

# A Statistical Inquiry into the Plausibility of Recursive Utility\*

A. Ronald Gallant	Han Hong
Fuqua School of Business	Department of Economics
Duke University	Duke University
Durham NC 27708-0120 USA	Durham NC 27708-0097 USA

First draft: April 2006  
This draft: July 2006  
Corrections: October 2006

## Abstract

We use purely statistical methods to determine if it is plausible that the pricing kernel can be represented as the intertemporal marginal rate of substitution of a representative agent in an endowment economy whose preferences are determined by recursive utility. Specifically we investigate regression implications of recursive utility such as the implication that the R-square of a log-linear regression of the pricing kernel on endowment growth and the return to wealth is one. We make no assumptions other than that a unique pricing kernel exists. We introduce a Bayesian statistical method that treats the pricing kernel as a latent variable and extracts it and its transition density from real returns on twenty-five Fama-French portfolios and bonds. Our priors are formed from an examination of a long simulation of a Bansal-Yaron economy. Using both monthly data and annual data, we determine the posterior distributions of the R-square of various regressions and compare these posteriors to the R-square that should have obtained were the pricing kernel determined by recursive utility. These regressions are invariant to seasonal adjustment. We find that the data support recursive utility. Incidentally we find that the data do not support external habit utility.

Keywords and Phrases: Recursive Utility, Asset Pricing, Statistical Methods.  
JEL Classification: E00, G12, C51, C52

---

\*Supported by the National Science Foundation.

# 1 Introduction

There has been a revival of interest in representative agent consumption based asset pricing. Much of this literature endows the representative agent with recursive utility (Kreps and Porteus, 1978; Epstein and Zin, 1989; Weil, 1990). Some recent examples of this strand of the literature are Bansal and Yaron (2004), Hansen, Heaton, and Li (2005), Kiku (2006), and Eraker (2006). Because so much of the new literature relies on recursive utility, it is natural to ask if recursive utility is plausible.

Trying to answer this question by implementing one or more of the models that have been proposed by contributors to this literature and fitting them statistically to consumption data and asset returns runs into several difficulties. The first is that macro economic data are sparse. For U.S. data all that are available are about 75 annual observations, 135 quarterly observations, and 550 monthly observations. The obvious way to combat sparse data is to use prior information and Bayesian statistical methods. But Bayesian methods bring into play two additional difficulties: The model must be solved and a likelihood must be synthesized. As in Bansal and Yaron (2004), Hansen, Heaton, and Li (2005), and Eraker (2006), most implementations rely on approximate solutions derived by linearization. Others rely on numerical methods (Kiku, 2006). In either event, the quality of these approximations is unknown, especially when considering that Bayesian MCMC methods require accurate solutions over a wide range of parameter values. To synthesize a likelihood, one can assume a VAR is an adequate approximation (Del Negro and Schorfheide, 2004), use a flexible functional form (Gallant and McCulloch, 2005), or use a particle filter (Fernandez-Villaverde and Rubio-Ramirez, 2004). The quality of the synthesis is unknown. In addition to errors introduced by solving the model and synthesizing a likelihood, there are intrinsic measurement errors in consumption data and distortions induced by seasonal adjustments to consumption data. The upshot is that a statistical inquiry along these lines carries with it the augmenting assumption of an assumed law of motion for consumption and perhaps other driving processes such as cash flows, error from solving the model, error from synthesizing a likelihood, and error from poor quality data. Answering the question “Is recursive utility plausible?” in the face of all this excess baggage would seem to be impossible.

We attack the problem from a different direction. We use Bayesian methods that are not new but are little known to directly extract the law of motion of the pricing kernel and the pricing kernel itself. We use both monthly and annual panel data comprised of returns on the twenty-five Fama-French (1993) portfolios and U.S. Treasury obligations. This extraction process does not rely on consumption data, assumed driving processes, a solution method, or a synthesized likelihood. All it assumes is that a pricing kernel  $\theta = (\theta_1, \dots, \theta_{n+1})$  exists. Given existence, a nonparametric likelihood that depends only on returns is derived directly from Euler equations. Neither a model nor driving processes are required to derive this likelihood.

An overview of the logical structure of our approach is as follows. The likelihood is hierarchical. The first component is denoted  $\mathcal{L}(\theta)$ , is fully nonparametric, depends only on returns, and is derived directly from Euler equations. The second component provides the law of motion of the pricing kernel given parameters  $\eta$  and is denoted  $f(\theta_t|\theta_{t-1}, \dots, \theta_1, \eta)$  or  $f(\cdot|\cdot, \eta)$  for short. The second component  $f(\cdot|\cdot, \eta)$  is seminonparametrically specified. The likelihood itself is the product  $\mathcal{L}(\theta, \eta) = \mathcal{L}(\theta) \prod_{t=1}^n f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$ . The prior is composite. The first part  $p(\eta)$  is substantive, important, and has three versions: tight, intermediate, and loose. The second part  $p_T(\theta, \eta)$  imposes technical conditions. The prior itself is the product  $p(\theta, \eta) = p(\eta)p_T(\theta, \eta)$ . Inference is conducted by using the posterior distribution  $P(\theta, \eta)$  to compute the posterior probabilities  $P(A) = \int \int I_A(\theta, \eta)P(d\theta, d\eta)$  of events  $A$  that represent hypotheses.

We derive our prior distribution by examining a Bansal-Yaron economy. Upon relaxing this prior, i.e. moving from the tight to the loose prior, our posterior probabilities indicate that Bansal-Yaron consumption dynamics are mildly misspecified: The law of motion of the marginal rate of intertemporal substitution, which is the pricing kernel in a Bansal-Yaron economy, has less predictability than the law of motion that we extract from the data. The misspecification seems to be mild because one of our tests of recursive utility relies on an augmenting hypothesis of Bansal-Yaron consumption dynamics yet it accepts. This same test does appear to have reasonable power because it rejects external habit. In contrast, some statistical tests cannot distinguish between a Bansal-Yaron economy and an external habit economy, e.g., the test of overidentifying restrictions used by Bansal, Gallant,

and Tauchen (2004). The cash flow process in Bansal-Yaron seems to be more severely misspecified because our test that relies on this augmenting hypothesis rejects. Our test that relies on neither augmenting hypothesis accepts.

The plan of the paper is as follows. In Section 2 we describe recursive utility, derive the law of motion  $f(\cdot|\cdot, \eta)$ , derive the substantive prior  $p(\eta)$ , and discuss the rationale for our tests. In Section 3 we describe our data. In Section 4 we provide the rationale for the first part of our likelihood  $\mathcal{L}(\theta)$ . In Section 5 we describe the technical portion of our prior  $p_T(\theta, \eta)$ . In Section 6 we present our empirical findings. We conclude in Section 7.

## 2 Recursive Utility

Let  $C_t$  denote the monthly consumption endowment. Let  $P_{ct}$  denote the price of an asset that pays the consumption endowment. Let  $R_{ct} = (P_{ct} + C_t)/P_{c,t-1}$  denote the gross return on the consumption endowment. Similarly, if an asset  $S$  pays  $D_{st}$  per month, then it has price  $P_{st}$  and gross return  $R_{st} = (P_{st} + D_{st})/P_{s,t-1}$ . Prices are real.

The recursive utility function is defined as

$$U_t = \left[ (1 - \delta) C_t^{(1-1/\psi)} + \delta (\mathcal{E}_t U_{t+1}^{1-\gamma})^{\frac{1-1/\psi}{1-\gamma}} \right]^{\frac{1}{1-1/\psi}},$$

where  $\delta$  is the time preference parameter,  $\gamma$  is the coefficient of risk aversion, and  $\psi$  is the elasticity of intertemporal substitution. Under the budget constraint  $W_{t+1} = (W_t - C_t)R_{c,t+1}$  where  $W_t$  is the representative agent's wealth, the intertemporal marginal rate of substitution of the agent implied by recursive utility is

$$M_{t,t+1} = \delta^\beta \left( \frac{C_{t+1}}{C_t} \right)^{-(\beta/\psi)} (R_{c,t+1})^{(\beta-1)},$$

where

$$\beta = \frac{1 - \gamma}{1 - 1/\psi}.$$

The gross return of the asset that pays the consumption endowment satisfies the Euler equation

$$1 = \mathcal{E}_t (M_{t,t+1} R_{c,t+1}) \tag{1}$$

and the gross return on an asset that pays  $D_{st}$  satisfies

$$1 = \mathcal{E}_t (M_{t,t+1} R_{s,t+1}). \tag{2}$$

Any pricing kernel  $\{\theta_t\}$  will satisfy these Euler equations; e.g.

$$1 = \mathcal{E}_t(\theta_{t+1}R_{s,t+1}). \quad (3)$$

Our goal is to use Bayesian methods to infer the posterior distribution of  $\{\theta_t\}$  using (3) and a panel of asset returns without assuming that  $\{\theta_t\}$  is necessarily of the form  $\theta_t = M_{t-1,t}$  implied by recursive utility. Our approach uses a hierarchical likelihood of the form

$$\mathcal{L}(\theta, \eta) = \mathcal{L}(\theta) \prod_{t=1}^n f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$$

where  $f(\cdot|\cdot, \eta)$  is a law of motion for  $\{\theta_t\}$ . Consequently we need a specification of  $f(\cdot|\cdot, \eta)$  and a prior  $p(\eta)$  for the hyperparameter  $\eta$ . To this end we provisionally adopt recursive utility and proceed as follows.

From the literature we accept plausible values for the recursive utility function parameters and a plausible specification of the joint distribution of the consumption endowment and dividend on the market portfolio. Once  $(\delta, \gamma, \psi)$  and the distribution of  $(C_t, D_{st})$  are specified, the distribution of  $M_{t,t+1}$  can be determined from a long simulation of the economy. We accept Kiku's (2006) specification of the distribution of  $C_t$  and  $D_{st}$ , which is

$$\begin{aligned} \log C_{t+1} &= \log C_t + \mu_c + x_t + \sigma_t \eta_{t+1} \\ \log D_{t+1} &= \mu + \phi x_t + \kappa \sigma_t u_{t+1} \\ x_{t+1} &= \rho x_t + \kappa_x \sigma_t \epsilon_{t+1} \\ \sigma_{t+1}^2 &= \sigma^2(1 - \nu) + \nu \sigma_t^2 + \sigma_w w_{t+1}. \end{aligned}$$

Here the errors are normally distributed with mean zero and unit variance and are independent both temporally and cross-sectionally with the exception that  $\text{Corr}(\eta_t, u_t) = \alpha \neq 0$ . Parameters were calibrated in Kiku (2006) by simulating at the monthly frequency, aggregating to the annual frequency, and matching moments computed from annual market returns and annual consumption data. Her parameter values are  $(\delta, \gamma, \psi) = (0.999, 10, 1.5)$ ,

$$(\mu_c, \rho, \kappa_x, \sigma, \nu, \sigma_w) = (0.0015, 0.98, 0.032, 0.0064, 0.99, 0.0000017),$$

and  $(\mu, \phi, \kappa, \alpha) = (0.0012, 2.8, 7.5, 0.55)$ . This specification is similar to Bansal and Yaron(2004) and Bansal, Gallant, and Tauchen(2004).

From a simulation of this calibrated economy kindly provided to us by Dana Kiku, we fitted the SNP expansion (Gallant and Tauchen, 1989) by maximum likelihood using a public domain program (<http://econ.duke.edu/webfiles/arg/snp>) with the number of terms in the expansion determined by the Schwarz (1978) BIC criterion. We interpret this fit as a Bayesian nonparametric estimate of the transition density obtained with a flat prior. The form of the fitted density of  $\theta_t = M_{t-1,t}$  is

$$f(\theta_t|\theta_{t-1}, \dots, \theta_1, \eta) = f(y_t|y_{t-1}, \dots, y_1, \eta) / \exp(y_t) \quad (4)$$

$$y_t = \log(\theta_t) \quad (5)$$

$$f(y_t|y_{t-1}, \dots, y_1, \eta) = \mathcal{P}^2(y_t|a) n(y_t | \mu_{t-1}, \sigma_{t-1}^2) \quad (6)$$

$$\eta = (a_1, a_2, a_3, a_4, b_0, b_1, r_0, r_1, r_2) \quad (7)$$

$$\mathcal{P}(y_t|a) = 1 + a_1 y_t + a_2 y_t^2 + a_3 y_t^3 + a_4 y_t^4 \quad (8)$$

$$\mu_{t-t} = b_0 + b_1 y_{t-1} \quad (9)$$

$$\sigma_{t-1}^2 = r_0^2 + r_1^2 \sigma_{t-2}^2 + r_2^2 (y_{t-1} - \mu_{t-2})^2 \quad (10)$$

where  $n(\cdot|\mu, \sigma^2)$  denotes the normal density function.

The parameter estimates and their standard errors are shown under the headings location and scale, respectively, in Table 1 for fits of (4) to simulations at both the monthly and annual frequency. The monthly estimates are determined from a simulation of length 5000. We interpret the monthly parameter estimates and their estimated standard errors as the prior opinion that one would form about the location and scale of the parameters of (4) after having observed this economy for  $5000/12=417$  years. The annual estimates are determined from a simulation of length 1600 which we interpret as the prior opinion one would form after 1600 years of observation. Operationally it turns out that if we take the scaling an order of magnitude smaller than shown in Table 1, the prior binds, and if we take it an order of magnitude larger, it does not. For this reason we call a prior in Table 1 with scale divided by ten tight, one with scale multiplied by ten loose, and one with scale as shown in Table 1 intermediate.

Table 1 about here

Lastly we note that if the pricing kernel that we extract from data does obtain from

recursive utility then it must satisfy the equation

$$\log(M_{t,t+1}) = c_1 + c_2 \log\left(\frac{C_{t+1}}{C_t}\right) + c_3 \log(R_{c,t+1}) \quad (11)$$

where  $\beta = c_3 + 1$ ,  $\psi = -\beta/c_2$ , and  $\delta = \exp(c_1/\beta)$ . Moreover, regardless of whether the pricing kernel is or is not recursive, if we can obtain a distribution for the process  $\frac{C_{t+1}}{C_t} M_{t,t+1}$ , then we can compute the price of consumption using

$$P_{ct} = \mathcal{E}_t \sum_{j=1}^{\infty} C_{t+j} M_{t,t+j} = C_0 \left( \prod_{k=1}^t \frac{C_k}{C_{k-1}} \right) \sum_{j=1}^{\infty} \mathcal{E}_t \prod_{k=1}^j \left( \frac{C_{t+k}}{C_{t+k-1}} M_{t-1+k,t+k} \right) \quad (12)$$

and compute the gross return on consumption using

$$R_{ct} = \frac{P_{ct} + C_t}{P_{c,t-1}} = \frac{\frac{C_{t-1}}{C_{t-2}} \sum_{j=1}^{\infty} \mathcal{E}_t \prod_{k=1}^j \left( \frac{C_{t+k}}{C_{t+k-1}} M_{t+k-1,t+k} \right) + \frac{C_t}{C_{t-1}} \frac{C_{t-1}}{C_{t-2}}}{\sum_{j=1}^{\infty} \mathcal{E}_{t-1} \prod_{k=1}^j \left( \frac{C_{t+k-1}}{C_{t+k-2}} M_{t+k-2,t+k-1} \right)}. \quad (13)$$

### 3 Data

We use two data sets. The first is 551 monthly observations from February 1959 through December 2004 on real returns including dividends on twenty-four of the twenty-five Fama-French (1993) portfolios, real returns on U.S. Treasury debt of ten year, one year, and thirty day maturities, real returns including dividends on the aggregate stock market, real, per-capita, consumption expenditure growth, and real, per-capita, labor income growth. The second is 75 annual observations from 1930 through 2004 on the same variables except U.S. Treasury obligations of ten and one year maturities. The exclusion of one of the twenty-five Fama-French portfolios was because there were missing values during the period of observation. Ten and one year Treasuries were excluded from the annual data for the same reason.

The raw monthly and annual Fama-French portfolios were obtained from Kenneth French's web site (<http://mba.tuck.dartmouth.edu/pages/faculty/ken.french>). The portfolios are the intersections of five portfolios formed on market equity and five portfolios formed on the ratio of book equity to market equity. The portfolios are for all NYSE, AMEX, and NASDAQ stocks for which equity data are not missing and book equity data are positive. The portfolios are constructed at the end of each June with breakpoints determined by the NYSE quintiles at the end of June. Complete details are at Kenneth French's web site. The

advantage of the Fama-French portfolios here is that they appear to isolate and exhaust the risk factors for holding equities (Fama and French, 1992, 1993).

The raw monthly and annual data for returns on U.S. Treasury debt are from the Center for Research in Security Prices data at the Wharton Research Data Services web site (<http://wrds.wharton.upenn.edu>).

The raw monthly consumption data are seasonally adjusted at annual rates personal consumption expenditures on nondurables and services obtained from the Bureau of Economic Analysis web site (<http://www.bea.gov>). The annual data are annual expenditures on nondurables and services from the same source.

The raw monthly labor income data are seasonally adjusted at annual rates and are the series entitled compensation of employees received at the Bureau of Economic Analysis web site. The annual data are the annual series from the same source.

Raw data are converted from nominal to real using the monthly and annual consumer price indexes obtained from the Bureau of Economic Analysis web site. Conversion to per capita is by means of the mid-year population data from the same source. Monthly population values are obtained from the annual by linear interpolation.

Simple statistics for these data are shown in Tables 2 and 3.

Table 2 about here

Table 3 about here

## 4 Estimation Methodology

In this section we provide the rationale behind our estimation strategy and the details of its implementation.

### 4.1 Rationale

The ideas behind the estimation methodology are apparently part of the folklore of Bayesian inference. We first learned of the method from James Berger, Duke University. We have seen it used by Duan and Mela (2006) who cite Romeo (2004). We do not know where a rationale

for the method proceeding from the notion of sufficiency (Fisher, 1925) can be found and so provide it here.

Consider estimating the mean  $\mu$  and variance  $\Sigma$  from a sample  $\{y_t\}_{t=1}^n$  of independent multivariate normal observations. The likelihood factors into the product of the Wishart distribution and the normal distribution as follows

$$\mathcal{L}(\mu, \Sigma) \propto |\Sigma|^{-\frac{1}{2}} |W|^{\frac{1}{2}(n-p-2)} \exp \left\{ -\frac{n}{2} \text{tr} [(\bar{y} - \mu)(\bar{y} - \mu)' + W] \Sigma^{-1} \right\}, \quad (14)$$

where

$$\begin{aligned} \bar{y} &= \frac{1}{n} \sum_{t=1}^n y_t \\ W &= \frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})(y_t - \bar{y})' \end{aligned}$$

are sufficient statistics by the factorization theorem (Fisher, 1925). The estimator that maximizes (14) is  $(\hat{\mu}_n, \hat{\Sigma}_n) = (\bar{y}, W)$ .

As is well known, one can use  $(\bar{y}, W)$  as data and the right hand side of (14) as the likelihood without affecting the conclusions of Bayesian inference. However, there is one subtlety that is usually irrelevant but is important later in this subsection and in Subsection 4.3. It is that information is lost because not every set  $A$  expressed in terms of the original data  $Y = (y_1, \dots, y_n)$  can be assigned probability. Probability can only be assigned to  $A$  if it is the preimage of a set that can be expressed in terms of  $(\bar{y}, W)$ . Specifically, if  $T$  denotes the map  $Y \mapsto (\bar{y}, W)$ , then  $A$  must satisfy  $A = T^{-1}(B)$  for some Borel set  $B$ . The probability assigned to  $A$  is the integral of the right hand side of (14) over  $B$ . This is just an application of the change of variables formula of integration. Although it is rarely necessary to deduce the form of  $A$ , it is easy to do in this instance: If  $\sigma = (\sigma_1, \dots, \sigma_n)$  denotes a permutation of the indices  $t = 1, \dots, n$ , and  $\mathcal{S}$  denotes the set of all such permutations, and  $Y_\sigma = (y_{\sigma_1}, \dots, y_{\sigma_n})$ , then sets  $A$  that satisfy  $A = T^{-1}(B)$  have the property that  $Y \in A \Rightarrow Y_\sigma \in A$  for all  $\sigma \in \mathcal{S}$ .

The estimator  $(\hat{\mu}_n, \hat{\Sigma}_n)$  can also be expressed in terms of the estimating equations

$$\Psi_n(\hat{\mu}_n, \hat{\Sigma}_n) = \begin{pmatrix} \bar{y} - \hat{\mu}_n \\ W - \hat{\Sigma}_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (15)$$

Evaluating the estimating equations at the true value  $(\mu^o, \Sigma^o)$ , we have

$$\Psi_n(\mu^o, \Sigma^o) = \begin{pmatrix} \bar{y} - \mu^o \\ W - \Sigma^o \end{pmatrix} = \begin{pmatrix} Z_{1n} \\ Z_{2n} \end{pmatrix}. \quad (16)$$

Here  $Z_{1n}$  and  $Z_{2n}$  are independent,  $Z_{1n}$  is  $N(0, \Sigma/n)$ , and  $Z_{2n}$  is a location shifted Wishart.

One can proceed in the opposite direction by viewing estimating equations such as (15) as the first order conditions of a maximum likelihood estimator and infer a likelihood by specifying a distribution for  $Z_{1n}$  and  $Z_{2n}$  given by (15). In this instance, if we specify that  $Z_{1n}$  and  $Z_{2n}$  are independent,  $Z_{1n}$  is normal, and  $Z_{2n}$  location shifted Wishart we would infer the likelihood  $\mathcal{L}(\mu, \Sigma)$  above. In general, either the direct approach of writing down the likelihood of the sample or the indirect approach of writing down a set of estimating equations and inferring a likelihood from the estimating equations will lead to a likelihood function  $\mathcal{L}(\mu, \Sigma)$  that can be used as the basis of statistical inference. When, as here, the statistics in the estimating equations are first and second moments and the data are normally distributed, the outcome is the same.

However, describing an estimator in terms of estimating equations based on first and second sample moments and inferring the likelihood  $\mathcal{L}(\mu, \Sigma)$  is actually a more general approach because all that is required to justify it that  $n$  is large enough for the central limit theorem to be a good approximation to the sampling distribution of the first moment and an assumption that the skewness of  $y_t$  is zero.

Carrying the logic further and restricting attention to  $\mu$ , one can formulate the estimating equations as

$$\Psi_n(\hat{\mu}_n) = W^{-1/2} \sqrt{n}(\bar{y}_n - \hat{\mu}_n) = 0$$

and assume that

$$\Psi_n(\mu^o) = W^{-1/2} \sqrt{n}(\bar{y}_n - \mu^o) = Z_n$$

is normal to infer the likelihood

$$\mathcal{L}(\mu) \propto \exp \left\{ -\frac{n}{2} (\bar{y} - \mu)' W^{-1} (\bar{y} - \mu) \right\}.$$

This eliminates the skewness assumption.

Now consider estimating equations for a potentially nonlinear situation

$$\Psi_n(\hat{\theta}_n) = [W_n(\hat{\theta}_n)]^{-1/2} \sqrt{n} \bar{m}_n(\hat{\theta}_n) = 0 \quad (17)$$

$$\bar{m}_n(\theta) = \frac{1}{n} \sum_{t=1}^n m_t(y_t, x_t, \theta) \quad (18)$$

where  $x_t$  denotes exogenous and/or predetermined variables. The parameter  $\theta$  is used advisedly here because later it will be the ex-post pricing kernel. If the  $m_t(y_t, x_t, \theta^o)$  are an uncorrelated mixing process, then

$$W_n(\theta) = \frac{1}{n} \sum_{i=1}^n [m_i(y_i, x_i, \theta) - \bar{m}_n(\theta)] [m_i(y_i, x_i, \theta) - \bar{m}_n(\theta)]'$$

If the  $m_t(y_t, x_t, \theta^o)$  are a correlated mixing process, then  $W_n(\theta)$  is a HAC estimator such as

$$W_n(\theta) = \sum_{\tau=-[n^{1/5}]}^{[n^{1/5}]} w\left(\frac{\tau}{[n^{1/5}]}\right) W_{n\tau}(\theta)$$

where

$$w(u) = \begin{cases} 1 - 6|u|^2 + 6|u|^3 & \text{if } 0 < u < \frac{1}{2} \\ 2(1 - |u|)^3 & \text{if } \frac{1}{2} \leq u < 1 \end{cases}$$

$$W_{n\tau}(\theta) = \begin{cases} \frac{1}{n} \sum_{t=1+\tau}^n [m(y_t, x_t, \theta) - \bar{m}_n(\theta)] [m(y_{t-\tau}, x_{t-\tau}, \theta) - \bar{m}_n(\theta)]' & \tau \geq 0 \\ W'_{n, -\tau}(\theta) & \tau < 0 \end{cases}$$

(Gallant, 1987). Putting

$$\Psi_n(\theta^o) = [W_n(\theta^o)]^{-1/2} \sqrt{n} \bar{m}_n(\theta^o) = Z_n$$

we have, as above, that a likelihood is

$$\mathcal{L}(\theta) \propto \exp\left\{-\frac{n}{2} \bar{m}_n(\theta)' [W_n(\theta)]^{-1} \bar{m}_n(\theta)\right\}. \quad (19)$$

Although (19) is the formula used in computing, it is a little misleading. One should view the formula as follows: One summarizes the data  $(Y, X)$  as  $t = T_\theta(Y, X)$  using the map  $T_\theta : (Y, X) \mapsto \sqrt{n} [W_n(\theta)]^{-1/2} \bar{m}_n(\theta)$ .  $\mathcal{L}(\theta)$  is to be viewed as a density in  $t$  over  $\mathfrak{R}^K$ . If one wishes, conditional probability can be assigned to sets  $C$  expressed in terms of  $(Y, X)$  that have the form  $C = T_\theta^{-1}(B_\theta)$ , where  $B_\theta$  is a Borel subset of  $\mathfrak{R}^K$ . Probability is assigned using  $P(C | \theta) = (2\pi)^{-K/2} \int \cdots \int I_{B_\theta}(t) \exp(-t't/2) dt_1 \cdots t_K$ .

This analysis provides the logical foundations for using (19) as the likelihood in Bayesian inference. We note in passing that (19) is the objective function of the continuously updating GMM estimator (Hansen, Heaton, and Yaron, 1996).

In our particular application these ideas require further extension. The extension is in Subsection 4.3.

## 4.2 Implementation

Our likelihood is hierarchical. Its first component is of the form of the likelihood (19). For specificity we describe the implementation for monthly data. Let  $s_t$  denote the vector of gross returns at time  $t$  on the twenty-four Fama-French portfolios S11 through S54 listed in Table 2. Let  $b_t$  denote the vector of gross returns at time  $t$  on the three Treasury debt issues t30ret, b1ret, and b10ret listed in Table 2. Let  $c_t$  denote consumption growth (i.e.  $C_t/C_{t-1}$ ) and  $l_t$  denote labor income growth. Define the instruments

$$Z_t = \begin{pmatrix} s_t - 1 \\ b_t - 1 \\ c_t - 1 \\ l_t - 1 \\ 1 \end{pmatrix},$$

where  $s_t - 1$  and  $b_t - 1$  denote 1 subtracted from each element of  $s_t$  and  $b_t$ . Let the parameter  $\theta$  denote the 551 ex-post values  $\theta_1, \dots, \theta_{551}$  of the pricing kernel. Denote the vector of Euler equation errors by

$$e_t(\theta) = e_t(s_{t+1}, b_{t+1}, \theta_{t+1}) = 1 - \theta_{t+1} \begin{pmatrix} s_{t+1} \\ b_{t+1} \end{pmatrix}, \quad (20)$$

where 1 denotes a vector of 1's of length twenty-seven. The moment function that determines (18) for our estimator is

$$m_t(\theta) = m_t(s_t, b_t, s_{t+1}, b_{t+1}, \theta_{t+1}) = Z_t \otimes e_t(s_{t+1}, b_{t+1}, \theta_{t+1}),$$

where  $t = 1, \dots, n = 550$ . The length of the vector  $m_t(\theta)$  is  $K = 756$  so that the number of overidentifying restrictions on  $\theta_2, \dots, \theta_{551}$  is 206. Note that  $\theta_1$  is not yet identified because  $\theta_1$  does not appear in any of the  $m_t(\theta)$ . This issue is addressed later in this subsection.

We assume that  $\theta_t(s_t, b_t)$  has a factor structure. There is one error common to all elements of  $\theta_t s_t$ , one error common to all elements of  $\theta_t b_t$ , and twenty-seven idiosyncratic errors, one for each element of  $\theta_t(s_t, b_t)$ . Denote this matrix by  $\Sigma_e$  (or by  $\Sigma_{e,t}$  if one wants to allow for heterogeneity, which makes no difference in what follows). A set of orthogonal eigen vectors  $U_e$  for  $\Sigma_e$  are easy to construct and can be used to diagonalize  $\Sigma_e$ . To illustrate, if there were three stocks and two bonds, then

$$U_e = \begin{pmatrix} 1/3 & 1/\sqrt{2} & 1/\sqrt{6} & 0 & 0 \\ 1/3 & -1/\sqrt{2} & 1/\sqrt{6} & 0 & 0 \\ 1/3 & 0 & -2/\sqrt{6} & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/\sqrt{2} \\ 0 & 0 & 0 & 1/2 & -1/\sqrt{2} \end{pmatrix}.$$

Similarly  $U_z$  and  $\Sigma_z$  for  $Z_t$ . (The last row and column of  $\Sigma_z$  are zeros whence  $U_z$  looks like  $U_e$  with an extra 2x2 block for  $(c_t, l_t)$  and zeros appended to the bottom and right and a one in the southeast corner.)

Now  $\mathcal{E}[Z_{i,t} e_{j,t}(s_{t+1}, b_{t+1}, \theta_{t+1}^o)] = 0$ . If we strengthen the zero correlation condition and assume  $\text{Var}[Z_t \otimes e_t(s_{t+1}, b_{t+1}, \theta_{t+1}^o)] = \Sigma_z \otimes \Sigma_e$ , then we can diagonalize  $m_t(\theta)$  by multiplying on the left by  $(U_z \otimes U_e)'$ . Taking into account the block structure of  $U_z$  and  $U_e$ , computing  $(U_z \otimes U_e)'m_t(\theta)$  requires  $(24^2 + 3^2 + 2^2 + 1) + (24^2 + 3^2) + 30 \times 27 = 1,985$  operations. It is the most significant component of the computational cost of the statistical methodology proposed here. Computation of  $U'm_t(\theta)$  for general  $U$  requires  $(27 \times 30)^2 = 656,100$  operations and is so costly that our proposed methodology becomes infeasible. (We learned this the hard way by trying to do it.) The upshot is that we either accept a diagonalization scheme based on the assumption that  $\text{Var}(Z_t, e(s_{t+1}, b_{t+1}, \theta_{t+1}^o)) = \Sigma_z \otimes \Sigma_e$  is a reasonable approximation or quit here.

Let  $V_t(\theta) = (U_z \otimes U_e)'m_t(\theta)$  with elements  $v_{i,t}(\theta)$  for  $i = 1, \dots, K = 756$ . Because diagonalization implies  $\mathcal{E}(v_{t,i}(\theta^o)v_{t,j}(\theta^o)) = 0$ , we can estimate the variance of  $V_t(\theta)$  by a diagonal matrix  $S_n(\theta)$  with elements

$$s_i(\theta) = \frac{1}{n} \sum_{t=1}^n \left( v_{t,i}(\theta) - \frac{1}{n} \sum_{t=1}^n v_{t,i}(\theta) \right)^2.$$

We correct for the mean as shown because we evaluate at proposed  $\theta$  in an MCMC scheme

rather than at a value of  $\hat{\theta}_n$  that tries to put  $\bar{m}_n(\theta) = 0$  as in frequentist inference. This is to prevent having an absurd value of  $\theta$  accepted merely because it makes  $s_i(\theta)$  large.

The likelihood for the pricing kernel  $\theta$  is, then,

$$\mathcal{L}(\theta) \propto \exp \left\{ -\frac{n}{2} \bar{m}_n(\theta)' (U_z \otimes U_e) S_n^{-1}(\theta) (U_z \otimes U_e)' \bar{m}_n(\theta) \right\} \quad (21)$$

$$\bar{m}_n(\theta) = \frac{1}{n} \sum_{t=1}^n Z_t \otimes e_t(s_{t+1}, b_{t+1}, \theta_{t+1}), \quad (22)$$

The full likelihood is the product

$$\mathcal{L}(\eta, \theta) \propto \mathcal{L}(\theta) \times \prod_{t=1}^n f(\theta_{t+1} | \theta_t, \theta_{t-1}, \dots, \theta_1, \eta). \quad (23)$$

where  $f(\cdot | \cdot, \eta)$  is the law of motion given by (4).

The diagonalization and scaling scheme in (21) (equivalently, the weighting matrix  $W_n(\theta)$  of (17)) determines the relative importance each of the elements of  $\bar{m}_n(\theta) = \frac{1}{n} \sum_{t=1}^n m_t(\theta)$  has in determining results. In frequentist inference this amounts to a choice of a loss function and can be a serious consideration in applications (Cochrane, 2001, Chapter 16). With the Bayesian methodology proposed here it is less of an issue because one can use a prior to counteract the deleterious effects of a particular diagonalization scheme, which we do in Section 5.

Lags are needed to prime the recursion for  $\sigma_t^2$  of (10) that appears in  $f(\cdot | \cdot, \eta)$  of (23). This is done by prepending a copy of  $\theta$  to  $\theta$  so that  $\sigma_t^2$  recurses through a copy of  $\theta$  before computation of  $f(\cdot | \cdot, \eta)$  begins. As noted above,  $\theta_1$  does not appear in  $\mathcal{L}(\theta)$  of (21). It is a backcasted lag identified by the law of motion  $f(\cdot | \cdot, \eta)$  in (23). The role of  $\theta_1$  is to buffer the transition from the prepended copy of  $\theta$  to the values of  $\theta$  that are used to compute  $\mathcal{L}(\theta)$ .

Given the prior density described in Section 5, a Bayesian posterior is obtained from  $\mathcal{L}(\eta, \theta)$  using a general purpose, public domain, parallelized implementation of the Chernozhukov and Hong (2003) MCMC method (<http://econ.duke.edu/webfiles/arg/emm>). It is a Metropolis-Hastings algorithm so that neither the likelihood nor the prior have to be normalized to integrate to one. The proposal density randomly chooses a component of  $(\eta, \theta)$  to move and then proposes a normal random walk move. Since the dimension of  $(\eta, \theta)$  is 560 a full cycle takes 560 draws on average making the MCMC chain highly correlated. We deal with this in the standard way by running a long chain and sampling it. After burning off

transients, we generate a chain of length 30,000,000 and retain every 1000th draw leaving 30,000 draws net. The exception is that the modes reported in Section 6 are taken over the 30,000,000 draws rather than the sample of 30,000 draws. The autocorrelations in the 30,000 draws are strongly influenced by the prior. For the tightest prior they are negligible after ten lags and for the loosest after 100. While our proposal strategy lacks finesse, it is a relatively safe strategy that tends to explore a posterior well and it allows us to use well tested parallel code. It is practicable on a 16 CPU Beowulf cluster, which is what we used for the computations. It may not actually be possible to do better than this within a reasonable expenditure of intellectual effort for a likelihood as nonlinear as  $\mathcal{L}(\eta, \theta)$ .

### 4.3 Further Rationale

The rationale offered in Subsection 4.1 does not justify the use of  $\mathcal{L}(\theta)$  as a likelihood without further explanation because  $\theta$  is not usually thought of as a parameter which determines the distribution of returns but rather is usually thought to be a random variable that covaries with returns. In this subsection we shall justify the use of  $\mathcal{L}(\theta)$  from the latter point of view.

We assume that the function  $\mathcal{L}(\theta)$  given by (21) expresses our subjective opinion about the ex-ante, joint distribution of  $\theta$  and returns. This is reasonable because the ex-ante expectation of  $\bar{m}_n$  given by (22) is zero and  $\bar{m}_n$  is presumed to satisfy an unconditional central limit theorem. Let  $R$  denote a panel of returns,  $\mathcal{R}$  all possible panels, and let

$$T(R, \theta) = \sqrt{n}S_n^{-1/2}(\theta)(U_z \otimes U_e)\bar{m}_n(\theta). \quad (24)$$

As discussed in Subsection 4.1, the sets  $A$  expressed in terms of  $(R, \theta)$  to which subjective probability can be assigned using  $\mathcal{L}(\theta)$  are those that are preimages  $A = T^{-1}(B)$  of a Borel set  $B \subset \mathfrak{R}^K$ . For such  $A$  one computes its subjective probability as  $P(A) = P_T(B)$ , where

$$P_T(B) = (2\pi)^{-K/2} \int \cdots \int I_B(t) \exp(-t't/2) dt_1 \cdots t_K. \quad (25)$$

Let  $T_\theta$  denote the map  $R \mapsto T(R, \theta)$  for  $\theta$  fixed. The sets  $C$  expressed in terms of  $R$  to which we can assign conditional probability knowing that  $\theta$  has occurred have the form  $C = T_\theta^{-1}(B_\theta)$  for some Borel set  $B_\theta \subset \mathfrak{R}^K$ . The principle of conditioning is that one assigns conditional probability proportionately to joint probability. Let  $\mathcal{R}_\theta$  be the Borel set for

which  $\mathcal{R} = T_\theta^{-1}(\mathcal{R}_\theta)$ . According to this principle, the conditional probability of  $C$  given  $\theta$  is computed as  $P(C|\theta) = P_T(B_\theta)/P_T(\mathcal{R}_\theta)$ , where  $P_T(\cdot)$  is given by (25). In this instance  $\mathcal{R}_\theta = \mathfrak{R}^K$  so that  $P_T(\mathcal{R}_\theta) = 1$ . Thus  $\mathcal{L}(\theta)$  is the conditional distribution for  $R$  given  $\theta$  and is therefore the likelihood.

A natural question is what is the marginal for  $\theta$ . The answer is that there is none because the joint distribution can only assign probability to a coarse collection of sets and rectangles are not in that collection. We have a concept of a joint distribution for  $(R, \theta)$  and a concept of a conditional distribution for  $R$  given  $\theta$ . We do not yet have enough information to deduce a marginal distribution.

Suppose that we supply a marginal density  $f(\theta)$  defined on  $\mathfrak{R}^{n+1}$  that integrates to one with respect to Lebesgue measure. This puts additional information into the analysis and allows us to assign probability to sets of the following form

$$E = \{(R, \theta) : R \in T_\theta^{-1}(B_\theta), \theta \in B\},$$

where  $B_\theta$  is a Borel subset of  $\mathfrak{R}^K$  that can depend on  $\theta$  and  $B$  is a Borel subset of  $\mathfrak{R}^{n+1}$ . Probabilities are assigned using

$$\begin{aligned} P(E) &= \int \cdots \int P_T(B_\theta) f(\theta) d\theta_1 \dots d\theta_{n+1} \\ &= (2\pi)^{-K/2} \int \cdots \int I_{B_\theta}(t) \exp(-t't/2) f(\theta) dt_1 \cdots dt_K d\theta_1 \dots d\theta_{n+1}. \end{aligned}$$

If we summarize the data using  $t = T(R, \theta)$ , then this expression says that we can use  $(2\pi)^{-K/2} \exp(-t't/2)$  as the conditional density for Bayesian inference.

Another way of viewing the foregoing argument is to note that we have verified that in our application  $T(R, \theta)$  is a pivotal and the method of deriving a conditional distribution from this fact is the same as is used to derive fiducial probability from a pivotal (Fisher, 1930). The difference is that for fiducial probability one would condition on  $R$  whereas we condition on  $\theta$ . But the argument is symmetric in  $R$  and  $\theta$  so that this does not matter. On this see also Hampel (2003) and Pitman (1957) and note that in our application the confusion in the fiducial literature as to whether  $\theta$  is to be regarded as fixed or random does not arise.

## 5 The Prior Distribution

In Section 2 we derived the substantive portion of the prior, which is  $p(\eta)$  defined in Table 1. It has three variants: tight, intermediate, and loose. Here we complete the description of the prior by specifying its technical component  $p_T(\theta, \eta)$ . The prior imposed on the estimation is the product  $p(\theta, \eta) = p(\eta)p_T(\theta, \eta)$ . We use the Metropolis-Hastings algorithm in our MCMC chain so there is no need to normalize  $p(\theta, \eta)$  to integrate to one. In the remainder of this section we describe each component of  $p_T(\theta, \eta)$  in isolation;  $p_T(\theta, \eta)$  itself is the product of these components.

The support conditions are that  $r_0, r_1, r_2$  of  $\eta$  be positive and that both (9) and (10) be mean reverting. The latter involves checking the largest eigen values of companion matrices that are automatically computed by the SNP software. The support conditions are implemented as an indicator function that is one when they are all true and zero else. This is the first component of  $p_T(\theta, \eta)$ .

As mentioned earlier, a criticism of GMM style inference is that the weighting matrix determines the loss function with the consequence that moments that have economic importance can get neglected (Cochrane, 2001, Chapter 16). Here the specific criticism is that because the variance of returns on stocks is larger than for bonds, the Euler equations for bonds unreasonably influence results. To mitigate this criticism we impose the loosest prior that forces all the Euler equation errors to be about the same. Let

$$e_n(\theta) = \frac{1}{n} \sum_{t=1}^n e_t(s_{t+1}, b_{t+1}, \theta_{t+1}),$$

where  $e_t(\cdot)$  is given by (20). For annual data we impose a normal prior such that  $P(-0.5 < e_{i,n}(\theta) < 0.5) = 0.95$  and for monthly data we impose  $P(-0.05 < e_{i,n}(\theta) < 0.05) = 0.95$ . For the annual data these are components two through twenty-six of  $p_T(\theta, \eta)$ ; for the monthly data two through twenty-nine.

In order to assess the plausibility of recursive utility we need to compute the gross return to the consumption endowment using (12) and for this we need

$$\sum_{j=1}^{\infty} \mathcal{E}_t \prod_{k=1}^j \left( \frac{C_{t+k}}{C_{t+k-1}} \theta_{t-1+k,t+k} \right)$$

to converge when computed from a law of motion estimated from the realization of  $\frac{C_{t+k}}{C_{t+k-1}}$  given by the data and  $\theta$  given by, for example, the mean of the posterior. The smaller are the elements of  $\theta$ , the better the chance of success. The magnitude of  $\theta$  can be quantified as the average

$$P_B = \frac{1}{n} \sum_{t=1}^n \theta_t,$$

which is positive because the support of  $f(\cdot|\cdot, \eta)$  in (23) is positive.  $P_B$  can be regarded as a crude approximation to the average price of a risk-free, one-period bond. Let  $f(P_B) \propto 1 + \cos(\alpha + \beta P_B)$  if  $a < P_B < b$  and zero else, where  $\beta = 2\pi/(b-a)$  and  $\alpha = \pi - \beta b$ . We choose  $a$  to correspond to a 4% bond and  $b$  to correspond to a 1% bond, taking compounding into account for the monthly bond. The exact values are  $(a, b) = (0.9967369, 0.99171)$  for the monthly data and  $(a, b) = (0.9615385, 0.990099)$  for the annual data. This has the net effect of making  $1/P_B$  about 0.5% higher than it would have been had we not imposed the prior. This is the last component of  $p_T(\theta, \eta)$ .

We now have all ingredients in place. The ex-ante, subjective, joint distribution of  $(R, \theta, \eta)$  is determined by a density that is proportional to

$$\mathcal{L}(\theta, \eta)p(\theta, \eta) = \mathcal{L}(\theta) \prod_{t=1}^n f(\theta_{t+1} | \theta_t, \eta)p(\theta, \eta),$$

where  $R$  denotes a panel of returns,  $\theta$  denotes the path of the stochastic discount factor,  $\eta$  denotes the parameters of the law of motion  $f(\cdot|\cdot, \eta)$ ,  $\mathcal{L}(\theta)$  denotes the likelihood for  $R$  given by (21), and  $p(\theta, \eta)$  denotes the prior for  $(\theta, \eta)$  described in this section. The data are summarized by the  $K$ -vector  $t = T(R, \theta)$ , where  $T$  is given by (24), and the interpretation of  $\mathcal{L}(\theta)$  reflects this fact. For instance, in order to compute the conditional probability of a set  $C$  expressed in terms of  $R$  using  $\mathcal{L}(\theta)$  one must use a change of variables procedure described in Subsection 4.3. In the next section we shall observe an ex-post panel of returns and use Bayes rule to assign probability to putative ex-post values for  $(\theta, \eta)$ .

## 6 Empirical Results

Tables 4 and 5 show the mean, mode, and standard deviation of the posterior distribution of the parameter  $\eta$  subject to the tight, intermediate, and loose priors. The parameter  $\eta$

determines the conditional density  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$  of the pricing kernel given by (4). We shall focus on the the posterior mode for  $\eta$  in our discussion of  $f(\cdot|\cdot, \eta)$  rather than the posterior mean because the posterior mode corresponds to a density  $f(\cdot|\cdot, \eta)$  that was drawn in the MCMC chain whereas  $f(\cdot|\cdot, \eta)$  evaluated at the posterior mean might not resemble any density that was drawn in the MCMC chain.

In both tables the important changes when moving from the posterior mode under the tight prior to the loose prior are (1) a decrease in the variance parameter  $r_1$  compensated by an increase in kurtosis via  $a_4$  and (2) a shift to more sensitivity to the immediate past represented by increases in the magnitudes of the AR parameter  $b_1$  and the ARCH parameter  $r_2$ . The standard deviations of the posterior distribution of  $\eta$  under the loose prior are an order of magnitude smaller than the standard deviations of the loose prior itself which suggests that loose prior on  $\eta$  does not have much influence on the posterior distribution. By the same logic, the tight prior binds and the intermediate prior is informative. Next we undertake a graphical analysis to confirm the characterization of the dynamics of the pricing kernel inferred from Tables 4 and 5.

Table 4 about here

Table 5 about here

Figures 1 and 2 plot the conditional densities  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$  with  $\eta$  set to the posterior mode under each of the three priors for various conditioning sequences  $(\theta_t, \dots, \theta_1)$ . When interpreting Figures 1 and 2, recall that the unconditional mean is constrained to be between 0.996 and 0.999 for the monthly data and between 0.96 and 0.99 for the annual data and be aware that the conditional mean is to the right of the conditional mode due to the heavy right tails.

As seen from Figures 1 and 2, the tight prior binds and effectively imposes the distribution of  $M_{t-1,t}$  determined by Kiku's (2006) simulation of a Bansal-Yaron (2004) economy on  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$  as evidenced by the fact that the black line, which is  $f(\cdot|\cdot, \eta)$  under the prior, and the dashed line, which is  $f(\cdot|\cdot, \eta)$  under the tight prior, overplot. The dashed line is only distinguishable by virtue of being plotted at a thicker line width than the solid line.

The posterior distribution of  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$  under the intermediate prior moves away from  $f(\cdot|\cdot, \eta)$  under the tight prior under some conditions, notably the crash of October 1987 for the monthly prior and the years 1934 and 1981 for the annual prior. Otherwise it is close to  $f(\cdot|\cdot, \eta)$  under the tight prior.

The loose prior on  $\eta$  appears to be non-informative in a practical sense. Under it, the movements of  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$  away from plots under the tight prior are large under all conditioning sets for both the monthly and annual data, sometimes dramatically. It is clear that there are features of the conditional distribution of the pricing kernel that a Bansal-Yaron economy fails to explain: (1) the conditional variance of the density implied by a Bansal-Yaron economy is too large and location and scale do not move sharply enough with changes in the conditioning set; (2) there is predictability in the pricing kernel that a Bansal-Yaron model misses.

Although it is obvious from Figures 1 and 2 that the model under the loose prior is preferred, we check numerically by assigning equal prior probability to the models under the tight, loose, and intermediate priors and computing the posterior probability attached to each model in Table 6. As seen from Table 6, the posterior probability of the loose model is 0.9975 from the monthly data and 0.9772 from the annual data. A posteriori the model under the loose prior is strongly preferred.

Figure 1 about here

Figure 2 about here

Table 6 about here

The sample paths  $\{\theta_t\}_{t=1}^n$  of the pricing kernel are of interest. Here we focus on the posterior mean  $\{\bar{\theta}_t\}$  because it is more representative of the entire ensemble of MCMC draws than any single draw such as the posterior mode, and there is no feature of an actual draw that has to be preserved as there was in our discussion of  $\eta$ . The posterior mean of the pricing kernel under the loose prior is plotted as the solid line in Figures 3 and 4 together with pointwise standard errors added and subtracted, which are shown as dotted lines. There

is considerable difference in the information content of the two posterior distributions. As measured by signal to noise, only 1% of the ratios  $\bar{\theta}_t/\sqrt{Var}(\theta_t)$  exceed one for the monthly data whereas 70% of the ratios exceed one for the annual data.

The most striking feature of Figures 3 and 4 is the periods of extreme volatility from 1970 through 1975 and 2000 through 2003. Also there is a discrepancy in 1981 and 1982 where  $\bar{\theta}_t$  is much higher for the annual data than for the monthly. This may be due to the differences between the debt portfolios in the annual and monthly data. The annual data panel includes only the thirty day Treasury obligation whereas the monthly data includes the one year and the ten year obligations as well. Annual returns on the thirty day obligation have a smooth path with a marked peak in 1981. Monthly returns on the thirty day obligation have a visually apparent but noisy peak in 1981. Monthly returns on the one year and ten year obligations have increased volatility in the early 1980's with no isolated peak. It may or not be the case that the annual posteriors give more weight to debt portfolios than do the monthly posteriors despite the prior on Euler equation errors that we imposed to prevent this, but it certainly is the case that monthly returns on debt are different visually from annual returns. There is the additional factor that events prior to 1960 are influencing results post 1960 in the annual data through the law of motion  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$ . There is evidence of different behavior of substance before and after 1960 a few paragraphs hence.

Figure 3 about here

Figure 4 about here

Recursive utility implies that the pricing kernel is

$$\theta_{t+1} = M_{t,t+1} = \delta^\theta \left( \frac{C_{t+1}}{C_t} \right)^{-(\theta/\psi)} \left( R_{c,t+1} \right)^{(\theta-1)}$$

which, in turn, implies that the regression

$$\log(\theta_{t+1}) = c_1 + c_2 \log \left( \frac{C_{t+1}}{C_t} \right) + c_3 \log \left( R_{c,t+1} \right)$$

should have an  $R^2$  of one. However, if  $C_t$  and  $R_{ct}$  are measured with error, then the  $R^2$  will be smaller than one. Also, if  $R_{ct}$  is omitted from the regression or a proxy is substituted for

$R_{ct}$ , then the  $R^2$  will be smaller than one. We use these observations to check the plausibility of recursive utility in three ways.

The first check is to compute the posterior distribution of the  $R^2$  from a regression of the log pricing kernel,  $\log(\theta_{t+1})$ , on log consumption growth,  $\log\left(\frac{C_{t+1}}{C_t}\right)$ , and compare these  $R^2$  to those that would obtain in a Bansal-Yaron economy.

For the monthly data we use contemporaneous log consumption growth and eleven lags in these regressions because consumption growth is seasonally adjusted. If constant seasonal weights  $w_1, \dots, w_{12}$  have been applied to a variable  $x_t$  and twelve consecutive values of the logarithm of that variable are entered into a regression, then the regressors are actually twelve consecutive values of the log of the unadjusted variable plus  $\sum_{j=1}^{12} \beta_j \log(w_j)$ , where the  $\beta_j$  are the regression coefficients of  $\log(x_{t-j})$ . The term  $\sum_{j=1}^{12} \beta_j \log(w_j)$  is a constant. U.S. agencies use X12 to seasonally adjust data which means that  $\sum_{j=1}^{12} \beta_j \log(w_j)$  will be slowly varying rather than constant. But we still expect twelve lags to substantially undo the effects of seasonal adjustment. For the annual data we regress the log of the pricing kernel on contemporaneous log consumption growth.

The  $R^2$  that one would expect to see in a Bansal-Yaron economy are shown in Tables 7 and 8. The posterior distribution of  $R^2$  that we do see are shown in Tables 9 and 10. Before implementing our first check on the plausibility of recursive utility, we comment on these tables.

Although we only need the  $R^2$  for contemporaneous consumption for the annual data, we computed  $R^2$  up to four lags in Tables 8 and 10. We see that predictability increases with lag in Table 10 more than would be expected from Table 8 implying that there is predictability in the pricing kernel that a Bansal-Yaron economy fails to explain. This is consistent with our findings with respect to the posterior distribution of  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$  above that there is more predictability in the pricing kernel than is implied by a Bansal-Yaron economy. The same phenomenon is noted when one compares Table 7 to Table 9 but we do not know how much predictability is induced by seasonal adjustment and how much is due to the Bansal-Yaron misspecification of  $f(\cdot|\cdot, \eta)$ .

The tables also imply that a consequence of this misspecification is to make a test that accepts when a posterior  $R^2$  is large a conservative test. To see this, note that  $R^2$  decrease

as one moves from left to right in Tables 9 and 10. That is,  $R^2$  decrease as one moves from the tight to the loose prior. The implication is that the  $R^2$  computed from a simulation of a Bansal-Yaron economy in Tables 7 and 8 tend to be too large.

Returning to the first check, the upper panels of Tables 11 and 12 show the posterior probabilities of recursive utility computed from the regression  $R^2$  that one would expect, shown in Tables 7 and 8, and the posterior distribution of  $R^2$  that one finds, shown in Tables 9 and 10. For instance, from the upper panel of Table 7 one expects an  $R^2$  of 0.0178 at lag eleven for 2% measurement error. From the upper panel of Table 9 under the tight prior at lag eleven the 25% quantile is 0.0149 and the 50% quantile is 0.0197. That means that the posterior probability of observing an  $R^2$  of 0.0178 or larger is somewhere between 50% and 75%. Using a finer grid (not shown) we get 60%.

For the monthly data the posterior probability attached to recursive utility is (<1%, 60%, 99%, 99%) for (1%, 2%, 5%, 10%) measurement error under all three priors as seen from the upper panel of Table 11. For the annual data posterior probabilities range from (<1%, 20%, 90%, 99%) under the tight prior to (<1%, 10%, 75%, 95%) under the loose prior as seen from the upper panel of Table 12. One can summarize by stating that recursive utility is implausible at 1% measurement error, somewhat plausible at 2% measurement error, and plausible at higher rates of measurement error according to our first check.

It is possible to use Tables 9 and 10 to assess the posterior plausibility of external habit utility as implemented by Campbell and Cochrane (1999). Using the parameter settings  $(g, \sigma, \phi, \delta, \gamma) = (0.01639, 0.0006887, 0.9861, 0.9955, 0.5726)$  taken from Gallant and McCulloch (2005) for the monthly data and the settings  $(g, \sigma, \phi, \delta, \gamma) = (0.01967, 0.02386, 0.8450, 0.9477, 0.5726)$  from the same source for the annual data we compute the  $R^2$  one expects in a regression of the log pricing kernel on log consumption growth (plus eleven lags in the case of monthly data) for an external habit model under measurement error in the same manner as Tables 7 and 8. The predicted  $R^2$  values obtained with measurement error of (0%, 1%, 2%, 5%, 10%) are (1, 0.0455, 0.0044, 0.0003, 4.3e-5) for the monthly data and (1, 0.8980, 0.6776, 0.1931, 0.0295) for the annual data. With these we can compute posterior probabilities from the upper panel of Tables 9 and 10 in the same manner as above. The implied posterior probabilities are (<1%, <1%, 10%, 85%) for the monthly data at measurement

errors of (1%, 2%, 5%, 10%). For the annual data they are (<1%, <1%, <1%, 45%). One can summarize by stating that an external habit utility is implausible at plausible rates of measurement error.

The second check is to compare the posterior distribution of the  $R^2$  from a regression of  $\log(\theta_{t+1})$  on  $\log\left(\frac{C_{t+1}}{C_t}\right)$  and  $\log(R_{s,t+1})$  (plus eleven lags in the case of monthly data), where  $R_{s,t+1}$  is the market return. The market return  $R_{s,t+1}$  is sometimes used as a proxy for the return to wealth  $R_{c,t+1}$  in implementations of recursive utility (Campbell, 1996, p. 301). For monthly data the posterior probability attached to a recursive utility is (<1%, <1%, <1%, 1%) for (1%, 2%, 5%, 10%) measurement error under all three priors as seen from the middle panel of Table 11. For the annual data it is (<1%, <1%, <1%, <1%) under all three priors as seen from the second panel of Table 12. One can summarize by stating that recursive utility is implausible according to our second check

Our third check requires model-free computation of  $R_{ct}$ , which we compute using (12) and (13) as follows. We fit an SNP density to the process

$$y_t = \log\left(\theta_t \frac{C_t}{C_{t-1}}\right),$$

where  $\theta_t$  is the posterior mean of the pricing kernel and compute  $P_{ct}$  using (12) by simulating  $y_t$  100 steps ahead and averaging over 1000 Monte Carlo repetitions conditional upon past values of  $y_t$ . We then compute  $R_{ct}$  using (13). Computing (12) is essentially a multi-step-ahead forecasting exercise. We view past values of  $y_t$  as an adequate conditioning set for this forecasting exercise because it is a well understood principle of forecasting that a multi-step-ahead forecast of a process from its own past is nearly always more accurate than a multi-step-ahead forecast of a vector process that includes that variable even if the specification of the vector process is correct.

The third check is to compare the posterior distribution of  $R^2$  for a regression of  $\log(\theta_{t+1})$  on  $\log\left(\frac{C_{t+1}}{C_t}\right)$  and  $\log(R_{c,t+1})$  (plus eleven lags in the case of monthly data). For monthly data the posterior probability attached to a recursive utility ranges from (15%, 99%, 99%, 99%) for (1%, 2%, 5%, 10%) measurement error under the tight prior to (55%, 99%, 99%, 99%) to under the loose prior as seen from the bottom panel of Table 11. For the annual data it ranges from (<1%, <1%, 80%, 99%) to (<1%, <1%, 50%, 99%) as seen from the

third panel of Table 12. For the annual data we recomputed the  $R^2$  using only the years 1960–2004. These posterior probabilities are shown in the bottom panel of Table 12. For the recomputed  $R^2$  the probabilities become (<1%, 60%, 99%, 99%,) to (<1%, 20%, 99%, 99%), which partly reconciles the discrepancy between the monthly and annual results. One can summarize by stating that recursive utility is somewhat plausible at 1% measurement error and plausible at higher rates of measurement error according to our third check.

Table 7 about here

Table 8 about here

Table 9 about here

Table 10 about here

Table 11 about here

Table 12 about here

Our own view is that the evidence supports recursive utility for the following reasons.

The first check is a joint test of the driving process for consumption growth in a Bansal-Yaron economy and recursive utility and therefore overly stringent for the purpose of testing recursive utility per se. But this overly stringent test is passed in both the monthly data and the annual data. Our posterior analysis of  $f(\theta_{t+1}|\theta_t, \dots, \theta_1, \eta)$  suggested that the processes driving consumption in a Bansal-Yaron economy are misspecified but it would seem that this misspecification is not severe in directions that the first test can detect.

The first check does seem to have some discriminatory power because it definitively prefers a Bansal-Yaron economy to an external habit economy. One should note that the marginal rate of substitution implied by external habit is computable directly from the endowment process and so our check of external habit does not use Campbell and Cochrane's specification of the driving process. It is a direct test whose only reliance on augmenting

hypotheses is our choice of parameter settings, which do not differ materially from Campbell and Cochrane’s calibration.

According to the third check, recursive utility is strongly supported by the monthly data and moderately supported by the annual data. The third check relies on our subsidiary wealth return computation (12). The reliance on the characteristics of a Bansal-Yaron economy and the correctness of Kiku’s computations is fairly minimal. One concern might be sensitivity of this computation to the forecast horizon. It is 100 months for the monthly data and 100 years for the annual data. The computation is so costly that extension of the forecast horizon much beyond these limits is impractical. However the computation does not appear to be sensitive to the forecast horizon. For the annual data and loose prior we recomputed at 10 and 25 year horizons. The largest change in the posterior quantiles shown in the first line of the lower panel of Table 10 was 12%. This is not enough of a change to materially affect posterior probabilities.

We think that the disagreement of the second check with the first and third checks can easily be explained. The  $R^2$  that we expect to see in the first check relies on Bansal-Yaron consumption dynamics. The  $R^2$  that we expect to see in the second check relies, in addition, on the Bansal-Yaron specification of dividend dynamics. As Campbell (1996) points out, it is implausible that the return to publicly traded equity is good proxy for the return to total wealth because publicly traded equities are a small fraction of total wealth. Yet we see from the middle and lower panels of Tables 7 and 8 that as measurement error increases the return to equity becomes a better proxy for the return to wealth than the return to wealth itself. This strikes us as an implausible description of reality. Thus, it would seem that the  $R^2$  shown in the middle panel of Tables 7 and 8 can be dismissed as implausibly large and therefore the fact that the second check contradicts the first and third is not surprising. We view the outcome of the second test as evidence that the dividend dynamics in a Bansal-Yaron economy are more seriously misspecified than the consumption dynamics.

## 7 Conclusion

We used a Bayesian statistical method that can extract the pricing kernel and its transition density from a panel of returns to determine if it is plausible that the pricing kernel can

be represented as the intertemporal marginal rate of substitution of a representative agent with recursive utility. We used two panels of U.S. equity and bond returns. The first was monthly from 1959–2004 and the second annual from 1930–2004. Our priors were formed from an examination of a long simulation of a calibrated Bansal-Yaron (2004) economy. We concluded that recursive utility is plausible a posteriori.

We also concluded from an analysis of the law of motion extracted by our Bayesian procedure that the Bansal-Yaron consumption dynamics are misspecified because they generate a pricing kernel that has less predictability than is seen in the dynamics of the pricing kernel that we extracted. The misspecification is not strong enough to cause our joint test of recursive utility and Bansal-Yaron consumption dynamics to reject. The misspecification of the Bansal-Yaron dividend process seems more severe because it did cause our joint test of recursive utility and Bansal-Yaron dynamics to reject.

Incidentally we determined that external habit utility in the form set forth in Campbell and Cochrane (1999) is implausible a posteriori.

Our methodology makes no assumptions other than that a unique pricing kernel exists and therefore the extensive tabular results that describe our posterior density are model free and provide a rich collection of statistics that we hope other researchers can use to conduct similar analyses with other dynamics and utility functions.

## 8 References

- Bansal, Ravi, A. Ronald Gallant, and George Tauchen (2004), “Rational Pessimism, Rational Exuberance, and Markets for Macro Risks,” Manuscript, Fuqua School of Business, Duke University, Durham NC.
- Bansal, Ravi, and Amir Yaron (2004) “Risks For the Long Run: A Potential Resolution of Asset Pricing Puzzles”, *Journal of Finance* 59, 1481–1509.
- Campbell, John Y., (1996), “Understanding Risk and Return,” *The Journal of Political Economy* 104, 298–345.
- Campbell, John Y., and John Cochrane (1999) “By Force of Habit: A Consumption-based

- Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy* 107, 205–251.
- Chernozhukov, Victor, and Han Hong (2003), “An MCMC Approach to Classical Estimation,” *Journal of Econometrics* 115, 293–346.
- Cochrane, John H., (2001), *Asset Pricing*, Princeton University Press, Princeton, NJ.
- Del Negro, Marco, and Frank Schorfheide (2004), “Priors from General Equilibrium Models for VARS,” *International Economic Review* 45, 643–673.
- Duan, Jason A., and Carl F. Mela (2006), “The Role of Spatial Demand on Outlet Location and Pricing,” Manuscript, Fuqua School of Business, Duke University, Durham NC.
- Epstein, L. G., and Stanley Zin (1989), “Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework,” *Econometrica* 57, 937–969.
- Eraker, Bjorn, (2006), “Affine General Equilibrium Models,” Manuscript, Department of Economics, Duke University, Durham NC.
- Fama, Eugene, and Kenneth French (1992), “The Cross-Section of Expected Stock Returns,” *Journal of Finance* 59, 427–465.
- Fama, Eugene, and Kenneth French (1993), “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics* 33, 3–56.
- Fernandez-Villaverde, Jesus, and Juan F. Rubio-Ramirez (2004), “Estimating Macroeconomics Models: A Likelihood Approach,” Manuscript, Department of Economics, Duke University, Durham NC.
- Fisher, Sir Ronald A., (1925), “Theory of Statistical Estimation,” *Proceedings of the Cambridge Philosophical Society* 22, 700–725.
- Fisher, Sir Ronald A., (1930), “Inverse Probability,” *Proceedings of the Cambridge Philosophical Society* 26, 528–535.

- Gallant, A. Ronald, (1987), *Nonlinear Statistical Models*, Wiley, New York.
- Gallant, A. Ronald, and Robert E. McCulloch (2005), “On the Determination of General Statistical Models with Application to Asset Pricing,” Manuscript, Fuqua School of Business, Duke University, Durham NC.
- Gallant, A. Ronald, and George Tauchen (1989), “Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications,” *Econometrica* 57, 1091–1120.
- Hampel, Frank, (2003), “The Proper Fiducial Argument,” Research Report No. 114, Eidgenössische Technische Hochschule (ETH), Zurich.
- Hansen, Lars Peter, John Heaton, and Nan Li (2005), “Consumption Strikes Back: Measuring Long-Run Risk,” Manuscript, Department of Economics, University of Chicago, Chicago IL.
- Hansen, Lars Peter, John Heaton, and Amir Yaron (1996), “Finite-Sample Properties of Some Alternative GMM Estimators,” *Journal of Business and Economic Statistics* 14, 262–280.
- Kiku, Dana, (2006), “Is the Value Premium a Puzzle?” Manuscript, Department of Economics, Duke University, Durham NC.
- Kreps, D. M., and E. L. Porteus (1978), “Temporal Resolution of Uncertainty and Dynamic Choice,” *Econometrica* 46, 185–200.
- Newton, Michael A., and Adrian E. Raftery (1994), “Approximate Bayesian Inference with the Weighted Likelihood Bootstrap,” *Journal of the Royal Statistical Society (B)* 56, 3–48.
- Pitman, E. J. G., (1957), “Statistics and Science,” *Journal of the American Statistical Association* 52, 322–330.

Romeo, Charles, (2004), “A Gibbs Sampler for Mixed Logit Analysis of Differentiated Product Markets Using Aggregated Data,” Manuscript, Economic Analysis Group, U.S. Department of Justice, Washington DC.

Schwarz, G., (1978), “Estimating the Dimension of a Model,” *Annals of Statistics* 6, 461–464.

Weil, Philippe, (1990), “Nonexpected Utility in Macroeconomics,” *The Quarterly Journal of Economics* 105, 29–42.

Table 1. The Prior Distribution on the Parameters of the Law of Motion of the Pricing Kernel

Parameter $\eta$	Monthly		Annual	
	Location	Scale	Location	Scale
$a_1$	-0.00776	0.02203	-0.03835	0.04292
$a_2$	0.05157	0.04670	0.04199	0.03535
$a_3$	0.01967	0.00722	0.00065	0.01323
$a_4$	0.03365	0.00941	0.07604	0.01354
$b_0$	0.01684	0.03800	0.08739	0.07855
$b_1$	-0.02091	0.01470	0.00002	0.02661
$r_1$	0.09952	0.01374	0.48020	0.09367
$r_2$	0.24192	0.02239	0.32158	0.05073
$r_3$	0.95979	0.00432	0.77915	0.08241

Shown are the location and scale of the prior distribution on the parameter  $\eta$  of the SNP transition density (4). Kernels  $\{\theta_t\}$  to which the law of motion (4) are applied are scaled as  $y_t = (\log \theta_t + 0.0149120)/0.0251032$  for the monthly sampling interval and  $y_t = (\log \theta_t + 0.162398)/0.290158$  for the annual;  $\eta$  as shown reflect this scaling. The prior for the monthly sampling rate is the prior that one would form after observing a Bansal-Yaron (2004) economy at the parameter settings of Kiku (2006) for 417 years. The prior for the annual sampling rate is the prior that one would form after 1600 years of observation. The priors used in the analysis are a normal prior with location as shown and scale divided by ten, a normal prior with location and scale as shown, and a normal prior with scale multiplied by ten. These three priors are called the tight, intermediate, and loose priors, respectively.

Table 2. Simple Statistics for the Monthly Data

series	mean	std dev	variance	skewness	kurtosis
S11	0.00706525	0.0802603	0.00644171	0.0131425	2.23479
S12	0.0124367	0.0687695	0.00472924	0.0795228	3.16875
S13	0.0127413	0.0586878	0.00344426	-0.111275	2.97526
S14	0.0150867	0.0547156	0.00299379	-0.106023	3.39949
S15	0.0159196	0.057735	0.00333333	-0.0725225	3.62517
S21	0.00840988	0.0727346	0.00529033	-0.288341	1.42296
S22	0.0110266	0.0591875	0.00350316	-0.432666	2.45035
S23	0.0132057	0.0521504	0.00271966	-0.487761	3.33043
S24	0.0140323	0.0506762	0.00256807	-0.358239	3.39873
S25	0.0149585	0.0561855	0.00315681	-0.271674	3.42558
S31	0.00877447	0.066831	0.00446639	-0.28299	1.27429
S32	0.0118487	0.053009	0.00280995	-0.568934	2.97323
S33	0.0114416	0.0479753	0.00230163	-0.565203	2.76251
S34	0.0130511	0.0470064	0.00220961	-0.249837	2.55519
S35	0.0142133	0.0529927	0.00280823	-0.299185	3.54204
S41	0.00966034	0.0594082	0.00352933	-0.144753	1.53655
S42	0.00980133	0.0500286	0.00250286	-0.521865	3.07117
S43	0.01196	0.0474238	0.00224902	-0.377258	2.82729
S44	0.0129354	0.0459593	0.00211226	-0.0630278	1.86861
S45	0.0129713	0.0525289	0.00275928	-0.148915	2.53809
S51	0.00872889	0.0474538	0.00225186	-0.196489	1.50865
S52	0.00920998	0.0445955	0.00198876	-0.305751	1.72442
S53	0.0099928	0.042268	0.00178658	-0.207737	2.4257
S54	0.00993414	0.0418485	0.0017513	0.0668961	1.3031
vwretd	0.0093053	0.0420116	0.00176498	-0.388699	2.05291
t30ret	0.00449958	0.00232253	5.39413e-06	1.16436	2.15888
b1ret	0.00546105	0.00553156	3.05982e-05	2.3569	16.0043
b10ret	0.00591054	0.0221827	0.000492074	0.363896	1.42381
cg	1.00205	0.00392879	1.54354e-05	0.0470574	1.28515
lg	1.0019	0.00592699	3.51292e-05	-1.27554	35.1467

The sample size is  $n = 551$ .  $S_{ij}$  denotes the Fama-French portfolio for size quintile  $i$  and book equity to market equity quintile  $j$ .  $vwretd$  denotes CRSP value weighted returns on NYSE, AMEX, and NASDAQ stocks.  $t30ret$ ,  $b1ret$ ,  $b10ret$  denote returns on U.S. Treasury thirty day, one year, and ten year debt, respectively.  $cg$  is consumption expenditure growth and  $lg$  is labor income growth. All returns are real;  $lg$  and  $cg$  are real, per-capita.

Table 3. Simple Statistics for the Annual Data

series	mean	std dev	variance	skewness	kurtosis
S11	0.0795453	0.366403	0.134251	0.400457	0.803673
S12	0.141255	0.364505	0.132864	0.488452	0.92771
S13	0.176035	0.335697	0.112693	0.217829	0.475074
S14	0.215857	0.429625	0.184577	2.78424	14.9469
S15	0.218001	0.357469	0.127784	0.603246	1.11051
S21	0.113784	0.304723	0.0928563	0.225833	-0.398236
S22	0.154365	0.291381	0.0849029	0.548953	1.79904
S23	0.176127	0.292259	0.0854154	0.70081	2.19993
S24	0.182411	0.313625	0.0983608	0.845125	3.2986
S25	0.191585	0.315609	0.0996087	0.211441	0.615325
S31	0.125573	0.29068	0.0844948	0.940837	4.05721
S32	0.148883	0.264816	0.0701273	0.447725	2.39609
S33	0.153194	0.257398	0.0662537	0.354567	1.04735
S34	0.166249	0.266329	0.0709312	0.125378	0.509677
S35	0.177283	0.320206	0.102532	0.176807	0.98299
S41	0.115437	0.222926	0.0496959	0.00815978	0.229493
S42	0.128749	0.241885	0.0585082	1.08298	5.26054
S43	0.150266	0.25514	0.0650964	0.421346	2.55637
S44	0.158126	0.264845	0.0701427	-0.078021	0.644494
S45	0.16623	0.336853	0.11347	1.12868	4.71723
S51	0.106063	0.198638	0.039457	-0.265064	-0.507773
S52	0.10374	0.18779	0.035265	-0.369049	0.434111
S53	0.119079	0.217012	0.0470941	-0.347299	3.28346
S54	0.124612	0.251266	0.0631345	0.199181	3.01312
vwretd	0.112128	0.194068	0.0376623	-0.387258	0.438181
t30ret	0.0360364	0.0303867	0.000923352	0.850801	0.330985
cg	1.02125	0.0229756	0.000527879	-1.37777	4.80132
lg	1.02503	0.0491977	0.00242041	0.184789	1.59688

The sample size is  $n = 75$ .  $S_{ij}$  denotes the Fama-French portfolio for size quintile  $i$  and book equity to market equity quintile  $j$ .  $vwretd$  denotes CRSP value weighted returns on NYSE, AMEX, and NASDAQ stocks.  $t30ret$  denotes returns on U.S. Treasury thirty day debt.  $cg$  is consumption expenditure growth and  $lg$  is labor income growth. All returns are real;  $lg$  and  $cg$  are real, per-capita.

Table 4. Posterior Distribution of the Parameters of the Law of Motion of the Monthly Pricing Kernel

$\eta$	Tight Prior			Intermediate Prior			Loose Prior		
	Mean	Mode	Std.Dev.	Mean	Mode	Std.Dev.	Mean	Mode	Std.Dev.
$a_1$	-0.0090738	-0.01053	0.0021428	-0.077195	-0.069801	0.01619	-0.1009	-0.050606	0.049225
$a_2$	0.053431	0.052584	0.0044374	0.054482	0.018227	0.028107	-0.083432	-0.024971	0.047376
$a_3$	0.019685	0.020624	0.00071004	0.023048	0.021965	0.0069036	0.15101	0.040962	0.039
$a_4$	0.033732	0.03377	0.00093304	0.03959	0.03318	0.0086597	0.18599	0.23356	0.041314
$b_0$	0.016074	0.013288	0.0037388	-0.027799	-0.038383	0.03545	-0.46205	-0.5447	0.1281
$b_1$	-0.020842	-0.020651	0.0014362	-0.0166	-0.013206	0.014301	0.02608	0.11015	0.062157
$r_1$	0.099588	0.098673	0.0013623	0.10183	0.094246	0.013287	0.24502	0.048576	0.056492
$r_2$	0.24352	0.24489	0.0020933	0.26038	0.28599	0.012799	0.30959	0.54743	0.023572
$r_3$	0.95994	0.96019	0.0004126	0.95868	0.95272	0.0032045	0.9323	0.83509	0.0086239

Shown are the mean, mode, and standard deviation of the posterior distribution of the parameters  $\eta_i$  of the SNP transition density (4) that describes the law of motion of the pricing kernel  $\theta_t$ . Before applying these priors, data are normalized by  $y_t = (\log \theta_t + 0.0149120)/0.0251032$  to conform to the conventions of Table 1. The tight prior uses the scaling of Table 1 divided by ten, the intermediate prior uses the scaling of Table 1, and the loose prior uses the scaling of Table 1 multiplied by ten.

Table 5. Posterior Distribution of the Parameters of the Law of Motion of the Annual Pricing Kernel

$\eta$	Tight Prior			Intermediate Prior			Loose Prior		
	Mean	Mode	Std.Dev.	Mean	Mode	Std.Dev.	Mean	Mode	Std.Dev.
$a_1$	-0.039511	-0.033577	0.004254	-0.099579	-0.097374	0.035642	-0.42682	-0.38172	0.21314
$a_2$	0.042694	0.042747	0.0035054	0.0639	0.080284	0.033432	0.045261	0.083427	0.21359
$a_3$	0.00065929	0.001091	0.0013239	0.0045749	0.0051956	0.012995	0.051852	0.40411	0.11287
$a_4$	0.076104	0.076042	0.0013435	0.076267	0.093651	0.013435	0.07014	0.27934	0.1279
$b_0$	0.085887	0.087822	0.007847	0.015376	0.0018082	0.074239	0.035331	-0.33724	0.42591
$b_1$	0.00013013	0.0022964	0.0026463	0.0089322	0.0048752	0.026314	0.12583	0.31019	0.15749
$r_1$	0.48361	0.47112	0.0093066	0.54138	0.47095	0.084174	0.95554	0.1963	0.25558
$r_2$	0.32297	0.31582	0.0050372	0.34567	0.3993	0.047824	0.33342	0.8081	0.18451
$r_3$	0.7839	0.78724	0.0081781	0.82938	0.7672	0.048072	0.43532	0.30476	0.21749

Shown are the mean, mode, and standard deviation of the posterior distribution of the parameters  $\eta_i$  of the SNP transition density (4) that describes the law of motion of the pricing kernel  $\theta_t$ . Before applying these priors, data are normalized by  $y_t = (\log \theta_t + 0.162398)/0.290158$  to conform to the conventions of Table 1. The tight prior uses the scaling of Table 1 divided by ten, the intermediate prior uses the scaling of Table 1, and the loose prior uses the scaling of Table 1 multiplied by ten.

Table 6. The Posterior Probabilities of the Models

Prior	Probability	
	Monthly Data	Annual Data
Tight	2.2e-9	0.0004
Intermediate	0.0025	0.0224
Loose	0.9975	0.9772

Shown are the posterior probabilities of the models under the tight, intermediate, and loose priors computed using Newton and Raftery's (1994)  $\hat{p}_4$  method for computing the marginal likelihood from an MCMC chain and assigning equal prior probability to each model.

Table 7. Regression  $R^2$  for the Monthly Pricing Kernel in a Bansal-Yaron Economy

Lag	Measurement Error				
	0%	1%	2%	5%	10%
Log Pricing Kernel on Log Consumption Growth					
0	0.1616	0.0509	0.0169	0.0032	0.0007
1	0.1622	0.0510	0.0170	0.0036	0.0010
2	0.1623	0.0512	0.0171	0.0036	0.0011
3	0.1626	0.0512	0.0171	0.0036	0.0011
4	0.1627	0.0513	0.0171	0.0038	0.0011
5	0.1629	0.0513	0.0172	0.0038	0.0012
6	0.1629	0.0514	0.0172	0.0038	0.0012
7	0.1629	0.0514	0.0172	0.0038	0.0012
8	0.1631	0.0514	0.0172	0.0038	0.0012
9	0.1631	0.0515	0.0173	0.0039	0.0012
10	0.1637	0.0516	0.0173	0.0039	0.0014
11	0.1639	0.0514	0.0178	0.0039	0.0014
Log Pricing Kernel on Log Consumption Growth and Log Market Return					
0	0.5946	0.5346	0.4417	0.2207	0.0772
1	0.5950	0.5346	0.4418	0.2207	0.0774
2	0.5952	0.5347	0.4418	0.2208	0.0775
3	0.5953	0.5348	0.4419	0.2208	0.0776
4	0.5954	0.5348	0.4419	0.2209	0.0776
5	0.5958	0.5349	0.4420	0.2210	0.0778
6	0.5958	0.5349	0.4420	0.2210	0.0780
7	0.5958	0.5349	0.4420	0.2212	0.0782
8	0.5961	0.5351	0.4420	0.2214	0.0783
9	0.5962	0.5352	0.4420	0.2214	0.0784
10	0.5964	0.5352	0.4421	0.2214	0.0784
11	0.5965	0.5354	0.4424	0.2215	0.0788
Log Pricing Kernel on Log Consumption Growth and Log Wealth Return					
0	1.0000	0.3217	0.1204	0.0193	0.0065
1	1.0000	0.3218	0.1206	0.0197	0.0070
2	1.0000	0.3222	0.1207	0.0198	0.0072
3	1.0000	0.3222	0.1207	0.0199	0.0072
4	1.0000	0.3222	0.1207	0.0201	0.0072
5	1.0000	0.3223	0.1208	0.0201	0.0073
6	1.0000	0.3225	0.1208	0.0201	0.0073
7	1.0000	0.3225	0.1208	0.0201	0.0074
8	1.0000	0.3227	0.1208	0.0202	0.0074
9	1.0000	0.3228	0.1209	0.0203	0.0075
10	1.0000	0.3230	0.1210	0.0204	0.0077
11	1.0000	0.3230	0.1214	0.0204	0.0077

Shown are the  $R^2$  for regressions of the log pricing kernel on log consumption growth, log consumption growth and log stock return, and log consumption growth and log wealth return using Kiku's (2006) simulation of a Bansal-Yaron (2004) economy. Lagged regressions include contemporaneous regressors and lower order lags. Normally distributed measurement error has been added to the regressors with standard deviation as shown; e.g. 1% is  $\sigma = 0.01$ . As these regressions are logarithmic, the measurement error is multiplicative.

Table 8. Regression  $R^2$  for the Annual Pricing Kernel in a Bansal-Yaron Economy

Measurement Error					
Lag	0%	1%	2%	5%	10%
Log Pricing Kernel on Log Consumption Growth					
0	0.2214	0.2041	0.1331	0.0406	0.0095
1	0.2308	0.2092	0.1368	0.0409	0.0103
2	0.2453	0.2168	0.1414	0.0410	0.0109
3	0.2564	0.2267	0.1494	0.0410	0.0110
4	0.2585	0.2306	0.1503	0.0419	0.0110
Log Pricing Kernel on Log Consumption Growth and Log Market Return					
0	0.5931	0.5866	0.5674	0.4922	0.3921
1	0.6036	0.5950	0.5750	0.4949	0.3946
2	0.6081	0.5974	0.5767	0.4953	0.3951
3	0.6176	0.6048	0.5824	0.4954	0.3956
4	0.6194	0.6082	0.5832	0.4965	0.3966
Log Pricing Kernel on Log Consumption Growth and Log Wealth Return					
0	1.0000	0.7768	0.5386	0.2529	0.0631
1	1.0000	0.7794	0.5428	0.2547	0.0641
2	1.0000	0.7822	0.5488	0.2562	0.0648
3	1.0000	0.7849	0.5562	0.2565	0.0648
4	1.0000	0.7871	0.5581	0.2573	0.0651

Shown are the  $R^2$  for regressions of the log pricing kernel on log consumption growth, log consumption growth and log stock return, and log consumption growth and log wealth return using Kiku's (2006) simulation of a Bansal-Yaron (2004) economy. Annual growth rates are obtained from Kiku's monthly simulations by multiplying growth rates in blocks of twelve. The annual pricing kernel is obtained by multiplying the monthly kernel in blocks of twelve. Lagged regressions include contemporaneous regressors and lower order lags. Normally distributed measurement error has been added to the regressors with standard deviation as shown; e.g. 1% is  $\sigma = 0.01$ . As these regressions are logarithmic, the measurement error is multiplicative.

Table 9. Posterior Distribution of Regression  $R^2$  for the Monthly Pricing Kernel

Lag	Tight Prior					Intermediate Prior					Loose Prior				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
Regression of Log Pricing Kernel on Log Consumption Growth															
0	5.9e-6	0.0001	0.0006	0.0017	0.0051	6.6e-6	0.0002	0.0007	0.0019	0.0052	5.7e-6	0.0001	0.0006	0.0016	0.0045
1	0.0001	0.0008	0.0019	0.0038	0.0081	0.0002	0.0008	0.0020	0.0039	0.0085	0.0001	0.0006	0.0015	0.0031	0.0065
2	0.0005	0.0020	0.0040	0.0070	0.0140	0.0007	0.0024	0.0045	0.0077	0.0143	0.0004	0.0016	0.0031	0.0054	0.0103
3	0.0013	0.0034	0.0059	0.0095	0.0162	0.0015	0.0039	0.0067	0.0107	0.0186	0.0011	0.0031	0.0053	0.0085	0.0144
4	0.0020	0.0047	0.0075	0.0114	0.0183	0.0022	0.0050	0.0081	0.0123	0.0204	0.0018	0.0042	0.0067	0.0101	0.0164
5	0.0028	0.0059	0.0091	0.0131	0.0209	0.0030	0.0062	0.0097	0.0141	0.0230	0.0023	0.0051	0.0079	0.0114	0.0178
6	0.0040	0.0078	0.0113	0.0157	0.0241	0.0041	0.0078	0.0116	0.0166	0.0257	0.0035	0.0068	0.0100	0.0139	0.0213
7	0.0048	0.0091	0.0127	0.0173	0.0261	0.0051	0.0092	0.0131	0.0183	0.0276	0.0044	0.0080	0.0115	0.0156	0.0232
8	0.0058	0.0103	0.0141	0.0188	0.0279	0.0061	0.0104	0.0147	0.0198	0.0294	0.0053	0.0091	0.0128	0.0172	0.0250
9	0.0069	0.0116	0.0158	0.0210	0.0310	0.0071	0.0119	0.0162	0.0216	0.0311	0.0061	0.0104	0.0142	0.0188	0.0270
10	0.0079	0.0129	0.0172	0.0225	0.0322	0.0082	0.0132	0.0177	0.0233	0.0329	0.0071	0.0117	0.0157	0.0204	0.0283
11	0.0095	0.0149	0.0197	0.0253	0.0361	0.0096	0.0153	0.0201	0.0258	0.0366	0.0093	0.0150	0.0199	0.0254	0.0353
Regression of Log Pricing Kernel on Log Consumption Growth and Log Market Return															
0	0.0002	0.0013	0.0030	0.0057	0.0112	0.0005	0.0021	0.0043	0.0076	0.0138	0.0005	0.0023	0.0046	0.0079	0.0142
1	0.0016	0.0039	0.0065	0.0101	0.0167	0.0020	0.0047	0.0077	0.0116	0.0194	0.0023	0.0056	0.0091	0.0135	0.0214
2	0.0035	0.0073	0.0112	0.0158	0.0240	0.0042	0.0083	0.0122	0.0175	0.0270	0.0051	0.0102	0.0149	0.0205	0.0295
3	0.0055	0.0100	0.0144	0.0195	0.0288	0.0064	0.0113	0.0160	0.0218	0.0319	0.0074	0.0135	0.0186	0.0243	0.0343
4	0.0076	0.0126	0.0174	0.0229	0.0323	0.0085	0.0139	0.0191	0.0253	0.0358	0.0097	0.0167	0.0225	0.0292	0.0406
5	0.0099	0.0156	0.0208	0.0268	0.0374	0.0107	0.0169	0.0227	0.0292	0.0400	0.0119	0.0190	0.0251	0.0318	0.0436
6	0.0122	0.0187	0.0243	0.0306	0.0417	0.0135	0.0201	0.0262	0.0331	0.0441	0.0146	0.0221	0.0286	0.0359	0.0478
7	0.0146	0.0213	0.0271	0.0339	0.0450	0.0159	0.0231	0.0294	0.0366	0.0482	0.0174	0.0255	0.0323	0.0404	0.0529
8	0.0176	0.0255	0.0322	0.0393	0.0515	0.0197	0.0279	0.0345	0.0425	0.0548	0.0220	0.0305	0.0374	0.0456	0.0588
9	0.0199	0.0285	0.0354	0.0429	0.0558	0.0222	0.0308	0.0376	0.0457	0.0586	0.0245	0.0335	0.0410	0.0494	0.0632
10	0.0246	0.0335	0.0411	0.0494	0.0633	0.0262	0.0355	0.0429	0.0514	0.0651	0.0278	0.0366	0.0443	0.0530	0.0673
11	0.0277	0.0372	0.0455	0.0540	0.0684	0.0291	0.0392	0.0469	0.0560	0.0699	0.0321	0.0421	0.0504	0.0597	0.0748
Regression of Log Pricing Kernel on Log Consumption Growth and Log Wealth Return															
0	0.0679	0.0894	0.1078	0.1262	0.1568	0.0724	0.1070	0.1329	0.1600	0.1990	0.0915	0.1264	0.1554	0.1852	0.2339
1	0.1139	0.1436	0.1662	0.1917	0.2273	0.1058	0.1440	0.1774	0.2112	0.2534	0.1288	0.1714	0.2061	0.2403	0.2869
2	0.1387	0.1698	0.1923	0.2179	0.2526	0.1237	0.1638	0.1967	0.2280	0.2725	0.1600	0.2016	0.2336	0.2674	0.3162
3	0.1552	0.1873	0.2107	0.2356	0.2723	0.1390	0.1810	0.2136	0.2442	0.2885	0.1817	0.2239	0.2559	0.2908	0.3407
4	0.1692	0.2018	0.2245	0.2498	0.2854	0.1519	0.1929	0.2245	0.2551	0.2996	0.1970	0.2387	0.2701	0.3039	0.3521
5	0.1797	0.2131	0.2361	0.2605	0.2965	0.1654	0.2058	0.2382	0.2662	0.3120	0.2075	0.2510	0.2840	0.3177	0.3680
6	0.1896	0.2229	0.2458	0.2706	0.3061	0.1728	0.2125	0.2455	0.2749	0.3205	0.2149	0.2617	0.2944	0.3273	0.3773
7	0.1991	0.2322	0.2552	0.2791	0.3140	0.1790	0.2213	0.2535	0.2845	0.3286	0.2271	0.2721	0.3036	0.3356	0.3857
8	0.2084	0.2407	0.2637	0.2885	0.3241	0.1877	0.2299	0.2619	0.2925	0.3388	0.2354	0.2810	0.3126	0.3452	0.3944
9	0.2138	0.2469	0.2704	0.2954	0.3314	0.1925	0.2356	0.2681	0.2986	0.3457	0.2415	0.2875	0.3208	0.3543	0.4040
10	0.2184	0.2515	0.2760	0.3008	0.3374	0.1978	0.2409	0.2727	0.3038	0.3507	0.2450	0.2924	0.3256	0.3598	0.4092
11	0.2241	0.2579	0.2816	0.3071	0.3435	0.2040	0.2475	0.2788	0.3088	0.3558	0.2512	0.2990	0.3321	0.3664	0.4163

Shown are quantiles of the posterior distribution of  $R^2$  for regressions of the log pricing kernel on lags of log consumption growth, log consumption growth and log stock return, and log consumption growth and log wealth return under the tight, intermediate, and loose priors.

Table 10. Posterior Distribution of Regression  $R^2$  for the Annual Pricing Kernel

Lag	Tight Prior					Intermediate Prior					Loose Prior				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
Regression of Log Pricing Kernel on Log Consumption Growth															
0	0.0342	0.0651	0.0915	0.1211	0.1701	0.0211	0.0480	0.0731	0.1021	0.1510	0.0168	0.0405	0.0650	0.0928	0.1425
1	0.0412	0.0727	0.1002	0.1295	0.1798	0.0287	0.0571	0.0829	0.1118	0.1624	0.0276	0.0518	0.0757	0.1028	0.1500
2	0.0500	0.0848	0.1139	0.1445	0.1940	0.0381	0.0697	0.0976	0.1285	0.1800	0.0373	0.0660	0.0916	0.1218	0.1699
3	0.0961	0.1424	0.1784	0.2184	0.2815	0.0750	0.1177	0.1537	0.1923	0.2552	0.0709	0.1102	0.1443	0.1837	0.2472
4	0.1012	0.1479	0.1864	0.2272	0.2920	0.0810	0.1248	0.1611	0.2011	0.2642	0.0768	0.1184	0.1532	0.1956	0.2603
Regression of Log Pricing Kernel on Log Consumption Growth and Log Market Return															
0	0.0920	0.1373	0.1690	0.2057	0.2601	0.0724	0.1140	0.1476	0.1838	0.2411	0.0710	0.1111	0.1426	0.1783	0.2323
1	0.1158	0.1613	0.1959	0.2345	0.2905	0.0904	0.1348	0.1693	0.2063	0.2664	0.0848	0.1243	0.1569	0.1951	0.2526
2	0.1368	0.1839	0.2215	0.2594	0.3181	0.1088	0.1566	0.1940	0.2326	0.2944	0.0999	0.1432	0.1775	0.2166	0.2764
3	0.1808	0.2355	0.2771	0.3189	0.3843	0.1471	0.2002	0.2406	0.2847	0.3532	0.1288	0.1800	0.2182	0.2617	0.3331
4	0.1911	0.2460	0.2876	0.3317	0.3980	0.1586	0.2129	0.2541	0.2983	0.3680	0.1402	0.1915	0.2313	0.2740	0.3461
Regression of Log Pricing Kernel on Log Consumption Growth and Log Wealth Return															
0	0.2133	0.2678	0.3104	0.3549	0.4190	0.1482	0.2015	0.2419	0.2823	0.3430	0.1436	0.2062	0.2524	0.3031	0.3837
1	0.2851	0.3482	0.3937	0.4402	0.5092	0.2339	0.2961	0.3396	0.3857	0.4570	0.2144	0.2825	0.3329	0.3877	0.4698
2	0.3466	0.4098	0.4568	0.5022	0.5692	0.3033	0.3659	0.4119	0.4574	0.5237	0.2752	0.3459	0.3962	0.4492	0.5262
3	0.3758	0.4398	0.4860	0.5308	0.5960	0.3394	0.4049	0.4520	0.4976	0.5618	0.3222	0.3933	0.4414	0.4929	0.5660
4	0.4058	0.4754	0.5217	0.5700	0.6346	0.3721	0.4403	0.4884	0.5349	0.6018	0.3576	0.4319	0.4838	0.5365	0.6053

Shown are quantiles of the posterior distribution of  $R^2$  for regressions of the log pricing kernel on lags of log consumption growth, log consumption growth and log stock returns, and log consumption growth and log wealth return under the tight, intermediate, and loose priors.

Table 11. Posterior Probabilities of Regression  $R^2$  for the Monthly Pricing Kernel

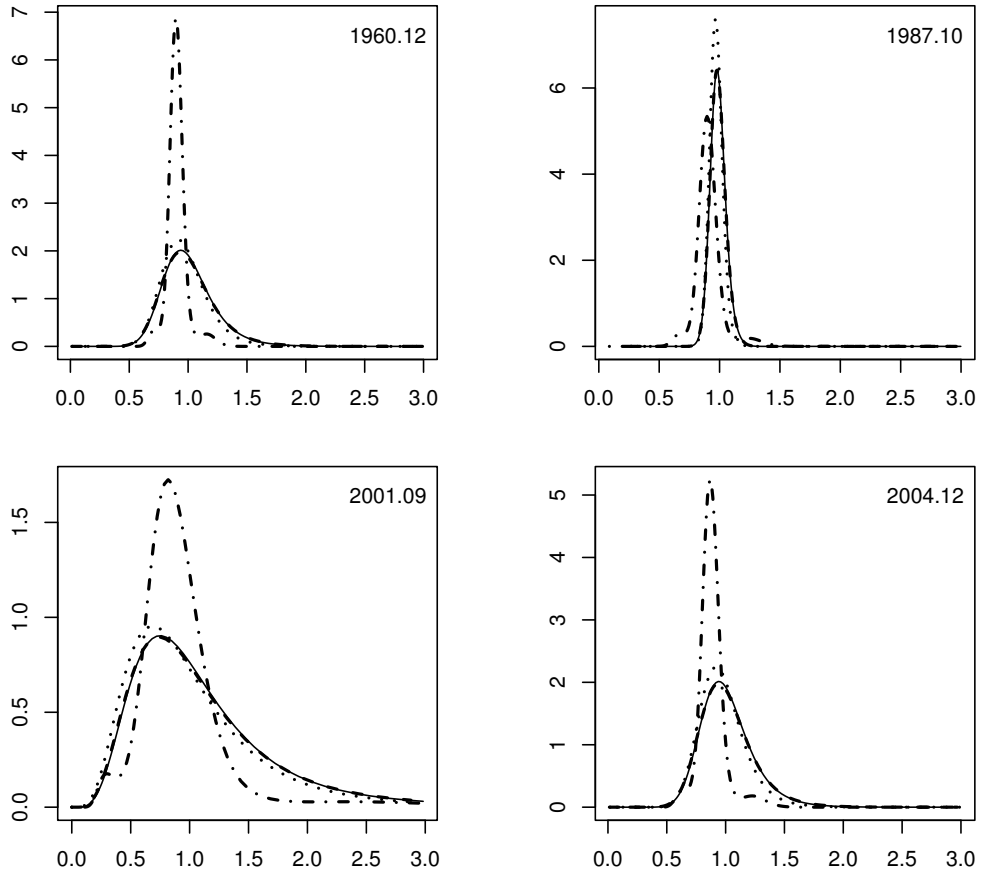
Prior	Measurement Error			
	1%	2%	5%	10%
Log Pricing Kernel on Log Consumption Growth				
Tight	<0.01	0.60	0.99	0.99
Intermediate	<0.01	0.60	0.99	0.99
Loose	<0.01	0.60	0.99	0.99
Log Pricing Kernel on Log Consumption Growth and Log Market Return				
Tight	<0.01	<0.01	<0.01	<0.01
Intermediate	<0.01	<0.01	<0.01	<0.01
Loose	<0.01	<0.01	<0.01	<0.01
Log Pricing Kernel on Log Consumption Growth and Log Wealth Return				
Tight	0.15	0.99	0.99	0.99
Intermediate	0.20	0.99	0.99	0.99
Loose	0.55	0.99	0.99	0.99

Shown are the posterior probabilities of  $R^2$  for regressions of the log pricing kernel on log consumption growth, log consumption growth and log stock return, and log consumption growth and log wealth return under the tight, intermediate, and loose priors.

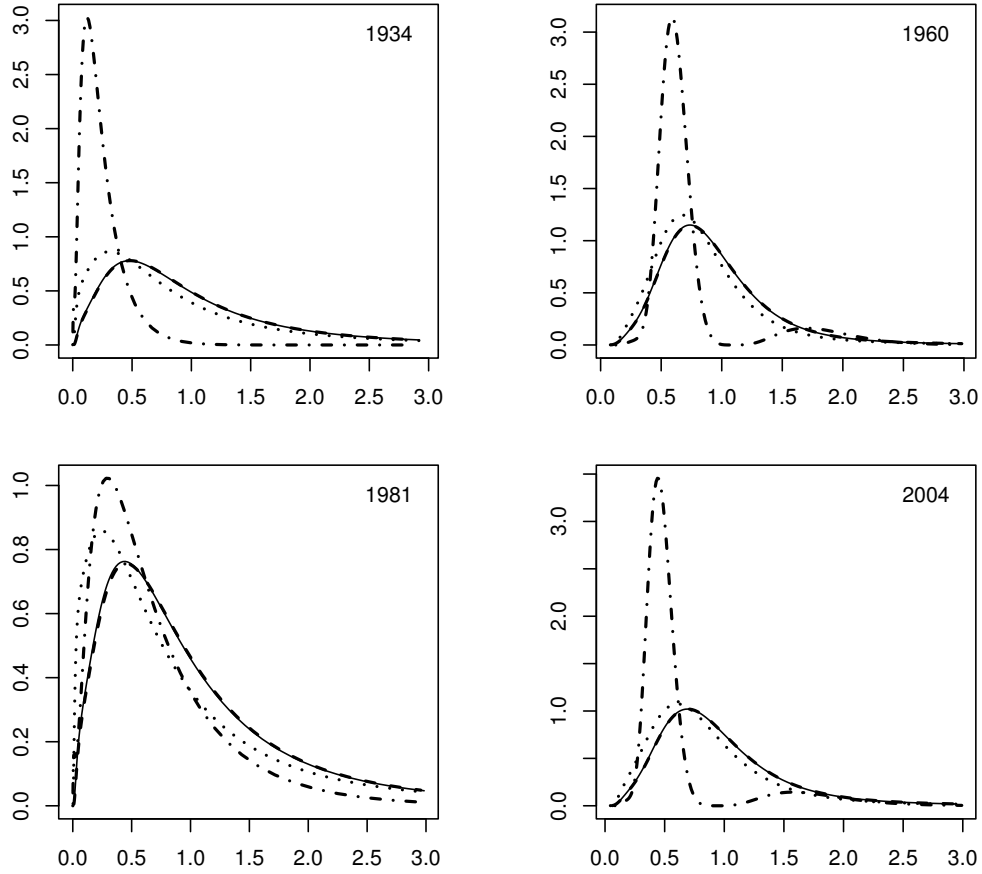
Table 12. Posterior Probabilities of Regression  $R^2$  for the Annual Pricing Kernel

Prior	Measurement Error			
	1%	2%	5%	10%
Log Pricing Kernel on Log Consumption Growth				
Tight	<0.01	0.20	0.90	0.99
Intermediate	<0.01	0.10	0.85	0.95
Loose	<0.01	0.10	0.75	0.95
Log Pricing Kernel on Log Consumption Growth and Log Market Return				
Tight	<0.01	<0.01	<0.01	<0.01
Intermediate	<0.01	<0.01	<0.01	<0.01
Loose	<0.01	<0.01	<0.01	<0.01
Log Pricing Kernel on Log Consumption Growth and Log Wealth Return				
Tight	<0.01	<0.01	0.80	0.99
Intermediate	<0.01	<0.01	0.40	0.99
Loose	<0.01	<0.01	0.50	0.90
Log Pricing Kernel on Log Consumption Growth and Log Wealth Return Years 1960–2004 only				
Tight	<0.01	0.60	0.99	0.99
Intermediate	<0.01	0.25	0.99	0.99
Loose	<0.01	0.20	0.99	0.99

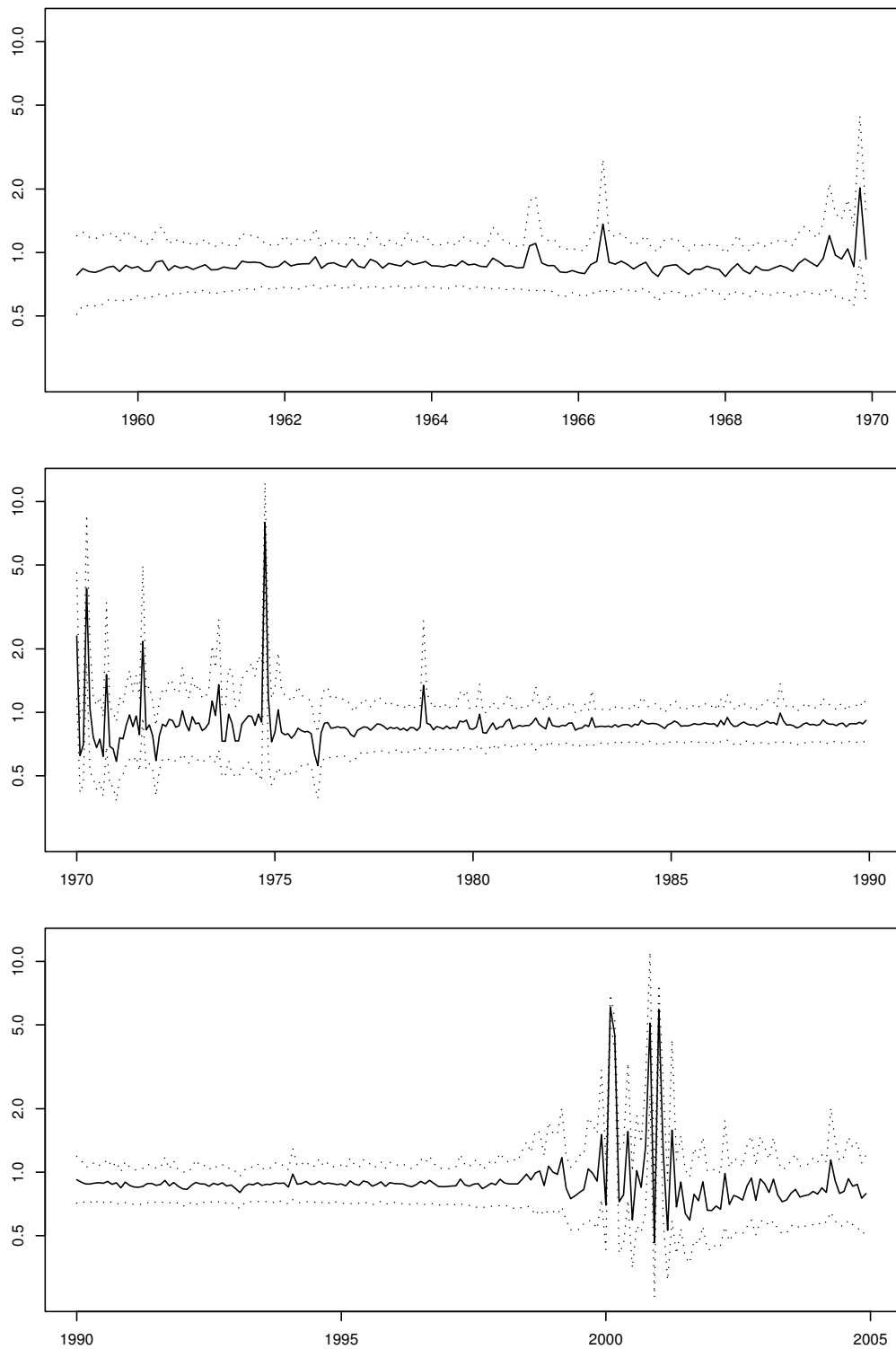
Shown are the posterior probabilities of  $R^2$  for regressions of the log pricing kernel on log consumption growth, log consumption growth and log stock return, and log consumption growth and log wealth return under the tight, intermediate, and loose priors.



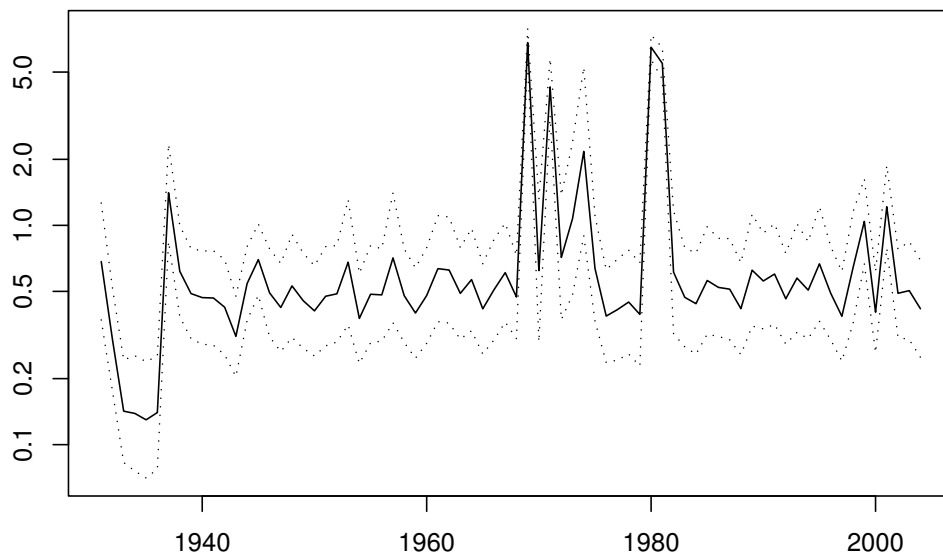
**Figure 1. The Law of Motion of the Monthly Pricing Kernel.** Plotted is the conditional density  $f(\theta_{t+1}|\theta_t, \dots, \theta_1|\eta)$  of the pricing kernel given by (4) conditional on  $\theta_t$  up to the year and month shown in the upper right of each panel. The solid line shows the density with  $\eta$  set to the location parameter of the prior in Table 1. The dashed line has  $\eta$  set to the mode of the posterior distribution under the tight prior. The dotted line has  $\eta$  set to the mode of the posterior distribution under the intermediate prior. The dotted-dashed line has  $\eta$  set to the mode of the posterior distribution under the loose prior. The conditioning set is the posterior mean of  $\theta_t$  under the tight prior in each case.



**Figure 2. The Law of Motion of the Annual Pricing Kernel.** Plotted is the conditional density  $f(\theta_{t+1}|\theta_t, \dots, \theta_1|\eta)$  of the pricing kernel given by (4) conditional on  $\theta_t$  up to the year shown in the upper right of each panel. The solid line shows the density with  $\eta$  set to the location parameter of the prior in Table 1. The dashed line has  $\eta$  set to the mode of the posterior distribution under the tight prior. The dotted line has  $\eta$  set to the mode of the posterior distribution under the intermediate prior. The dotted-dashed line has  $\eta$  set to the mode of the posterior distribution under the loose prior. The conditioning set is the posterior mean of  $\theta_t$  under the tight prior in each case.



**Figure 3. The Posterior Mean of the Monthly Pricing Kernel.** Plotted as the solid line is the posterior mean of  $\log(\theta_2), \dots, \log(\theta_{551})$  under the loose prior. The dotted lines are plus and minus one standard deviation. The units of the vertical axis are the exponential of the plotted quantity.



**Figure 4. The Posterior Mean of the Annual Pricing Kernel.** Plotted as the solid line is the posterior mean of  $\log(\theta_2), \dots, \log(\theta_{75})$  under the loose prior. The dotted lines are plus and minus one standard deviation. The units of the vertical axis are the exponential of the plotted quantity.