1$ to ensure that the implied map expressed in (14) is computed correctly.

The algorithm for the $\theta$-chain is as follows. Given a current $\theta^o$ and the corresponding $\zeta^o = g(\theta^o)$, obtain the next pair $(\theta', \zeta')$ as follows:

1. Draw $\theta^*$ according to $q(\theta^o, \theta^*)$.
2. Draw $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$ according to $p(y_t|x_{t-1}, \theta^*)$.
3. Compute $\zeta^* = g(\theta^*)$ and the functional $\varrho^*$ from the simulation $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$.
4. Compute $\alpha = \min\left(1, \frac{\varrho^*(\theta^*) \pi(\theta^*, \varrho^*) q(\theta^*, \theta^o)}{\varrho(\theta^o) \pi(\theta^o, \varrho^o) q(\theta^o, \theta^*)}\right)$. 

13
5. With probability $\alpha$, set $(\theta', \zeta') = (\theta^*, \zeta^*)$, otherwise set $(\theta', \zeta') = (\theta^o, \zeta^o)$.

It is at step 3 that AG made an important modification to the algorithm proposed by GM. At that point one has putative pairs $(\theta^*, \zeta^*)$ and $(\theta^o, \zeta^o)$ and corresponding simulations $\{\hat{y}^*_t, \hat{x}^*_{t-1}\}_{t=1}^N$ and $\{\hat{y}^o_t, \hat{x}^o_{t-1}\}_{t=1}^N$. AG use $\zeta^*$ as a start and recompute $\zeta^o$ using the BFGS algorithm, obtaining $\hat{\zeta}^o$. If

$$\sum_{t=1}^N \log f(\hat{y}^o_t | \hat{x}^o_{t-1}, \hat{\zeta}^o) > \sum_{t=1}^N \log f(\hat{y}^o_t | \hat{x}^o_{t-1}, \zeta^o),$$

then $\hat{\zeta}^o$ replaces $\zeta^o$. In the same fashion, $\zeta^*$ is recomputed using $\zeta^o$ as a start. Once computed, a $(\theta, \zeta)$ pair is never discarded. Neither are the corresponding $L(\theta)$ and $\pi(\theta, \varrho)$. Because the support of the proposal density is discrete, points in the $\theta$-chain will often recur, in which case $g(\theta)$, $L(\theta)$, and $\pi(\theta, \varrho)$ are retrieved from storage rather than computed afresh. If the modification just described results in an improved $(\theta^o, \zeta^o)$, that pair and corresponding $L(\theta^o)$ and $\pi(\theta^o, \varrho^o)$ replace the values in storage; similarly for $(\theta^*, \zeta^*)$. The upshot is that the values for $g(\theta)$ used at step 4 will be optima computed from many different random starts after the chain has run awhile.

4.1 Relative Model Comparison

Relative model comparison is standard Bayesian inference. The posterior probabilities of the models with and without ambiguity aversion are computed using the Newton and Raftery (1994) $\hat{p}$ method for computing the marginal likelihood from an MCMC chain when assigning equal prior probability to each model. The advantage of that method is that knowledge of the normalizing constants of the likelihood $L(\theta)$ and the prior $\pi(\theta)$ are not required. We do not know these normalizing constants due to the imposition of support conditions. It is important, however, that the auxiliary model be the same for both models. Otherwise the normalizing constant of $L(\theta)$ would be required. One divides the marginal density for each model by the sum for both models to get the probabilities for relative model assessment.

4.2 Forecasts

A forecast is a functional $Y : f(\cdot | \cdot, \zeta) \mapsto \upsilon$ of the auxiliary model that can be computed from $f(\cdot | \cdot, \zeta)$ either analytically or by simulation. Due to the map $\zeta = g(\theta)$, we view such a forecast as both a forecast from the structural model and as a function of $\theta$. Viewing it as a function of $\theta$, we can compute $\upsilon$ at each draw in the posterior MCMC chain for $\theta$ which results in an MCMC chain for
From the latter chain and the mean and standard deviation of \( \nu \) can be computed. The same quantities can be computed for draws from the prior. Examples are in Figure 4.

### 4.3 The Auxiliary Model

The observed data are \( y_t \) for \( t = 1, \ldots, n \), where \( y_t \) is a vector of dimension three in our application. We use the notation \( x_{t-1} = \{ y_{t-1}, \ldots, y_{t-L} \} \), if the auxiliary model is Markovian, and \( x_{t-1} = \{ y_{t-1}, \ldots, y_1 \} \) if it is not.\(^5\) Either way, \( x_{t-1} \) serves as a shorthand for lagged values of \( y_t \) vector. In this application, the auxiliary model is not Markovian due to the recursion in expression (17).

The data are modeled as

\[
y_t = \mu_{x_{t-1}} + R_{x_{t-1}} z_t
\]

where

\[
\mu_{x_{t-1}} = b_0 + B x_{t-1},
\]

which is the location function of a vector auto-regressive (VAR) specification, and \( R_{x_{t-1}} \) is the Cholesky factor of

\[
\Sigma_{x_{t-1}} = R_0 R'_0 + Q \Sigma_{x_{t-2}} Q' + P' (y_{t-1} - \mu_{x_{t-2}}) (y_{t-1} - \mu_{x_{t-2}})' P' + \max[0, V(y_{t-1} - \mu_{x_{t-2}})] \max[0, V(y_{t-1} - \mu_{x_{t-2}})]'.
\]

In computations, \( \max(0, x) \) in expression (19), which is applied elementwise, is replaced by a twice differentiable cubic spline approximation that plots slightly above \( \max(0, x) \) over \((0.00, 0.10)\) and coincides elsewhere.

The density \( h(z) \) of the i.i.d. \( z_t \) is the square of a Hermite polynomial times a normal density, the idea being that the class of such \( h \) is dense in Hellenger norm and can therefore approximate a density to within arbitrary accuracy in Kullback-Leibler distance (see Gallant and Nychka (1987)). Such approximations are often called sieves; Gallant and Nychka term this particular sieve semi-nonparametric maximum likelihood estimator, or SNP. The density \( h(z) \) is the normal when the degree of the Hermite polynomial is zero. In addition, the constant term of the Hermite polynomial

\(^5\) Refer to Gallant and Long (1997) for the properties of estimators of the form used in this section when the model is not Markovian.
can be a linear function of $x_{t-1}$. This has the effect of adding a nonlinear term to the location function (15) and the variance function (16). It also causes the higher moments of $h(z)$ to depend on $x_{t-1}$ as well. The SNP auxiliary model is determined statistically by adding terms as indicated by the Bayesian information criteria (BIC) protocol for selecting the terms that comprise a sieve, see Schwarz (1978).

In our specification, $R_0$ is an upper triangular matrix, $P$ and $V$ are diagonal matrices, and $Q$ is scalar. The degree of the SNP $h(z)$ density is four. The constant term of the SNP density does not depend on the past.

The auxiliary model chosen for our analysis, based on BIC, has 1 lag in the conditional mean component, 1 lag in each of autoregressive conditional heteroscedasticity (ARCH) and generalized autoregressive conditional heteroscedasticity (GARCH) terms. The model admits leverage effect in the ARCH term. The auxiliary model has 37 estimated parameters.

The implied error distributions in GSM estimation can differ significantly from those used for solving the structural model. For example, we numerically solve the two structural models in sections 3.1 and 3.2 assuming normal distributions for error terms in equations (1) and (2). This is a simplifying assumption to ease numerical solutions. The empirical error distributions of these models are non-Gaussian. For example, in addition to GARCH and leverage effects, the three-dimensional error distribution implied by the benchmark smooth ambiguity aversion model is skewed in all three components and has fat-tails for consumption growth and stock returns and thin tails for bond returns.

The auxiliary model is determined from simulations of the structural model so issues of data sparsity do not arise; one can make the simulation length $N$ as large as necessary to determine the parameters of the auxiliary model accurately. As stated above, we used $N = 1000$ and found that using larger values of $N$ did not change results other than increase run times.

### 4.4 The Prior and Its Support

Both the benchmark and alternative models are richly parameterized. For the benchmark model, parameter is

$$\theta = (\beta, \gamma, \psi, \eta, p_{11}, p_{22}, \kappa_1, \kappa_2, \lambda, \sigma_{\Delta c}, \sigma_{\Delta d}).$$
The prior is the combination of the product of independent normal density functions and support condition. For annual data, the benchmark model prior location parameters are

\[
\theta^\# = (0.9750, 2.0, 0.6667, 8.8640, 0.9780, 0.5160, -0.06785, 0.02251, 2.7400, 0.03127, 0.1200).
\]

The scale parameters, i.e., the standard deviations, are \( \sigma^\# = (0.90/1.96)\theta^\# \). The implication of this choice of standard deviation is that prior probability satisfies \( P(|\theta_i - \theta_i^\#|/|\theta_i^\#| < 0.90) = 0.95 \), i.e., the probability being within ninety percent of \( \theta_i^\# \) is 0.95. This is a loose prior so that the major determinant of the prior are support conditions described next and the exclusion of parameter values for which an equilibrium does not exist. Imposition of exclusion restrictions provides room for the theory to contribute to the estimated parameters' identification. Due to the support and exclusion restrictions, the effective prior is not an independence prior.

We constrain the subjective discount factor \( \beta \) to be between 0.00 and 1.00. We bound the coefficient of risk aversion \( \gamma \) to be above 0.00 and below 15.00. In line with recommendations of Mehra and Prescott (1985), the mean \( \theta_2^\# \) of the prior distribution for this coefficient is set to 2.00 but, as noted above, all that matters is the support conditions because \( \sigma_2^\# \) is large.

Fully parameterized Kreps and Porteus (1978) and Epstein and Zin (1989) preferences imply a separation between risk aversion and intertemporal substitution.\(^6\) To avoid a pathological case when per chance \( \gamma = 1/\psi \), we require that intertemporal elasticity of substitution parameter \( \psi \) cannot assume 0.00. We set \( \psi \) to be above 0.20.

Based on the findings in previous empirical studies such as Aldrich and Gallant (2011) and theoretical considerations of the Ju and Miao (2012) model, we require positive leverage in the model. To this end, we impose \( \lambda > 0 \).

We bound \( \eta \) between 2.00 and 100.00. We impose \( \eta > \gamma \) on coefficient of ambiguity aversion \( \eta \) in the benchmark model. We need this restriction for ambiguity aversion to exist. Hayashi and Miao (2011) and Ju and Miao (2012) furnish detailed discussions of this requirement. Briefly, with \( \eta \leq \gamma \), compound predictive probability distributions are reduced to an ordinary predictive probability, removing ambiguity from the model and leaving no room for ambiguity aversion to play a part in agent’s allocations. To preserve concavity in attitudes toward ambiguity, we need \( \eta > \gamma \). When estimating the alternative model, we impose the linear restriction \( \eta = \gamma \) to obtain

\(^6\) If \( \gamma = 1/\psi \), Kreps-Porteus and Epstein-Zin preferences collapse to power utility. When \( \gamma \neq 1/\psi \), that is when these preferences are fully parameterized, risk aversion and intertemporal elasticity of substitution are separate.
ambiguity neutrality, in which case \( \theta_2^# = \theta_4^# = 5.432 \) and \( \sigma_2^# = \sigma_3^# = (0.9/1.96)(5.432) \); all else is the same.

We constrain \( 0.93960 < p_{1,1} < 0.99962 \), \( 0.2514 < p_{2,2} < 0.7806 \), \( 0.01596 < \kappa_1 < 0.02906 \), \( -0.1055 < \kappa_2 < -0.0302 \), \( 0.0 < \lambda \), \( 0.02646 < \sigma_{\Delta e} < 0.03608 \), and \( 0.06542 < \sigma_{\Delta d} < 0.1746 \) based on numerical experience acquired with these model in previous studies.

The net result is a fairly loose prior: Compare location and scale between the posterior an prior distributions in Table 2. There are considerable shifts in both location and scale: The data has a strong influence on our results. This observation is reassuring, since an important concern in Bayesian estimation is identification of the estimated parameters. In other words, one wants to know the relative contribution of priors and support conditions versus the contribution of the data. Figures 2 – 5 provide a visual representation of the contribution of our data. It is immediately clear that for almost all parameters, posterior densities shift significantly in comparison with prior densities. Moreover, shifts in posterior densities of the key parameters in our study, \( \eta, \gamma, \psi \), and \( \lambda \), are large enough across models and sampling frequencies. The size of these shifts dispel the notion that the identification of these parameters solely rest on tight priors. Indeed, even after combining our relatively loose priors with support conditions for parameters and imposition of exclusion restrictions regarding existence of an equilibrium given a given estimated \( \theta^* \), the main source of identification is still data as seen on Figures 2 – 5.

In general, this prior is in line with other Bayesian macroeconomic studies. Measures of location and scale for the prior for the benchmark and alternative model are reported in columns 2-4 and 8-10 of Table 2, respectively. The mean is what is typically reported. However, the model is never actually simulated at the mean. The mode has the advantage that it must satisfy support conditions and the model has been simulated at that value.

For quarterly data, an appropriately rescaled version of the annual prior is used. See the lower panel of Table 2.

### 4.5 Empirical Results

We report estimation results for the benchmark model featuring ambiguity aversion and the alternative model with ambiguity neutral agent in Table 2 for both annual and quarterly-sampled data. We estimate a total of 11 parameters for the benchmark model. Log dividend growth process, \( \Delta d_t \), is the latent variable in our estimation. Estimation results for the benchmark structural model
featuring ambiguity aversion and learning are reported in columns 5 to 7 in Table 2. We report estimated parameters for the alternative structural model with no ambiguity aversion in the last three columns of Table 2. Estimation results for annually sampled data appear in Panel A of Table 2, and those for quarterly sampled data appear in Panel B.

Estimates of mode and mean measures of subjective discount factor $\beta$ are stable across our benchmark and alternative models and between annual and quarterly-sampled results. Moreover, they are reasonably close to values reported by Aldrich and Gallant (2011) and Bansal et al. (2007). Thus, they do not cause any concern for us and imply precise estimation of the target parameter.

We observe the following regularities in estimated parameters: The magnitude of risk aversion parameter, $\gamma$, is sensitive to the presence of ambiguity aversion. The mode and the mean of the posterior of the estimated $\gamma$ based on annual data in the alternative model are over 3 times larger than the value of the corresponding estimate in the benchmark model with ambiguity. When we use quarterly-sampled data, estimated mean and mode of $\gamma$ in alternative model with no ambiguity aversion are an order of magnitude larger than those for the benchmark model featuring ambiguity aversion. This result is related to estimation outcomes reported in Jeong et al. (2015) for their baseline models II (recursive utility) and III (MPU), where aggregate wealth consists of financial wealth alone. Using MPU, Jeong et al. (2015) report $\gamma$ ranging between 0.20 to 2.90, while using only recursive utility this value is 4.90. Thus, mode and mean values for $\gamma$ equal to 1.6172 and 1.6264 (annual) and 1.3672 and 1.3237 (quarterly) are in agreement with Jeong et al. (2015) estimates. In comparison with Aldrich and Gallant (2011), our estimates for $\gamma$ in the benchmark model – regardless of the sampling frequency – are smaller than what they report for both habit formation and long-run risk models, but similar to prospect theory-based results. The difference between estimates of mode and mean of posterior values of $\gamma$ in our alternative model and the LRR model in Aldrich and Gallant (2011) are non-negligible. Our quarterly estimates are larger and our annual estimates are smaller.

Under the benchmark model, we obtain estimates for the mode and the mean of IES, $\psi$, that are larger than unity as advocated by the long run risk literature. This regularity holds across estimates based on annual and quarterly-sampled data, with the difference that these estimated values are much larger than those reported by Aldrich and Gallant (2011), which are in the neighborhood of 1.50. Our mode and mean estimates under both benchmark and alternative models are significantly

We force $\eta = \gamma$ throughout the estimation. Thus, parameter space $\theta$ for the alternative model featuring ambiguity neutrality has 10 parameters.
more stable than values reported by Jeong et al. (2015). Their estimated $\psi$s, across five models and two assumptions for volatility dynamics (time-varying volatility and nonlinear stochastic volatility) range between 0.00 to $\infty$. In their benchmark MPU model, $\psi$ estimates (when all parameters are estimated from the data and wealth is only a function of returns to financial investments) are equal to 0.68 with time-varying volatility and 11.16 with nonlinear stochastic volatility. When wealth is assumed as the sum of financial and labor income, and hence some parameters are calibrated, their EIS parameter estimates range between 1.17 to 15.13. As such, our estimates of posterior means that range between 2.6172 and 4.9141 are both more stable and more plausible, especially in comparison with EIS estimates in financial wealth-only, fully estimated cases in Jeong et al. (2015) that are directly comparable to our findings.

The EIS parameter $\psi$, as discussed in Liu and Miao (2015), is a crucial parameter for matching macroeconomic and financial moments. The risk aversion parameter and the EIS both determine the representative agent’s preference for the timing of resolution of uncertainty. If $\gamma > 1/\psi$, the agent prefers an early resolution of uncertainty (see Epstein and Zin (1989) and Bansal and Yaron (2004)). By this measure, both benchmark and alternative models point to a representative agent who desires an early resolution of uncertainty. Based on our estimation results, and as expected due to the impact of ignoring ambiguity aversion on the estimated value of risk aversion parameter, this effect is stronger in the alternative model.

We report the first direct estimates of the ambiguity aversion parameter $\eta$ based on the U.S. financial data. The posterior mode and mean values of this parameter are 29.80 and 30.33 for annually sampled data and 57.08 and 55.93 for quarterly sampled data in our benchmark model. In comparison with calibration exercises in the literature, our estimates are larger than calibrated values in an endowment economy ($\eta = 8.86$ in Ju and Miao (2012)) and in a production economy ($\eta = 19$ in Jahan-Parvar and Liu (2014)). Obviously, our estimated results meet the criteria that $\eta \gg \gamma$. Our estimated results for ambiguity aversion parameter imply that the mapping between annual and quarterly estimates of $\eta$ may not be linear and that agents are more ambiguity averse in higher sampling frequencies. Intuitively, a higher sampling frequency for data collection implies a shorter decision horizon and more frequent shocks to beliefs ($\tilde{\mu}_t$) for the agent, thus more room

---

8 In the empirical literature, some papers (e.g. Hall (1988) and Ludvigson (1999)) find that the EIS estimate is close to zero using aggregate consumption data. Other papers find higher values using cohort- or household-level data (e.g., Attanasio and Weber (1993) and Vissing-Jorgensen (2002)). Attanasio and Vissing-Jorgensen (2003) find that the EIS estimate for stockholders is typically above 1. Bansal and Yaron (2004) argue that estimates of the EIS based on aggregate data will be biased down if the usual assumption that consumption growth and asset returns are homoscedastic is relaxed.
Ahn et al. (2014) study smooth ambiguity aversion as part of their experimental research on ambiguity. They report values for the ambiguity aversion parameter based on the static formulation of smooth ambiguity aversion ranging between 0.00 and 2.00, with mean value of 0.207 for all subjects in their experiment population. They choose their population such that both “ambiguity neutral” and “ambiguity loving” subjects are represented. Thus, estimates for ambiguity aversion parameter for 5 to 50th percentiles of their population are zero. In addition, the value of ambiguity aversion parameters – regardless of whether smooth or kinked specifications are estimated – are at least an order of magnitude smaller than dynamic model-based estimates such as ours.\footnote{We find that the difference in the magnitude of these estimates is similar to the difference between static estimates of Gul (1991) disappointment aversion parameter reported by Choi, Fisman, Gale, and Kariv (2007) and dynamic estimates reported by Feunou, Jahan-Parvar, and Tédongap (2013). Thus, these differences are more likely to be an outcome of ignoring the dynamics in the data, rather than a result of using alternative estimation methods. For example, while we use GSM Bayesian methodology, Feunou et al. (2013) implement a frequentist maximum likelihood estimator. Yet, similar discrepancies between statically and dynamically estimated preference parameters are present.} We believe that ignoring intertemporal dimensions of choice under ambiguity explains these differences in the magnitude of estimated parameters. As mentioned earlier, while estimated ambiguity aversion parameters based on kinked specification are statistically different from zero in Ahn et al. (2014), the estimated smooth ambiguity parameters are not. In comparison with these findings, $\eta$ in our benchmark model is tightly estimated – standard deviations of the posterior are less than 15 percent of the value of the posterior mean.

Thimme and Völkert (2015) use quarterly data to estimate smooth ambiguity parameter based on partial calibration of other parameters (specifically, they calibrate $\psi$) and a linearized SDF. Their estimates based on $\psi = 2 \ (I_{ES} = 0.5)$ and $\psi = \infty \ (I_{ES} = 0.00)$, yield $\eta = 35.09$ and 61.03 respectively. These partially estimated values are comparable in magnitude with our quarterly estimation results, but fall deep in the tails of the posterior distribution of $\eta$.

Jeong et al. (2015) report an indirect measure for ambiguity aversion based on MPU equal to 0.34. Their estimation result is not directly comparable to ours, since it is based on a very different functional form and underlying assumptions. In fact, it is best described as a lower bound on the beliefs about the probability of the state of the economy in the MPU model, and not a measure for ambiguity aversion. In other words, this estimated value is the bound for what is called the density generator, a stochastic process defining the Girsanov density of the subjective probability measure of an agent with respect to the objective probability measure. Similarly, Viale et al. (2014) characterize cross-sectional ambiguity as the likelihood ratio between MPU-distorted and reference
models. Ilut and Schneider (2014) and Bianchi et al. (2014) present estimates of MPU ambiguity measures in DSGE settings. Ilut and Schneider (2014) posits an autoregressive and exogenous process for ambiguity, while Bianchi et al. (2014) – similar to Jeong et al. (2015) – find bounds for beliefs about uncertainty. As mentioned earlier, these findings are not directly comparable with ours.

Our estimates of the transition probabilities \((p_{11} \text{ and } p_{22})\), low and high mean consumption growth \((\kappa_1 \text{ and } \kappa_2)\), and the volatility of consumption growth \((\sigma_{\Delta c})\) are close to empirical results reported in other studies such as Cecchetti, Lam, and Mark (2000) for both benchmark and alternative models. The main difference lies with the benchmark model which yields lower estimated values for \(p_{22}, \kappa_1\) and \(\sigma_{\Delta c}\). The difference between Cecchetti et al. (2000) estimates of these parameters and ours are not negligible. In particular, the difference between Cecchetti et al. (2000) estimate of \(p_{22}\) and ours points to about 50% difference. However, this is not surprising. Cecchetti et al. (2000) estimates (reported in Table 2, page 790 in their paper) are based on a reduced-form fitting of the data. Ours are based on a structural model. Besides, we use different data and methodology. Our data set is both more recent (1929-2011 vs. 1871-1993) and shorter.

Our estimates for leverage ratio \((\lambda)\) and the volatility of dividend process \((\sigma_{\Delta d})\) are not directly comparable to estimates reported in Aldrich and Gallant (2011), Gallant and McCulloch (2009), or Bansal et al. (2007) due to different specifications for modeling dividend growth.\(^{10}\) Specifically, the LRR model features time variation in the volatility of fundamentals, while we rely on Markov-switching and distortions in state beliefs to deliver the time-variation in the volatility of returns.\(^{11}\) However, our estimated annual \(\sigma_{\Delta d}\) is close in magnitude to the volatility of dividend process in the prospect theory model reported by Aldrich and Gallant (2011). In their formulation of a prospect theory-based asset pricing model, Aldrich and Gallant posit constant volatility for this process. Thus, magnitudes of estimated parameters are comparable, since estimation methodology is essentially the same in both studies.

In summary, apart from estimates for risk aversion parameter, \(\gamma\), and ambiguity aversion parameter, \(\eta\), other estimated parameters in our study are remarkably stable and are generally comparable in magnitude to values reported by other empirical asset pricing studies. Thus, it is reasonable to believe that risk aversion and ambiguity aversion parameters deserve special attention. The

\(^{10}\)Thimme and Völkert (2015) do not report estimates for the leverage ratio.

\(^{11}\)Jahan-Parvar and Liu (2014) discuss this feature of asset pricing models based on smooth ambiguity aversion preferences in detail both theoretically and based on simulation exercises.
rest of the estimated parameters are stable, comparable to previous estimates in the literature (especially those in the long-run risk literature), and have negligible influence in identification and model comparison when it comes to model featuring smooth ambiguity aversion preferences. We thus conclude that: (a) Estimates of the ambiguity aversion parameter are sensitive to sampling frequency; (b) the mapping of \( \eta \) estimates across different sampling frequencies is most likely non-linear; (c) ignoring ambiguity aversion results in asset pricing models leads to biased estimates of the risk aversion parameter, \( \gamma \); and finally, (d) the size of this bias in estimates of \( \gamma \) increases with sampling frequency.

5 Model Comparisons and Model Implications

5.1 Relative Model Comparison

Relative model comparison is standard Bayesian inference as described in Subsection 4.1. The computed odds ratio is \( 1/8.431e-107 \) for annual data-based estimation and \( 1/2.192e-95 \) for quarterly data-based estimates, which strongly favors the benchmark model over the alternative model. This ratio implies that our benchmark model provides a better description of the available data in the framework of the equilibrium model discussed in Section 3. Given the values of the log likelihoods for the benchmark and alternative models reported in Table 2 one hardly needs to bother with odds ratios. The verdict is obvious.

5.2 Forecasts

Forecasts are constructed as described in Subsection 4.2. Prior forecasts (not shown) do not differ much between pre- and post-Great Recession periods. There are, however, differences between prior forecasts based on the benchmark model and the alternative model. The main difference is the disparity in the level of benchmark and alternative model-based forecasts of the short rate. The benchmark model forecasts a higher level for the short rate (and wider posterior standard deviations) than the alternative model. These forecasts are counterintuitive, since we expect the agent to have a higher demand for a safe asset that pays the short rate, and hence a lower short rate. Once we take the forecast standard deviations into account, they appear less counterintuitive. The second difference is the slight increase in consumption growth path forecast by the benchmark model, against the drop in consumption growth path forecasts by the alternative model.
Prior forecasts are not a measure of a model’s success in capturing the data dynamics. For that purpose, we rely on posterior forecasts, which we report in Figure 4.\textsuperscript{12} As Figure 4 shows, consumption growth forecasts differ both between the benchmark and the alternative model, and across pre- and post-Great Recession episodes. Both observations are in line with the theory and our expectations. Our discussion is based on comparing mean posterior forecasts. The posterior forecast paths generated by both modes are on average similar. However, the benchmark model implies slightly more variation in consumption growth in the future.

The benchmark model yields little variation in consumption growth forecasts. Both benchmark and alternative models forecast a slight drop in consumption growth for pre-recession period. Similarly, both models forecast a relatively flat trajectory for consumption growth based on available information by the end of 2011.

In pre-recession period, the benchmark model forecasts a steeper drop in stock returns in comparison with the alternative model, roughly 10% drop against 5%, respectively. For the post-Great Recession period, the benchmark model yields stock return posterior forecasts that are 50% lower than the alternative model. Simply put, ambiguity aversion implies a modest equity premium (around 5%) for the foreseeable future, while ambiguity neutral alternative model predicts a bull market – equity premium close to 10%. The difference, reflects the role of the third term in SDF equation (6) for the benchmark model, and its absence in the SDF for the alternative model, both between models and across forecast periods. This gap in forecasts between the two models is in line with earlier findings. For example, Aldrich and Gallant (2011) report forecasts of roughly 6% for equity returns for 2009-2013 period for the long-run risk model of Bansal and Yaron (2004), based on data ending in the Great Recession period, which may be viewed as high given the recent past experience.

It is clear from this figure that the benchmark model predicts both a drop in the risk-free rate and an overall lower risk-free rate in comparison with the predictions of the alternative model, across pre- and post-Great Recession periods.\textsuperscript{13} This empirical regularity echoes the findings of earlier theoretical research. Ambiguity aversion implies a more pessimistic attitude, and as result, a higher precautionary saving demand than the precautionary demand that Epstein and Zin utility induces.

\textsuperscript{12}There is a rather dramatic change in the standard errors of the forecasts between prior and posterior that suggests that the data is quite informative. Compare also the location and scales for the prior and posterior values reported in Table 2 and prior and posterior densities of estimated parameters reported in Figures 2 to 5.

\textsuperscript{13}While this observation is in line with the zero-lower bound environment since the Great Recession, they should not be viewed as synonymous. We are forecasting real risk-free rates. They are not influenced by fiscal or monetary policy, and are endogenously determined.
in the alternative model. As a result, the prices of the risk-free bonds are higher, leading to lower yields. The difference between forecasts are not negligible: In both pre- and post-recession periods, it amounts to roughly 1% in real interest. Given recent announcements by various practitioners, academicians and former policy makers about likelihood of interest rates reverting back to “old normal” levels, our benchmark model forecasts seem reasonable.\footnote{For example, according to Reuters on May 16, 2014, former Federal Reserve Chairman Ben Bernanke opined that low interest rate environment is likely to continue beyond many current forecasts.}

Given that Bayesian model comparison prefers the benchmark model over ambiguity neutral alternative model, these forecasts merit attention. The two models lead to very different dynamics for consumption growth and asset returns. If we indeed live in a world populated by ambiguity averse agents – implied by our results – then policy and decision makers need to be aware of the inherently different attitudes (and hence, market behavior) of agents endowed with ambiguity aversion preferences, as opposed to those assumed in standard rational expectation models.

5.3 Asset Pricing Implications

In this section, we discuss the asset pricing implications of our estimated models. As mentioned earlier, our benchmark model is the exchange economy of Ju and Miao (2012). We are interested in how closely the benchmark and alternative model, calibrated at the equilibrium using our estimated parameters, can generate salient features of the historical equity premium which did not inform our estimation exercise. If our calibration study is reasonably successful in capturing such empirical features and/or improves the asset pricing model introduced by Ju and Miao in directions not explicitly targeted by our study, then the dynamics of the model implied by our estimated parameters are reasonably close to the underlying data generating process (DGP). In the first step, we examine how closely asset pricing quantities generated by calibrating the benchmark and alternative models using estimated parameters, reported in Table 2, match sample moments reported in Table 1. In addition, we compare the performance of this calibration with results reported in the original study by Ju and Miao. Besides, we study the properties of conditional moments of equity returns implied by our estimated parameters. Next, we investigate quantities such as variance risk premium – the premium that the agent demands to bear changes in volatility of returns – and its moments implied by our estimated models, as formulated and studied by Bollerslev et al. (2009). Finally, we study the scope for ambiguity aversion, given estimated parameters, using detection error probabilities (DEP) of Anderson et al. (2003) and Hansen and Sargent (2010). In what follows, we report
model generated values for asset pricing quantities, based on models simulated at mean parameter estimates reported in Table 2.

In the first four rows of Table 3, we report model generated unconditional means and standard deviations of risk-free rate, $r_t^f$ and equity premium, $r_t^e - r_t^f$. We compare these values with those reported in columns 2 and 4, Table 1. In comparison with sample data presented in Table 1, and with Ju and Miao (2012), we observe the following.\textsuperscript{15} First, the benchmark model generates risk-free rate moments that are much closer to sample moments than the alternative model.\textsuperscript{16} While the volatility of risk-free rate is largely controlled by the magnitude of the IES parameter, introduction of ambiguity aversion seems to improve the model generated volatility in risk-free rates compared with the alternative model.

Second, while both benchmark and alternative models yield model generated volatilities for equity premium that are close to the sample volatility for this quantity, they differ dramatically in terms of their performance in matching the mean equity premium. In our sample, the mean equity premium value is 8.54% for annual data and 8.76% for quarterly data. At 7.31%, the benchmark model based on annual data generates a large mean equity premium. The corresponding value for quarterly data is even higher at 12.68%, which may be viewed as unrealistic. However, both values are more realistic and much closer to sample means than 1.36% annual and 18.89% quarterly mean equity premia generated by the alternative model. As documented in Bansal and Yaron (2004), it is a well-known fact that without high risk aversion parameter values, Epstein and Zin (1989) recursive preferences have difficulty in matching the mean equity premium. In their study, Bansal and Yaron have to set $\gamma$ equal to 10 – at the end of admissible range suggested by Mehra and Prescott (1985) – to match the mean equity premium. Since the estimated $\gamma$ for the annual alternative model is smaller ($\mathbb{E}(\gamma) = 6.32$ and $\sigma(\gamma) = 0.26$), the alternative model falls into this well-documented trap.

In rows 5 and 6 of Table 3, we report Sharpe ratios and the ratio of unconditional volatility of the SDF to unconditional expected value of the SDF, $\sigma(M_t)/\mathbb{E}(M_t)$, which is interpreted as the market price of risk. The Sharpe ratio implied by annual data in our sample is 0.37. The same quantity for quarterly data equals 0.47. Thus, it is immediately clear that the benchmark model generates Sharpe ratios that are far more realistic than those implied by the alternative model.

\textsuperscript{15}Ju and Miao (2012) use data from 1871 – 1993 sample period. Thus, our findings are not directly comparable with theirs.

\textsuperscript{16}The negative values for expected risk-free rates in quarterly data is driven by low values of the estimated mean consumption growth parameter, $\kappa_2$. Changing this value to a slightly larger value, say $-0.15$, would eliminate these negative expected real risk-free rates.
The market price of risk, defined as \( \sigma(M)/\mathbb{E}(M) \), is closely related to the moments of asset returns via the Hansen-Jagannathan bound. Ju and Miao (2012) find that the market price of risk is about 0.60 as implied by the calibration with ambiguity aversion. In a production setting, Jahan-Parvar and Liu (2014) find that the market price of risk is about 0.94 with ambiguity aversion. In the data, the Sharpe ratio range is roughly between 0.40 – 0.50. The alternative model implies that the market price of risk is 0.28 for annual data and 7.45 for quarterly data. Thus, the model implied market price of risk for annual data is lower than the empirical Sharpe ratio and violates the Hansen-Jagannathan bound. On the other hand, the annual benchmark model generates a market price of risk of 2.63, which satisfies the Hansen-Jagannathan bound and also enables the model to match the key financial moments.

In rows 7 to 10 of Table 3, we report the unconditional moments of variance risk premium (VRP) as defined by Bollerslev et al. (2009), implied by the models calibrated at means of estimated parameters in our study. Miao, Wei, and Zhou (2012) and Liu and Zhang (2015) study the properties of the VRP in consumption-based and production-based asset pricing models featuring ambiguity averse agents, respectively. Briefly, VRP is defined as the difference between option-implied, risk neutral volatility and expected physical volatility measures of stock or aggregate index returns. In practice, risk-neutral volatility is most often measured through using the Chicago Board Options Exchange’s volatility index (VIX). The most common measures of expected physical volatility are forecasts of realized volatility. VRP is a well-studied, significant predictor of short term (typically one-month to three-months ahead) market index returns. Liu and Zhang (2015) provide a discussion of the difficulties that standard consumption- and production-based asset pricing models encounter in reproducing stylized facts observed in VRP data. The first problem is the inability of standard models to reproduce large enough the average and standard deviation of the VRP. This problem is often referred to as the “variance risk premium puzzle.” Second, the empirical distribution of the VRP has non-zero skewness and high kurtosis. We follow the methodologies of Miao et al. (2012) and Liu and Zhang (2015) to produce VRP in our study.

The annualized unconditional first four moments of monthly VRP, based on 1990 – 2013 data are 17.97, 21.56, 3.84 and 25.18, respectively.\(^\text{17}\) While these numbers are not directly comparable with our computations, they provide a benchmark for expected magnitudes of the computed VRP moments. As is seen on Table 3, VRP moments based on annual data parameter estimates are low:

\(^{17}\)We thank Hao Zhou for making the time-series of VRP available on his website.
Computed volatilities of the VRP are an order of magnitude smaller than their expected values for the benchmark model, and an order of magnitude smaller than the two higher moments for the alternative model. These problems do not seem to afflict quarterly data-based expected VRP and standard deviations of VRP for both benchmark and alternative models. While one may argue whether the size of these quantities are reasonable, the ratio of their magnitudes are reasonable. Similarly, the reported kurtosis values for both benchmark and alternative model generated VRP at annual or quarterly frequencies are reasonable, from the relative magnitude point of view. In contrast, computed skewness for annual and quarterly VRP measures across both benchmark and alternative models are of the same order of magnitude as mean VRP, in contrast to Zhou’s reported values.

Figure 6 provides additional visual evidence in support of adequacy of the benchmark model to capture the un-targeted features of underlying variables’ dynamics. In Panel B of this figure, we report the expected consumption growth path, implied by the benchmark model and calibrated for quarterly data parameter estimates. It is clear that first, the expected consumption growth is highly correlated with ambiguity-distorted beliefs reported on Panel A. Second, they both are nicely correlated with the NBER recessions. That is, the size of ambiguity-distortions and drops in conditional consumption growth rates closely follow the severity of NBER recessions. This implies that the main driver of expectations about future consumption growth is the ambiguity-induced distortions in beliefs about the future states of the economy. Not surprisingly, as observed on Panel C of this figure, the dynamics discussed earlier imply conditional equity premia that are highly correlated with the severity of NBER recessions – and mainly driven by ambiguity aversion. Similarly, the expected path of the VRP implied by the quarterly benchmark model implies significant rises in the conditional VRP during recession periods that are proportional to severity of these downturns.

Finally, an important question is: Do our structural estimations imply reasonable magnitudes for ambiguity aversion? To address this question, we use detection-error probabilities (DEP) to assess the room allowed for ambiguity aversion based on our structural estimation results. This exercise is meaningful because our estimation is grounded in the data and thus is more informative about the behavior of economic agents and the dynamics of economic variables. For comparison, we also calculate detection-error probabilities using the calibrated parameter values in Ju and Miao (2012).
Detection-error probabilities are an approach developed by Anderson, Hansen, and Sargent (2003) and Hansen and Sargent (2010) to assess the likelihood of making errors in selecting statistically “close” (in terms of relative entropy) data generating processes (DGP). In this paper, a DGP refers to a Markov switching model for consumption growth as specified in Equation (1). Without ambiguity aversion, transition probabilities are defined as in the transition matrix $P$ in Section 3.1, and in this case we obtain the reference DGP. However, ambiguity aversion implies distortion to the transition probabilities in a pessimistic way and thus gives rise to the distorted DGP. The Appendix shows that the reference DGP and the distorted DGP differ only in terms of transition probabilities. We adapt to the current endowment economy the approach of computing detection-error probabilities in Jahan-Parvar and Liu (2014). This approach enables us to simulate artificial data from the reference and distorted DGPs and to evaluate the likelihood explicitly. Details of the algorithm of computing detection-error probabilities are included in the Appendix.

A sizable detection-error probability associated with a certain value of the ambiguity aversion parameter, $\eta$, implies that there is a large chance of making mistakes in distinguishing the reference DGP from the distorted DGP, and thus ample room exists for ambiguity aversion. We report computed DEP values – when applicable – in the row above expected risk-free rate on Table 3. Based on our annual and quarterly data-based estimated parameters reported on Table 2, the detection-error probability is 10.22% for annual data and 13% for quarterly data. In comparison, using the parameter values that Ju and Miao used in their calibration exercise yields a DEP equal to 0.45%. We conclude that using our estimated parameter values admits a larger scope for ambiguity aversion in asset pricing models. This finding implies that the ambiguity-averse agent in our estimation faces significant difficulty is distinguishing between the reference and distorted DGPs.

6 Conclusion

Smooth ambiguity preferences of Klibanoff et al. (2005, 2009) have gained considerable popularity in recent years. In part, this popularity is due to clear separation between ambiguity – a characteristic of a representative agent’s subjective beliefs – and ambiguity aversion that derives from the agent’s tastes. In this paper, we estimate the endowment equilibrium asset pricing model with smooth ambiguity preferences proposed by Ju and Miao (2012) using U.S. data and GSM Bayesian
estimation methodology of Gallant and McCulloch (2009) to a) investigate the empirical properties of such an asset pricing model as an adequate characterization of the returns and consumption growth data and, b) provide an empirical estimation of the ambiguity aversion parameter and its relationship with other structural parameters in the model. Our study contributes to the existing literature by providing a formal empirical investigation for adequacy of this class of preferences for economic modeling, and presenting estimations for the structural parameters of this model. The estimated structural parameters are in line with theoretical expectations, and are comparable with estimated parameters in studies using similar or related estimation methods. With respect to measurement of ambiguity aversion, our results show a marked improvement over the existing literature. The existing empirical literature either provides measures of ambiguity (which is usually the size of the set of priors in the MPU framework) – but not ambiguity aversion of the agent – or statistically implausible estimates for smooth ambiguity aversion parameters. Our study addresses both shortcoming in the extant literature.

We find that Bayesian model comparison strongly favors the benchmark model over the alternative model featuring Epstein and Zin recursive preferences. In addition, we find that the estimated ambiguity aversion parameter is higher than the values in calibration studies. We explore forecasting and asset pricing implications using our estimated model. We find marked differences in forecasts generated from the benchmark model and the alternative model. In short, the benchmark model generates more conservative forecasts for equity premium and real interest rates in comparison with the alternative model. We find that based on the estimated parameters, the equilibrium asset pricing model can successfully reproduce main stylized facts about asset returns. In addition, detection-error probabilities computed using the estimated parameters imply ample scope for ambiguity aversion.
References


7 Appendix: Detection Error Probabilities

• In constructing distorted transition probabilities, we consider a “full information model”, where the agent is ambiguity averse but state $z_t$ is observable. In this case, the Euler equation is

$$0 = p_{11} E_{1,t} \left[ M_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right] + (1 - p_{11}) E_{2,t} \left[ M_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right]$$

for $z_t = 1$ and

$$0 = (1 - p_{22}) E_{1,t} \left[ M_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right] + p_{22} E_{2,t} \left[ M_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right]$$

for $z_t = 2$. The Euler equation can be rewritten as

$$0 = \tilde{p}_{11} E_{1,t} \left[ M^{EZ}_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right] + (1 - \tilde{p}_{11}) E_{2,t} \left[ M^{EZ}_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right]

0 = (1 - \tilde{p}_{22}) E_{1,t} \left[ M^{EZ}_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right] + \tilde{p}_{22} E_{2,t} \left[ M^{EZ}_{z_{t+1},t+1} (R_{e,t+1} - R_{f,t}) \right]$$

where $M^{EZ}_{z, t+1}$ is the SDF under recursive utility without ambiguity aversion, and $\tilde{p}_{11}$ and $\tilde{p}_{22}$ are distorted transition probabilities and are given by

$$\tilde{p}_{11} = \frac{p_{11}}{p_{11} + (1 - p_{11}) \left( \frac{E_{2} \left[ V_{z_{t+1},t+1}^{1-\gamma} \right]}{E_{1} \left[ V_{z_{t+1},t+1}^{1-\gamma} \right]} \right)^{\frac{-\beta}{1-\gamma}}}$$

$$\tilde{p}_{22} = \frac{p_{22}}{(1 - p_{22}) \left( \frac{E_{1} \left[ V_{z_{t+1},t+1}^{1-\gamma} \right]}{E_{2} \left[ V_{z_{t+1},t+1}^{1-\gamma} \right]} \right)^{\frac{-\beta}{1-\gamma}}} + p_{22}$$

where $V_{z,t}, (z_t = 1, 2)$ are solutions to the following value function under full information:

$$V_{z,t}(C) = \left[ (1 - \beta) C_t^{1-\frac{1}{\psi}} + \beta \left\{ R_{z,t} \left( V_{z_{t+1},t+1}(C) \right) \right\}^{1-\frac{1}{\psi}} \right]^{1-\frac{1}{\psi}}$$

$$R_{z,t} \left( V_{z_{t+1},t+1}(C) \right) = \left( E_{z_t} \left[ \left( E_{z_{t+1},t} \left[ V_{z_{t+1},t+1}^{1-\gamma} \right] \right)^{\frac{1-\beta}{\gamma}} \right] \right)^{\frac{1}{1-\gamma}}$$

• The numerical algorithm of calculating detection-error probabilities takes the following steps:

1. Repeatedly draw $\{\Delta c_t\}_{t=1}^T$ under the reference data generating process (DGP), which is the two-state Markov switching model with transition probabilities $p_{11}$ and $p_{22}$.

2. Evaluate the log likelihood function under the reference DGP by computing

$$\ln L_T^r = \sum_{t=1}^{T} \ln \left\{ \sum_{z_{t+1}=1}^{2} f(\Delta c_{t}|z_{t}) \Pr(z_{t}|\Omega_{t-1}) \right\}$$

where $\pi_{t-1} = \Pr(z_{t} = 1|\Omega_{t-1})$ are filtered probabilities implied by the Markov switching model.
3. Evaluate the log likelihood function under the distorted DGP by computing

\[
\ln L^d_T = \sum_{t=1}^{T} \ln \left\{ \sum_{z_t=1}^{2} f(\Delta c_t | z_t) \Pr(z_t | \Omega_{t-1}) \right\}
\]

where \( \Pr(z_t | \Omega_{t-1}) \) are the filtered probabilities that are obtained by applying the distorted transition probabilities \( \tilde{p}_{11,t} \) and \( \tilde{p}_{22,t} \) (in place of the constant transition probabilities \( p_{11} \) and \( p_{22} \)) to the Markov switching model’s filter.

4. Compute the fraction of simulations for which \( \ln \left( \frac{L^d_T}{L^r_T} \right) > 0 \) and denote it as \( p_r \). The fraction approximates the probability that the econometrician believes that the distorted DGP generated the data, while the data are actually generated by the reference DGP.

5. Do a symmetrical computation and simulate \( \{\Delta c_t\}_{t=1}^{T} \) under the distorted DGP. Compute the fraction of simulations for which \( \ln \left( \frac{L^r_T}{L^d_T} \right) > 0 \) and denote it as \( p_d \). This fraction approximates the probability that the reference DGP generated the data when actually the distorted DGP generates the data.

Assuming an equal prior on the reference and the distorted DGP, the detection error probability is defined by (see Anderson et al. (2003)):

\[
p(\eta) = \frac{1}{2} (p_r + p_d) . \tag{22}
\]

In the approximation, we set \( T = 100 \) years and simulate 20,000 samples of artificial data.
This table reports summary statistics for annual (1929-2013) and quarterly (1947:Q2-2014:Q2) U.S. data. 1-year Treasury Bill rate ($r_t^e$), aggregate equity returns ($r_t^I$), excess returns ($r_t^e - r_t^f$), and real, per capita, log consumption growth ($\Delta c_t$) are expressed in percentages. Mean and standard deviation of quarterly data are annualized. The row titled “J-B test” reports the $p$-values of Jarque and Bera (1980) test of normality.

<table>
<thead>
<tr>
<th></th>
<th>$r_t^e$</th>
<th>$r_t^I$</th>
<th>$r_t^e - r_t^f$</th>
<th>$\Delta c_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1929-2013</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>8.54</td>
<td>1.07</td>
<td>7.47</td>
<td>1.85</td>
</tr>
<tr>
<td>St Dev</td>
<td>20.35</td>
<td>0.06</td>
<td>20.35</td>
<td>2.15</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.29</td>
<td>0.60</td>
<td>-0.29</td>
<td>-1.49</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.72</td>
<td>1.32</td>
<td>-0.72</td>
<td>5.01</td>
</tr>
<tr>
<td>J-B test</td>
<td>0.3938</td>
<td>0.0133</td>
<td>0.4012</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>1947:Q2-2014:Q2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>8.76</td>
<td>1.05</td>
<td>7.71</td>
<td>1.91</td>
</tr>
<tr>
<td>St Dev</td>
<td>16.43</td>
<td>0.02</td>
<td>16.44</td>
<td>1.02</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.57</td>
<td>0.99</td>
<td>-0.57</td>
<td>-0.42</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.90</td>
<td>1.27</td>
<td>1.90</td>
<td>1.11</td>
</tr>
<tr>
<td>J-B test</td>
<td>0.0013</td>
<td>0.0001</td>
<td>0.0013</td>
<td>0.0017</td>
</tr>
</tbody>
</table>
Table 2: GSM Estimation Results for the Benchmark and Alternative Models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mode</td>
<td>Mean</td>
<td>St Dev</td>
<td>Mode</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9766</td>
<td>0.6267</td>
<td>0.2340</td>
<td>0.9428</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.8750</td>
<td>1.9808</td>
<td>0.8972</td>
<td>1.6172</td>
</tr>
<tr>
<td>$\psi$</td>
<td>1.3750</td>
<td>1.7765</td>
<td>0.5147</td>
<td>2.6172</td>
</tr>
<tr>
<td>$\eta$</td>
<td>10.0000</td>
<td>9.3388</td>
<td>3.6839</td>
<td>29.7969</td>
</tr>
<tr>
<td>$p_{1,1}$</td>
<td>0.9805</td>
<td>0.9682</td>
<td>0.0179</td>
<td>0.9399</td>
</tr>
<tr>
<td>$p_{2,2}$</td>
<td>0.4688</td>
<td>0.5133</td>
<td>0.0036</td>
<td>0.0201</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>0.0217</td>
<td>0.0225</td>
<td>0.0036</td>
<td>0.0201</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>-0.0615</td>
<td>-0.0662</td>
<td>0.0199</td>
<td>-0.0683</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>2.8750</td>
<td>2.5191</td>
<td>0.8536</td>
<td>1.0781</td>
</tr>
<tr>
<td>$\sigma_{\Delta c}$</td>
<td>0.0308</td>
<td>0.0310</td>
<td>0.0028</td>
<td>0.0267</td>
</tr>
<tr>
<td>Log Lik Stat Mod</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>-392.2910</td>
</tr>
<tr>
<td>Log Lik Sci Mod</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1929-2013</td>
</tr>
</tbody>
</table>

This table reports priors and posteriors on mode, mean, and standard deviation of preference, dividend growth, and consumption growth parameters for the benchmark model featuring ambiguity aversion and the alternative model with ambiguity neutral agents. The difference between the benchmark model and the alternative model is that we impose $\eta = \gamma$ restriction in the alternative model. Preference parameters ($\beta, \gamma, \psi, \eta$) represent subjective discount factor, coefficients of risk aversion, intertemporal elasticity of substitution, and ambiguity aversion respectively. $p_{1,1}$ and $p_{2,2}$ are transition probabilities from good-to-good and bad-to-bad states, respectively. $\kappa_1$ and $\kappa_2$ are good and bad state mean consumption growth parameters, respectively. $\lambda$ is the leverage ratio in the model, and $\sigma_{\Delta c}$ and $\sigma_{\Delta d}$ are volatility parameters for consumption and dividend growth, respectively. We use 1929-2013 annual real-valued data to obtain annual estimates reported in Panel A. 1929-1949 data are used for priming the estimation process, and 1950-2013 data yield the estimated parameters. We use 1947Q2-2014Q2 real-valued data to obtain quarterly estimates reported in Panel B. We prime the estimation process using 1947Q2-1955Q2. 1955Q3-2014Q2 data yield the estimated parameters. We report log likelihood and log posterior values (when relevant).
Table 3: Asset Pricing Implications

<table>
<thead>
<tr>
<th></th>
<th>Annual Data</th>
<th>Quarterly Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bench Mark Model</td>
<td>Alternative Model</td>
</tr>
<tr>
<td>$\mu_{1}$</td>
<td>0.98</td>
<td>1.44</td>
</tr>
<tr>
<td>$\sigma_{1}$</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>$\mu_{2} - \mu_{1}$</td>
<td>7.31</td>
<td>1.36</td>
</tr>
<tr>
<td>$\sigma_{2} - \sigma_{1}$</td>
<td>17.41</td>
<td>17.66</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>0.42</td>
<td>0.08</td>
</tr>
<tr>
<td>$\sigma(M_t)/\mu(M_t)$</td>
<td>2.63</td>
<td>0.28</td>
</tr>
<tr>
<td>$\Upsilon(VRP)$</td>
<td>1.21</td>
<td>0.35</td>
</tr>
<tr>
<td>$\sigma(VRP)$</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>Skewal(VRP)</td>
<td>2.40</td>
<td>2.41</td>
</tr>
<tr>
<td>Kurtal(VRP)</td>
<td>4.61</td>
<td>5.68</td>
</tr>
</tbody>
</table>

We present the asset pricing quantities implied by calibrating the model in Ju and Miao (2012) and the alternative model featuring ambiguity neutrality using the estimated parameters presented on Table 2. In addition, we present the original quantities reported by Ju and Miao (2012) for 1871–1993 sampling period. Unconditional means and standard deviations, $\mu(x_t)$ and $\sigma(x_t)$, are in percents.
The figure shows, from top to bottom, annual returns of CRSP-Compustat value weighted index returns, 1-year Treasury Bill rates, excess returns over 1-year T-Bill rate, and annual real per capita log consumption growth for 1929-2013 period.
Figure 2: Prior and Posterior Densities of Estimated Parameters of the Benchmark Model, Annual Data

This figure plots prior and posterior densities of the benchmark model featuring ambiguity aversion. The solid lines depict posterior densities and dotted lines do the same for prior densities. The parameters are: “bet” is discount factor $\beta$, “gma” is coefficient of risk aversion $\gamma$, “psi” is the inverse of the elasticity of intertemporal substitution parameter $\psi$, “eta” is the ambiguity aversion parameter, “P” is the good-to-good state transition probability $p_{1,1}$, “Q” is the bad-to-bad state transition probability $p_{2,2}$, “KaC(1)” is the bad state mean real per capita consumption growth rate $\kappa_2$, “KaC(2)” is the good state mean real per capita consumption growth rate $\kappa_1$, “lam” is the leverage ratio $\lambda$, “SigC” stands for the volatility of real per capita consumption growth $\sigma_{\Delta c}$, and finally “ESigD” represents the volatility of the real per capita dividend growth $\sigma_{\Delta d}$. The plotted values are based on 1929–2013 data.
This figure depicts prior and posterior densities of estimated parameters in the alternative model featuring ambiguity neutrality. This model is nested by the benchmark model. The difference between the two models is the imposition of $\eta = \gamma$ constraint in the alternative model. As a result, $\eta$ is not estimated in this model. Otherwise, the description of plotted densities is the same as in Figure 2. The plotted values are based on 1929–2013 data.
Figure 4: Prior and Posterior Densities of Estimated Parameters of the Benchmark Model, Quarterly Data

Refer to Figure 2 for a description and the model and parameters depicted. Solid lines represent posterior densities and the dotted lines are prior densities. The plotted values are based on 1947–2014 quarterly data.
Figure 5: Prior and Posterior Densities of Estimated Parameters of the Alternative Model, Quarterly Data

Refer to Figure 3 for a description and the model and parameters depicted. Solid lines represent posterior densities and the dotted lines are prior densities. The plotted values are based on 1947–2014 quarterly data.
The figures show posterior forecasts for consumption growth, equity returns, and the short rate for both our benchmark model featuring ambiguity aversion and the alternative model with ambiguity neutral agents. The left hand panel contains forecasts for the benchmark model and the right hand panel does the same for the alternative model. The dashed lines are the ±1.96 posterior standard deviations.
Figure 6: Quantitative Implications of Ambiguity Aversion

Panel A: Filtered probabilities and distorted probabilities

Panel B: $E_t(\Delta c)$

Panel C: Conditional equity premium

Panel D: Variance risk premium

The figure plots Bayesian beliefs, defined in Equation (3), and ambiguity-distorted beliefs, defined in Equation (8), expected real consumption growth, conditional equity premium, and the variance risk premium paths, generated based on quarterly estimated parameter values. Shaded areas represent NBER recessions.
Figure 7: Conditional moments: Annual estimates

Panel A: Equity premium

Panel B: Equity volatility

Panel C: Price of risk

Panel D: P/D

Conditional moments plotted against the belief of the good state. Same below.
Figure 8: Conditional moments: Quarterly estimates

Panel A: Equity premium

Panel B: Equity volatility

Panel C: Price of risk

Panel D: P/D