Habit, Long Run Risks, Prospect? A Statistical Inquiry.*

Eric M. Aldrich
Duke University
Department of Economics
Durham NC 27708-0097 USA

A. Ronald Gallant
Duke University
Fuqua School of Business
Durham NC 27708-0120 USA

First draft: June 2009
This draft: September 2009

Abstract

We use Bayesian statistical methods to compare the habit persistence asset pricing model of Campbell and Cochrane, the long run risks model of Bansal and Yaron, and the prospect theory model of Barberis, Huang, and Santos. We undertake two types of comparisons, relative and absolute, over two sample periods, 1930–2008 and 1950–2008, using two series, univariate U.S. stock returns and bivariate U.S. consumption growth and stock returns. The prior for each model is that the real interest rate be within 1% of 0.896 with probability 0.95 together with a preference for model parameters that are near their published values. For the univariate series and for both sample periods, the models perform about the same in the relative comparison and fit the data reasonably well in the absolute assessment. For the bivariate series, in the relative comparison the long run risks model overwhelmingly dominates over the 1930–2008 period, while the habit persistence model overwhelmingly dominates over the 1950–2008 period; in the absolute assessment, the habit model fails definitively in the 1930–2008 period and the prospect theory model fails definitively in the 1950–2008 period. Out-of-sample, the models show interesting differences in their forecasts over the 2009–2013 horizon. In-sample, they differ mainly in their ability to track the conditional volatility of consumption growth and the conditional correlation between consumption growth and stock returns.

Keywords and Phrases: Statistical Tests, Habit, Long Run Risks, Prospect Theory, Asset Pricing.
JEL Classification: E00, G12, C51, C52

*Corresponding author: A. Ronald Gallant, Duke University, Fuqua School of Business, DUMC 90120, Durham NC 27708-0120, USA; phone 919-660-7700; email aronldg@gmail.com. Most recent version at www.duke.edu/∼arg. Supported by National Science Foundation Grant Number SES 0438174.

1 Introduction

The goal of this paper is to fill a void in the literature. There are, to our knowledge, no head-to-head comparisons of asset pricing models from macro/finance that adhere to the principles of statistical science. This paper fills the void.

The reasons for this void are twofold: The first is the attack on the use of statistics in macro/finance by Edward Prescott and his followers; see Kydland and Prescott (1996) and the references therein. That attack has deterred inquiries that adhere to statistical principles. Contemporary macro economics, particularly the real business cycle literature, is a vast statistical wasteland. The second is that practicable statistical methods to compare non-nested models whose likelihood is not directly available on sparse data have only recently become available.

It is a matter of current debate as to whether or not the principles of statistics are relevant to economic research; see Hansen and Heckman (1996) and the references therein. We have nothing to say about that debate here. We do maintain, however, that the existence of a debate is no reason to suppress facts. An informed debate is better than a debate based on surmise. In this paper, we provide facts that relate to consumption-based asset pricing models. The methodology that we introduce is generally applicable and can be applied to other macro/finance models.

The asset pricing models considered are the habit persistence model of Campbell and Cochrane (1999), the long run risks model of Bansal and Yaron (2004), and the prospect theory model of Barberis, Huang, and Santos (2001). There are two reason for this choice: These three models are arguably the leading contenders; And the authors have provided descriptions of their computational methods that are sufficiently detailed to allow replication of their simulations.

We know of only one other study that attempts a head-to-head statistical comparison of asset pricing models: Bansal, Gallant, and Tauchen (2007). It compared the habit model to the long run risks model using frequentist methods. The statistical methods employed could not distinguish between these two models because frequentist non-nested model comparison methods require abundant data. Abundant data are not available in macro/finance. The

typical sampling frequency used to calibrate and assess macro/finance models is annual and there are only about 80 annual observations available on the U.S. economy. The three papers cited above use annual data. Barberis, Huang, and Santos (2001) insist that annual is the only frequency that is appropriate to their model. Using higher frequency data to compare these three models is not an option. They were not designed to explain high frequency data.

Failing to achieve a definitive statistical result, Bansal, Gallant, and Tauchen (2007) proceeded to compare the models using the more traditional methods of macro/finance which consist of enumerating some moment conditions and checking model simulations against them either formally by means of GMM or informally by means of calibration (Kydland and Prescott, 1996). On the basis of a battery of such tests, Bansal, Gallant, and Tauchen conclude that the long run risks model is preferred.

In addition to the fact that their comparison, in the end, was not statistical, there are other concerns. Bansal, Gallant, and Tauchen did not actually compare the models proposed by Campbell and Cochrane (1999) and Bansal and Yaron (2004). They modified them to impose cointegration on macro variables that ought not diverge. They also used a general purpose method to solve them; specifically, a Bubnov-Galerkin method (Miranda and Fackler, 2002, p 152–3). In our experience, simulations from models of the sort considered here are sensitive to the method used to solve them. Our view is that fairness dictates that one use both the same model that was proposed and the same solution method. To state our view succinctly, the structural model is the simulation algorithm proposed by the originator; it is not the mathematical equations that suggested the algorithm.

Lastly, Bansal, Gallant, and Tauchen used a dividend series and we do not. Although both the frequentist methods that they used and the Bayesian methods used here allow dividends to be latent in principle, the sparseness of the data compelled Bansal, Gallant, and Tauchen to use dividends. We do not use dividend data because, aside from the fact that it is difficult to properly adjust dividend payouts for stock repurchases and other distortions caused by tax policy, we want to focus solely on asset prices and consumption. Whether or not the models we consider can explain dividend payouts is of little interest. In fact, the long run risks model, which would seem to have the best chance of explaining dividends, apparently cannot (Gallant and Hong, 2007).

The data we use are annual, per capita, real, U.S., consumption growth and stock returns from 1925–2008. The comparisons are for the period 1930–2008 and 1950–2008. The data from 1925–1929 are only used to prime recursions because they are of lower quality than the data from 1930 onwards. The data are plotted in Figure 1. Note in the figure that consumption growth is far more volatile in the 1930–1949 period than in the period from 1950–2008. The volatility of stock returns is not much different. It turns out that the difference in consumption growth volatility dramatically influences results.

Figure 1 about here

Gallant and McCulloch (2009) introduced a Bayesian method for fitting a structural model for which a likelihood is not readily available to sparse data such as that shown in Figure 1. They synthesize a likelihood by means of an auxiliary model and simulation from the structural model along lines that are similar to indirect inference (Smith, 1993; Gourieroux, Monfort, and Renault, 1993) and efficient method of moments (Gallant and Tauchen, 1996). Dejong, Ingram, and Whiteman (2000) and Del Negro and Schorfheide (2004) are closely related to Gallant and McCulloch (2009) and use similar ideas. What is new in Gallant and McCulloch are the computational methods that allow extension of the ideas to highly nonlinear structural and auxiliary models. If Gallant and McCulloch's assumptions are satisfied, the synthesized likelihood is identical to the likelihood of the structural model, were it available. The Bayesian paradigm allows one to use prior information to compensate for data sparseness. Model comparison is by means of posterior probabilities. In short, the methodology used here is classical Bayesian statistics. It is only the computational methods that are not standard.

In the Gallant and McCulloch framework, the auxiliary model must encompass the structural model for the methodology to be logically correct. This is a departure from the indirect inference and efficient method of moments literatures where this requirement is not logically necessary. It may or may not be desirable. That is a subject of debate (see Gallant and Tauchen (2009) and the references therein). In this debate the auxiliary model that encompasses the structural model is presumed to be simpler and easier to fit to simulations from the structural model than is an auxiliary model that fits the data. That is not true here.

An auxiliary model that can encompasses the habit persistence, long run risks, and prospect theory models is far more complex than time series models customarily fit to the data in Figure 1. Bizarre might be a better word than complex. In view of the fact that theory flies in the face of common sense, we conduct a sensitivity analysis employing six auxiliary models of differing complexity to determine if the choice of auxiliary model makes a difference to our results. The six models are shown in Table 1. Model f_1 is closest to that used by Gallant and McCulloch; f_5 is the encompassing model.

Table 1 about here

Auxiliary models f_2 through f_5 are more complex than the models considered by Gallant and McCulloch (2009). We find that the computational methods that they proposed are not sufficiently accurate for models of this complexity. A contribution of this paper is a refinement of their methods that increases accuracy to the point that auxiliary models as complex as f_2 through f_5 can be used in applications.

Fairness requires that one use the model that was proposed together with the proposed solution method, as noted above. Fairness also requires evenhandedness with regard to the prior. Our prior for each model is that the real, risk-free, interest rate be within 1% of 0.896 with probability 0.95 together with a preference for model parameters that are near their published values. Campbell (2003) notes that any reasonable asset pricing model must incorporate the indirect evidence that the risk-free rate is very low with low volatility. Campbell's evidence suggests that the mean risk-free rate for the U.S. is 0.896 percent per annum. Bansal, Gallant, and Tauchen (2007) argue that imposing the risk free rate *a priori* is likely to produce better estimates than using an *ex ante* risk free rate series that is mostly noise due to the difficulty of determining *ex ante* inflation (Mishkin, 1981). This prior appears to strike the right balance. It is tight enough to achieve MCMC chains that mix well despite the use of sparse data but loose enough to allow the data to be influential with regard to the equity premium, the standard deviation of equity returns, and the conditional dynamics of consumption growth and equity returns.

2 Models Considered

In this section we describe the habit persistence model, the long run risks model, and the prospect theory model with a focus on the algorithms used to solve them. The algorithms for the habit persistence model are described in Campbell and Cochrane (1999) and in supplemental materials on John Cochrane's website, particularly an online appendix and Gauss code. The algorithms for the long run risks model are described in Bansal and Yaron (2004); useful supplements are Kiku (2006) and Bansal, Kiku, and Yaron (2006). The algorithms for the prospect theory model are described in Barberis, Huang, and Santos (2001).

2.1 The Habit Persistence Model

The driving processes for the habit persistence model are

$$\text{Consumption:}\quad c_t - c_{t-1} = g + v_t$$

$$\text{Dividends:}\quad d_t - d_{t-1} = g + w_t$$

$$\text{Random shocks:}\quad \begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim \text{NID}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\sigma_w \\ \rho\sigma\sigma_w & \sigma_w^2 \end{pmatrix}\right]$$

The time increment is one month. Lower case denotes the logarithm of an upper case quantity; i.e. $c_t = \log(C_t)$, $d_t = \log(D_t)$. The utility function is

$$\mathcal{E}_0\left(\sum_{t=0}^{\infty} \delta^t \, \frac{(S_t C_t)^{1-\gamma} - 1}{1-\gamma}\right), \tag{1}$$

where habit persistence is implemented by two equations:

$$\text{Surplus ratio:}\quad s_t - \bar{s} = \phi\,(s_{t-1} - \bar{s}) + \lambda(s_{t-1}) v_{t-1} \tag{2}$$

$$\text{Sensitivity function:}\quad \lambda(s) = \begin{cases} \frac{1}{\bar{S}}\sqrt{1 - 2(s-\bar{s})} - 1 & s \le s_{\max} \\ 0 & s > s_{\max} \end{cases} \tag{3}$$

$\mathcal{E}_t$ is conditional expectation with respect to S_t, which is the state variable; $s_t = \log(S_t)$. The quantities $\bar{S}$ and $s_{\max}$ can be computed from model parameters $\theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma)$

as $\bar{S} = \sigma\sqrt{\gamma/(1-\phi)}$ and $s_{\max} = \bar{s} + (1-\bar{S}^2)/2$. If one substitutes $S_t C_t = C_t - X_t$ in (1), where X_t is external habit, one obtains the habit persistence utility function as it is usually written. The form above is more convenient for computations.

Let V be the map from the state S to the price-dividend ratio P/D of the equity asset. It is determined by the Euler equation

$$V(S_t) = \mathcal{E}_t\left\{\delta\left(\frac{S_{t+1}C_{t+1}}{S_t C_t}\right)^{-\gamma}\left(\frac{D_{t+1}}{D_t}\right)[1+V(S_{t+1})]\right\}. \tag{4}$$

The logarithmic return on the equity asset, $r_{dt} = \log(P_{dt}+D_t) - \log(P_{d,t-1})$, is obtained from $V(S)$ using

$$r_{dt} = \log\left[\frac{1+V(S_t)}{V(S_{t-1})}\left(\frac{D_t}{D_{t-1}}\right)\right]. \tag{5}$$

In (4), the dividend shock can be integrated out analytically leaving an expression in the consumption shock. The consumption shock can be integrated numerically using Gauss-Hermite quadrature. We used a five point rule, which integrates polynomials up to degree nine exactly.

Having a means to compute the integral, one can solve (4) by approximating the log policy function

$$v(s_t) = \log V(e^{s_t}) \tag{6}$$

by a piecewise linear function $\hat{v}(s)$. Campbell and Cochrane set the join points $\{s_i\}_{i=1}^I$ of $\hat{v}(s)$ at $\bar{s}$, $s_{\max}$, $s_{\max} - 0.01$, $s_{\max} - 0.02$, $s_{\max} - 0.03$, $s_{\max} - 0.04$, and $\log(kS_{\max}/11)$ for $k = 1, \ldots, 10$. We added the abscissae of the Gauss-Hermite quadrature formula at the maximum and minimum of the above join points then deleted all points less than 0.001 apart.

To solve (4) for $V(S) = \exp[v(\log S)]$, start at guesses $\{\hat{v}_i^{(0)}\}_{i=1}^I$ for the ordinates $\hat{v}_i = \hat{v}(s_i)$ at each of the join points s_i, e.g, scale $\hat{v}_i^{(0)}$ off Figure 3 of Campbell and Cochrane (1999). Substitute the piecewise linear approximation $\hat{v}^{(0)}(s)$ determined by $\{\hat{v}_i^{(0)}, s_i\}_{i=1}^I$ into the right hand side of (4) and compute new ordinates $\{\hat{v}_i^{(1)}\}_{i=1}^I$. Repeat until the $\hat{v}_i^{(j)}$ converge. Substitute $\hat{v}_i^{(J)}(s)$ determined by the converged values $\{\hat{v}_i^{(J)}, s_i\}$ in (5) to compute returns.

The logarithmic return r_{ft} on an asset that pays one dollar one month hence with certainty is given by

$$r_{ft} = -\log\left\{\mathcal{E}_t\left[\delta\left(\frac{S_{t+1}C_{t+1}}{S_t C_t}\right)^{-\gamma}\right]\right\}$$

The integral involves only the consumption shock and may be computed by Gauss-Hermite quadrature.

Given the habit model's parameters

$$\theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma), \tag{7}$$

$\{C_t, r_{dt}, r_{ft}\}_{t=1}^{12N}$ are simulated at the monthly frequency and aggregated to the annual frequency

$$\begin{aligned} C_t^a &= \sum_{k=0}^{11} C_{12t-k}, & (8)\\ c_t^a &= \log(C_t^a), & (9)\\ r_{dt}^a &= \sum_{k=0}^{11} r_{d,12t-k}, & (10)\\ r_{ft}^a &= \sum_{k=0}^{11} r_{f,12t-k}, & (11) \end{aligned}$$

where N is the annual simulation size.

Our prior is

$$\pi(\theta) = \mathrm{N}\left[r_f \,|\, 0.896, \left(\frac{1}{1.96}\right)^2\right] \prod_{i=1}^{p} \mathrm{N}\left[\theta_i \,|\, \theta_i^*, \left(\frac{0.1\theta_i^*}{1.96}\right)^2\right] \tag{12}$$

where the θ_i^* are the calibrated values from Campbell and Cochrane (1999). The scale factor actually used for ϕ and δ is 0.001 rather than the 0.1 shown in (12) to overcome an identification problem. The MCMC chain will not mix when the scale factor for ϕ and δ in (12) is 0.1 because a move in ϕ can be nearly exactly offset by a move in δ. The value 0.001 is the largest value for which the MCMC chain that draws from the prior only will mix. This is not an independence prior as seen from the correlations in Table 2. Measures of location and scale for the prior and posterior distributions are shown in Table 3. The prior and posterior densities of the risk free rate, equity premium, equity returns, and the standard deviation of equity returns are shown in Figure 2. Overall, Table 3 and Figure 2 suggest that the prior is sufficiently informative to fill in where data are sparse but it allows the data to move the posterior where data are informative. The information content of the data is most apparent in Figure 2.

Table 2 about here

Table 3 about here

Figure 2 about here

Where differences in the three models are the most obvious visually is in their out-of-sample forecasts for the next five years. The mean posterior forecast for the habit persistence model, computed as described in Subsection 3.6, is shown in in Figure 3. The habit model predicts an end to the current recession in 2009 and return to steady-state growth by 2010. Stock returns are predicted to be high in 2009 with a return to steady-state returns by 2013.

Figure 3 about here

2.2 The Long Run Risks Model

The driving processes for the long run risks model are

$$\begin{aligned}
\text{Consumption:} \quad & c_{t+1} - c_t = \mu_c + x_t + \sigma_t \eta_{t+1} \\
\text{Long Run Risks:} \quad & x_{t+1} = \rho x_t + \phi_e \sigma_t e_{t+1} \\
\text{Stochastic Volatility:} \quad & \sigma_{t+1}^2 = \bar{\sigma}^2 + \nu(\sigma_t^2 - \bar{\sigma}^2) + \sigma_w w_{t+1} \\
\text{Dividends:} \quad & d_{t+1} - d_t = \mu_d + \phi_d x_t + \pi_d \sigma_t \eta_{t+1} + \phi_u \sigma_t u_{t+1} \\
\text{Random Shocks:} \quad & \begin{pmatrix} \eta_t \\ e_t \\ w_t \\ u_t \end{pmatrix} \sim \text{NID} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right]
\end{aligned}$$

The time increment is one month. Lower case denotes the logarithm of an upper case quantity; e.g. $c_t = \log(C_t)$, $d_t = \log(D_t)$. The notational conventions are from Bansal, Kiku, and Yaron (2007), which differ slightly from Bansal and Yaron (2004). The conventions above are more convenient for computation.

The long run risks model derives its name from the random shifts in the location of consumption and dividends due to x_t. It also incorporates stochastic volatility which should

help with respect to tracking the conditional dynamics that our relative comparisons and absolute assessments take into account.

The long run risks model uses the Epstein-Zin (1989) utility function

$$U_t = \left[(1-\delta)C_t^{\frac{\psi-1}{\psi}} + \delta\left(\mathcal{E}_t\, U_{t+1}^{1-\gamma}\right)^{\frac{\psi-1}{\psi(1-\gamma)}}\right]^{\frac{\psi}{\psi-1}}. \tag{13}$$

γ is the coefficient of risk aversion and ψ is the elasticity of inter temporal substitution. $\mathcal{E}_t$ is the conditional expectation with respect to x_t and σ_t, which are the state variables. Precursors of (13) are Kreps and Porteus (1978) and Weil (1989).

The long run risks model is richly parametrized

$$\theta = (\delta, \gamma, \psi, \mu_c, \rho, \phi_e, \bar{\sigma}^2, \nu, \sigma_w, \mu_d, \phi_d, \pi_d, \phi_u). \tag{14}$$

It is so richly parametrized that identification would have to come from the prior even were data abundant because most of the auxiliary models that we use have fewer parameters than θ. The two autoregressive parameters ρ and η cause additional problems. The solution method proposed by Bansal and Yaron (2004) degrades as ρ and η deviate from their published values. (The published values are approximately the same as the first column of Table 5.) As the degradation is continuous in ρ and η there is no logical threshold that one can impose on ρ and η to completely prevent degradation. In our prior, described below, we try to strike a reasonable balance that prevents likelihoods computed in our MCMC chains from having occasional absurdly small values but does permit some occasional extremely small values. In a sense, Bansal and Yaron are being punished for their choice of solution method, which is based on log-linear approximations. A log-linear approximation is, in the end, a Taylor's expansion. As is well known, Taylor's expansions have a limited radius of validity and that shows up in our work.

The marginal rate of substitution,

$$\text{mrs}_{t+1} = \delta^{\frac{1-\gamma}{1-1/\psi}} \exp\left[-\left(\frac{1-\gamma}{\psi-1}\right)(c_{t+1}-c_t) + \left(\frac{1-\gamma}{1-1/\psi}-1\right) r_{c,t+1}\right], \tag{15}$$

depends on the return $r_{ct} = \log(P_{ct} + C_t) - \log(P_{c,t-1})$, where P_{ct} is the price of the asset that pays the consumption stream. Let V_C be the map from the state (x, σ) to the price-consumption ratio P_c/C . It is determined by the Euler equation

$$V_C(x_t, \sigma_t) = \mathcal{E}_t\left\{\text{mrs}_{t+1}\left(\frac{C_{t+1}}{C_t}\right)\left[1 + V_C(x_{t+1}, \sigma_{t+1})\right]\right\}. \tag{16}$$

r_{ct} is defined by

$$r_{ct} = \log\left[\frac{1 + V_C(x_t, \sigma_t)}{V_C(x_{t-1}, \sigma_{t-1})}\left(\frac{C_t}{C_{t-1}}\right)\right].$$

To compute r_{ct}, Bansal and Yaron (2004) use the log linear approximation

$$\begin{aligned} r_{c,t+1} &\doteq \kappa_0 + \kappa_1 z_{t+1} + c_{t+1} - c_t - z_t \\ \kappa_1 &= [\exp(\bar{z})]/[1 + \exp(\bar{z})] \\ k_0 &= \log[1 + \exp(\bar{z})] - \kappa_1 \bar{z} \end{aligned}$$

where $z_t = \log(P_{c,t}/C_t)$ and $\bar{z}$ is its endogenous mean. To compute z_t, use the approximation

$$z_t \doteq A_0(\bar{z}) + A_1(\bar{z})\, x_t + A_2(\bar{z})\, \sigma_t^2, \tag{17}$$

where the $A_i(\bar{z})$ are tedious expressions in model parameters and $\bar{z}$ given in the Appendix to Bansal and Yaron (2004). (See also Bansal, Kiku, and Yaron (2006).) To compute $\bar{z}$, solve the fixed point problem

$$\bar{z} = A_0(\bar{z}) + A_2(\bar{z})\, \bar{\sigma}_t^2.$$

The return to equities, $r_{dt} = \log(P_{dt} + D_t) - \log(P_{d,t-1})$, is computed similarly. The Euler equation is

$$V_D(x_t, \sigma_t) = \mathcal{E}_t\left\{\mathrm{mrs}_{t+1}\left(\frac{D_{t+1}}{D_t}\right)[1 + V_D(x_{t+1}, \sigma_{t+1})]\right\}$$

and the logarithmic return is

$$r_{dt} = \log\left[\frac{1 + V_D(x_t, \sigma_t)}{V_D(x_{t-1}, \sigma_{t-1})}\left(\frac{D_t}{D_{t-1}}\right)\right].$$

Again similarly, one can compute the logarithmic risk free rate $r_{ft} = -\log \mathcal{E}_t\left(\mathrm{mrs}_{t+1}\right)$ using expressions given in Bansal and Yaron.

As with the habit model, the long run risks model is simulated at the monthly frequency and aggregated to the annual using expressions (8) through (11).

Our prior distribution is

$$\pi(\theta) = \mathrm{N}\left[r_f \,|\, 0.896, \left(\frac{1}{1.96}\right)^2\right] \prod_{i=1}^{p} \mathrm{N}\left[\theta_i \,|\, \theta_i^*, \left(\frac{0.1\theta_i^*}{1.96}\right)^2\right] \tag{18}$$

where the θ_i^* are the calibrated values from Kiku (2006). For the reasons discussed above, the scale factor for ρ and ν actually used is 0.01 rather than 0.1.

Prior correlations are shown in Table 4. Measures of location and scale for the prior and posterior distributions are shown in Table 5. The prior and posterior densities of the risk free rate, equity premium, equity returns, and the standard deviation of equity returns are shown in Figure 4. As for the habit model, Table 5 and Figure 4 suggest that the prior is sufficiently informative to fill in where data are sparse but it allows the data to move the posterior where data are informative.

Table 4 about here

Table 5 about here

Figure 4 about here

The mean posterior forecast for the long run risks model is shown in in Figure 5. The the long run risks model predicts an end to the current recession in 2010 and slow increase in the growth rate thereafter. Stock returns are predicted to be approximately at their steady-state values over the entire forecast period.

Figure 5 about here

2.3 The Prospect Theory Model

The driving processes for the prospect theory model are

$$\text{Aggregate Consumption: } \bar{c}_{t+1} - \bar{c}_t = g_C + \sigma_C \eta_{t+1}$$

$$\text{Dividends: } d_{t+1} - d_t = g_D + \sigma_D \epsilon_{t+1}$$

$$\text{Random Shocks: } \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim \text{NID} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix} \right]$$

The time increment is one year. $\bar{C}_t$ is aggregate, per-capita, consumption which is exogenous to the agent. Lower case denotes logarithms of upper case quantities; e.g., $\bar{c}_t = \log(\bar{C}_t)$, $d_t = \log(D_t)$. All variables are real.

In addition to these variables, let R_t denote the gross stock return; let R_f denote the gross risk free rate; let S_t denote the share of wealth allocated to the risky asset; let

$$X_{t+1} = S_t(R_{t+1} - R_f) \tag{19}$$

denote the relative gain or loss on the risky asset; let

$$z_{t+1} = \eta \left(z_t \frac{\bar{R}}{R_{t+1}} \right) + (1 - \eta) \tag{20}$$

denote the benchmark level, where $\bar{R}$ is chosen to make median $\{z_t\} = 1$; and let C_t denote the agent's consumption.

The prospect theory utility function is

$$\mathcal{E}_0 \left[\sum_{t=0}^{\infty} \left(\rho^t \frac{C_t^{1-\gamma} - 1}{1 - \gamma} + b_0 \bar{C}_t^{-\gamma} \rho^{t+1} S_t \, \hat{v}(R_{t+1}, z_t) \right) \right], \tag{21}$$

where

$$\hat{v}(R_{t+1}, z_t) = \begin{cases} R_{t+1} - R_f & z_t \leq 1, R_{t+1} \geq z_t R_f \\ (z_t R_f - R_f) + \lambda(R_{t+1} - z_t R_f) & z_t \leq 1, R_{t+1} < z_t R_f \\ R_{t+1} - R_f & z_t > 1, R_{t+1} \geq R_f \\ \lambda(z_t)(R_{t+1} - R_f) & z_t > 1, R_{t+1} < R_f \end{cases} \tag{22}$$

and

$$\lambda(z_t) = \lambda + k(z_t - 1). \tag{23}$$

$\mathcal{E}_t$ is the conditional expectation with respect to the benchmark level z_t, which is the state variable.

The first term of the prospect theory utility function is standard CRRA utility which involves the agent's consumption C_t, the discount factor ρ, and the risk aversion parameter γ. The second term is the utility from gains or losses. It is $v(R_{t+1}, z_t)$ weighted by $b_0 \bar{C}_t^{-\gamma} \rho^{t+1} S_t$, where b_0 is a scale factor, $\bar{C}_t$ is aggregate consumption, and S_t is the share of wealth allocated to the risky asset. To understand $v(R_{t+1}, z_t)$, consider Figure 6. As seen in the figure, when there are no prior gains and losses ($z = 1$), agents dislike losses more than they appreciate gains. When there are prior losses ($z > 1$) the dislike intensifies. When there are prior gains ($z < 1$), an agent is "playing on the house's money" and does not begin to feel pain until the "house's money has been lost".

Figure 6 about here

Note that the parameter η in (20) controls sensitivity to past gains and losses. When η is zero, its lower bound, the benchmark does not depend at all on past gains and losses. The dependence increases as η approaches its upper bound of one. Agents always dislike losses more than they appreciate gains; η just determines the extent to which this dislike is delayed. See Barberis, Huang, and Santos (2001) for a more detailed motivation of the prospect theory utility function and its relation to the psychology literature.

Let V be the map from the state z to the price-dividend ratio P/D of the equity asset. It is determined by the Euler equation

$$\begin{aligned} 1 &= \rho \exp\left(g_D - \gamma g_C + \gamma^2 \sigma_C^2 (1-\omega^2)/2\right) \\ &\quad \times \mathcal{E}_t \left[\frac{1+V(z_{t+1})}{V(z_t)} \exp[(\sigma_D - \gamma\omega\sigma_C)\epsilon_{t+1}]\right] \\ &+ \quad b_0 \rho \, \mathcal{E}_t \left[\hat{v}\left(\frac{1+V(z_{t+1})}{V(z_t)} \exp(g_D + \sigma_D \epsilon_{t+1}), z_t\right)\right]. \end{aligned} \tag{24}$$

The logarithmic return on the equity asset, $r_{dt} = \log(P_{dt} + D_t) - \log(P_{d,t-1})$, is obtained from $V(z)$ using

$$r_{dt} = \log\left[\frac{1+V(z_t)}{V(z_{t-1})} \exp(g_D + \sigma_D \epsilon_t)\right]. \tag{25}$$

The Euler equation (24) depends on three self-referential equations

$$z_{t+1} = \eta\left(z_t \frac{\bar{R}}{R_{t+1}}\right) + (1-\eta), \tag{26}$$

$$R_{t+1} = \frac{1+V(z_{t+1})}{V(z_t)} \exp(g_D + \sigma_D \epsilon_{t+1}), \tag{27}$$

$$1 = \text{median}\{z_t\}. \tag{28}$$

The solution proceeds as follows. Approximate V by a piecewise linear function $V^{(0)}(z)$ in (27). (We use twenty-five join points equally spaced between 0 and 4, inclusive.) Put z_t to the first join point of $V^{(0)}(z)$. Approximate $\bar{R}$ in (26) by $(1+V(1))\exp(g_D)/V(1)$, which is a departure from Barberis, Huang, and Santos (2001). We find that this departure has a negligible effect on results but does save considerable computational time. Define $h^{(0)}$ such that $z_{t+1} = h^{(0)}(z_t, \epsilon_{t+1})$ solves (26) and (27). This is a root finding problem. We use Brent's method. Substitute $V^{(0)}(h^{(0)}(z_t, \epsilon_{t+1}))$ for $V(z_{t+1})$ in (24). Use Gauss-Hermite quadrature

to integrate out ϵ_{t+1} in (24). (We use a nine point rule which integrates polynomials up to degree seventeen exactly.) One is left with an expression that involves $V(z_t)$. Solve for $V(z_t)$ and let $V^{(1)}(z_t)$ be the solution. Solving for $V(z_t)$ is a root finding problem. Repeat for the remaining join points. Let $V^{(1)}(z)$ be the linear function that interpolates the points $(z_t, V^{(1)}(z_t))$. Repeat $h^{(i)} \to V^{(i+1)}$ until convergence.

Oddly enough, despite its complexity, the algorithm for solving the prospect theory model appears to be more stable then the algorithms for solving the habit persistence and long run risks models above.

The risk free rate has an explicit formula

$$r_f = \log\left[\rho^{-1}\exp\Big(\gamma g_C - \gamma^2\sigma_C^2/2\Big)\right]. \tag{29}$$

r_{ft} is the logarithmic return on an asset that pays one dollar one year hence with certainty.

Given model parameters

$$\theta = (g_C, g_D, \sigma_C, \sigma_D, \omega, \gamma, \rho, \lambda, k, b_0, \eta) \tag{30}$$

simulate annually and set

$$c_t^a = \log(C_t),$$

$$r_{dt}^a = r_{dt},$$

$$r_{ft}^a = r_f.$$

As with the long run risks model, the prospect theory model is richly parametrized and, even were data abundant, identification would have to come from the prior for most of the auxiliary models considered.

Our prior is

$$\pi(\theta) = \mathrm{N}\left[r_f \,|\, 0.896, \left(\frac{1}{1.96}\right)^2\right] \prod_{i=1}^{p} \mathrm{N}\left[\theta_i \,|\, \theta_i^*, \left(\frac{0.1\theta_i^*}{1.96}\right)^2\right] \tag{31}$$

where the θ_i^* are calibrated values from Barberis, Huang, and Santos (2001). They present several sets of parameter values. We selected the set with the most reasonable risk free rate and equity premium.

Prior correlations are shown in Table 6. Measures of location and scale for the prior and posterior distributions are shown in Table 7. The prior and posterior densities of the risk free rate, equity premium, equity returns, and the standard deviation of equity returns are shown in Figure 7. As previously, Table 7 and Figure 7 suggest that the prior is sufficiently informative to fill in where data are sparse but it allows the data to move the posterior where data are informative.

Table 6 about here

Table 7 about here

Figure 7 about here

The mean posterior forecast for the prospect theory model is shown in in Figure 8. The the prospect theory model predicts steady-state growth throughout the forecast period. Stock returns are predicted to be double their steady-state value in 2009, reach steady-state by 2011, and remain at steady-state thereafter.

Figure 8 about here

3 Inference for General Scientific Models

We briefly describe the Bayesian methods proposed by Gallant and McCulloch (2009) and the modifications that we found necessary. They used statistical terms in their discussion. Here we are writing for a audience that is likely to be familiar with the indirect inference and efficient method of moments literature so we shall use the terminology of that literature instead. Public domain code implementing the methods discussed in this section and a User's Guide are available at `http://econ.duke.edu/webfiles/arg/gsm`.

3.1 Estimation of Structural Model Parameters

Let the transition density of the structural model (called the scientific model by Gallant and McCulloch) be denoted by

$$p(y_t|x_{t-1},\theta), \quad \theta \in \Theta, \tag{32}$$

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

We assume that there is a transition density

$$f(y_t|x_{t-1}, \eta), \quad \eta \in H \tag{33}$$

and a map

$$g : \theta \mapsto \eta \tag{34}$$

such that

$$p(y_t|x_{t-1}, \theta) = f(y_t|x_{t-1}, g(\theta)) \qquad \theta \in \Theta. \tag{35}$$

We assume that $f(y|x, \eta)$ and its gradient $(\partial/\partial\eta)f(y|x, \eta)$ are easy to evaluate. f is called the auxiliary model here and g is called the binding function. (Gallant and McCulloch call these the statistical model and implied map, respectively.) Whenever we need the likelihood $\prod_{t=1}^n p(y_t|x_{t-1}, \theta)$, we use

$$\mathcal{L}(\theta) = \prod_{t=1}^{n} f(y_t|x_{t-1}, g(\theta)) \tag{36}$$

instead, where $\{y_t, x_{t-1}\}_{t=1}^n$ are the data and n is the sample size. In theory x_{t-1} is the same for both the structural and auxiliary models. In practice they might not be the same. We actually only need to know what lags are in the x_{t-1} for the auxiliary model.

After substituting $\mathcal{L}(\theta)$ for $\prod_{t=1}^n p(y_t|x_{t-1}, \theta)$, standard Bayesian MCMC methods become applicable. The difficulty is computing the binding function g accurately enough that the accept/reject decision in an MCMC chain (step 5 in the algorithm below) is correct when f is a complex nonlinear model, as it is in our application.

Given θ, the corresponding $\eta = g(\theta)$ is computed by minimizing Kullback-Leibler divergence

$$d(f, p) = \iint \left[\log p(y|x, \theta) - \log f(y|x, \eta)\right] \, p(y|x, \theta) \, dy \, p(x|\theta) \, dx$$

with respect to η. The advantage of Kullback-Leibler divergence over other distance measures is that the part that depends on the unknown $p(\cdot|\cdot, \theta)$, $\iint \log p(y|x, \theta) \, p(y|x, \theta) \, dy \, p(x|\theta) \, dx$,

does not have to be computed to solve this minimization problem. We approximate the integral that does have to be computed by

$$\iint \log f(y|x,\eta)\, p(y|x,\theta)\, dy\, p(x|\theta)\, dx \approx \frac{1}{N}\sum_{t=1}^{N} \log f(\hat{y}_t|\hat{x}_{t-1},\eta),$$

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 binding function is computed as

$$g : \theta \mapsto \operatorname*{argmax}_{\eta} \sum_{t=1}^{N} \log f(\hat{y}_t \,|\, \hat{x}_{t-1},\eta). \tag{37}$$

That is, one computes the maximum likelihood estimator of η for the "data" $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^{N}$. We use $N = 5000$, which requires 60000 monthly simulations in the case of the habit and long run risks models. Results (posterior mean, posterior standard deviation, etc.) are not sensitive to N; doubling N makes no difference other than doubling computational time. By accident we once set $N = 60000$ in the prospect theory model; this also made no difference. It is essential that the same seed be used to start these simulations so that the same θ always produces the same simulation.

Gallant and McCulloch (2009) run a Markov chain $\{\eta_t\}_{t=1}^{K}$ of length K to compute η. There are two other Markov chains discussed below so, to help distinguish among them, this chain is called the η-subchain. While the η-subchain must be run to provide the scaling for the model assessment method that Gallant and McCulloch propose, the $\hat{\eta}$ that corresponds to the maximum of $\sum_{t=1}^{N} \log f(\hat{y}_t \,|\, \hat{x}_{t-1},\eta)$ over the η-subchain is not a sufficiently accurate evaluation of $g(\theta)$ for our auxiliary models. This is mainly because our auxiliary models use the BEKK multivariate generalization of GARCH (Engle and Kroner, 1995). Likelihoods incorporating BEKK are notoriously difficult to optimize. We use $\hat{\eta}$ as a starting value and maximize (37) using the BFGS algorithm (Fletcher, 1987, 26–40). This too 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 θ. We use $R = 25000$. It is called the θ-chain. Its computation proceeds as follows.

The θ-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 θ. To implement it, we require a likelihood, a prior, and transition density in θ called the proposal density. The likelihood is (36).

The prior may require quantities computed from the simulation $\{\hat{y}_t, \hat{x}_{t-1}\}_{t-1}^N$ used to compute (36). Our prior requires r_f^a. When $\{\hat{y}_t, \hat{x}_{t-1}\}_{t-1}^N$ is computed $\{\hat{r}_{ft}^a\}_{t-1}^N$ is available from (11) in the case of the habit model, a similar expression in the case of the long run risks model, or (29) in the case of the prospect theory model. The risk free rate for the prior is the average $r_f^a = \frac{1}{N}\sum_{t-1}^N \hat{r}_{ft}^a$. (For the habit and prospect models, r_{ft}^a is constant over the simulation.) Quantities computed in this fashion can be interpreted as the evaluation of a functional of the structural model of the form $\Psi : p(\cdot|\cdot, \theta) \mapsto \psi$. Thus, our prior is a function of the form $\pi(\theta, \psi)$. However, the functional ψ is a composite function, $\theta \mapsto p(\cdot|\cdot, \theta) \mapsto \psi$, so that $\pi(\theta, \psi)$ is ultimately a function of θ only. Therefore, we will only write $\pi(\theta, \psi)$ when it is necessary to call attention to the subsidiary computation $p(\cdot|\cdot, \theta) \mapsto \psi$.

Let q denote the proposal density. For a given θ, $q(\theta, \theta^*)$ defines a distribution of potential new values θ^*. We use a move-one-at-a-time, random-walk, proposal density that puts its mass on discrete, separated points. The details are not required to understand our results; they are in Gallant and McCulloch (2009). However, two of these details 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. As an example, the long run risks model is not sensitive to the risk aversion parameter so that values of the risk aversion parameter could be separated as much as 1/4 without making any difference to the usefulness of the θ-chain. A constraint is that the separation usually cannot be more than a standard deviation of the proposal density. In the work reported here, we typically used 1/8 of a standard deviation. As a rough guide, proposal standard deviations are usually no more than the same order of magnitude as the posterior standard deviations that we report and no less than one order of magnitude smaller. The second detail 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 fact is irrelevant. Similarly for the joint distribution $f(y|x, g(\theta))\pi(\theta)$ considered as a function of θ; $f(y|x, \eta)$ must be properly normalized as a function of y, at least to the extent that (37) is computed correctly.

The algorithm for the θ-chain is as follows. Given a current θ^o and the corresponding

$\eta^o = g(\theta^o)$, we obtain the next pair (θ', η') as follows:

1. Draw θ^* 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 $\eta^* = g(\theta^*)$ and the functional ψ^* from the simulation $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$.

4. Compute $\alpha = \min\left(1, \frac{\mathcal{L}(\theta^*)\,\pi(\theta^*,\psi^*)\,q(\theta^*,\theta^o)}{\mathcal{L}(\theta^o)\,\pi(\theta^o,\psi^o)\,q(\theta^o,\theta^*)}\right)$.

5. With probability α, set $(\theta', \eta') = (\theta^*, \eta^*)$, otherwise set $(\theta', \eta') = (\theta^o, \eta^o)$.

It is at step 3 that we make our second modification. At that point we have putative pairs (θ^*, η^*) and (θ^o, η^o) and corresponding simulations $\{\hat{y}_t^*, \hat{x}_{t-1}^*\}_{t=1}^N$ and $\{\hat{y}_t^o, \hat{x}_{t-1}^o\}_{t=1}^N$. We use η^* as a start and recompute η^o using the BFGS algorithm, obtaining $\hat{\eta}^o$. If

$$\sum_{t=1}^{N} \log f(\hat{y}_t^o \mid \hat{x}_{t-1}^o, \hat{\eta}^o) > \sum_{t=1}^{N} \log f(\hat{y}_t^o \mid \hat{x}_{t-1}^o, \eta^o),$$

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

To provide the scaling for the prior used in absolute model assessment, there is a subsidiary computation that needs to be carried out at step 3. It is as follows. Initialize S_η and L to zero. Each time the η-subchain $\{\eta_t\}_{t=1}^K$ is run, increment L, replace S_η by $S_\eta + (\eta_{K/2} - \eta_K)(\eta_{K/2} - \eta_K)'$ and set

$$\Sigma_\eta = \frac{1}{L} S_\eta. \tag{38}$$

We use $K = 200$. All that is important is that transients have died out by the time the midpoint $K/2$ of the η-subchain has been reached and that $\eta_{K/2}$ and η_K are nearly uncorrelated.

If the proposed θ in step 1 violates a support condition that can be checked without running step 2, one skips step 2 because α in step 4 will be zero. One interprets simulation failure at step 2 as violation of a support condition and puts $\alpha = 0$ in step 4. The typical cause of failure in the sort of algorithms used to simulate asset pricing models is lack of convergence of a fixed point computation. Simulation failure appears never to have happened in the results reported here.

We compute posterior probabilities for relative model comparisons reported using method f_5 of Gamerman and Lopes (2006, section 7.2.1). That method requires one to save the values θ', $\mathcal{L}(\theta')$, $\pi(\theta', \psi')$ available at step 5. It also requires that these same values for a chain that draws from the prior for θ be saved. To draw from the prior, replace α at step 4 by $\alpha = \min\left(1, \frac{\pi(\theta^*,\psi^*)\, q(\theta^*,\theta^o)}{\pi(\theta^o,\psi^o)\, q(\theta^o,\theta^*)}\right)$.

There are now two scaling matrices Σ_η available: the one that comes from the θ-chain for the posterior and the one that comes from the θ-chain for the prior. The Σ_η that comes from the θ-chain for the prior is the one that should be used in the prior for absolute model assessment because the other has been tainted by the data.

The algorithm for the η-subchain is as follows. We use a move-one-at-a-time, random walk proposal density with continuous support. Given the current η^o, obtain the next value η' in the chain as follows;

1. Draw η^* according to $q(\eta^o, \eta^*)$.
2. Compute $\alpha = \min\left(1, \frac{\left[\prod_{t=1}^N f(\hat{y}_t|\hat{x}_{t-1},\eta^*)\right] q(\eta^*,\eta^o)}{\left[\prod_{t=1}^N f(\hat{y}_t|\hat{x}_{t-1},\eta^o)\right] q(\eta^o,\eta^*)}\right)$.
3. With probability α, set $\eta' = \eta^*$, otherwise set $\eta' = \eta^o$.

In Subsection 3.3 we shall require another chain, called the η-chain, that is computed from the data and a prior π_κ. The algorithm for that chain replaces α with

$$\alpha = \min\left(1, \frac{\left[\prod_{t=1}^n f(y_t|x_{t-1},\eta^*)\right]\, \pi_\kappa(\eta^*)\, q(\eta^*,\, \eta^o)}{\left[\prod_{t=1}^n f(y_t|x_{t-1},\eta^o)\right]\, \pi_\kappa(\eta^o)\, q(\eta^o,\eta^*)}\right).$$

Draws from the prior are also required. This is done by putting $\alpha = \min\left(1, \frac{\pi_\kappa(\eta^*)\, q(\eta^*,\eta^o)}{\pi_\kappa(\eta^o)\, q(\eta^o,\eta^*)}\right)$.

3.2 Relative Model Comparison

Relative model comparison is standard Bayesian inference although there are a few details that need to be discussed in order to connect it to Subsection 3.1.

One computes the predictive density, $\int \prod_{t=1}^{n} f(y_t|x_{t-1}, g(\theta))\, \pi(\theta)\, d\theta$, for the three structural models $p_1(y|x,\theta_1)$, $p_2(y|x,\theta_2)$, $p_3(y|x,\theta_3)$ with respective priors $\pi_1(\theta_1)$, $\pi_2(\theta_2)$, $\pi_3(\theta_3)$ using method f_5 of Gamerman and Lopes (2006, section 7.2.1). The advantage of that method is that knowledge of the normalizing constants of $f(\cdot|\cdot,\eta)$ and $\pi(\theta)$ are not required and it appears to be accurate in tests that we conducted. The computation is straightforward because the relevant information from the θ-chains for the prior and posterior are available after completion of the computations discussed in Subsection 3.1. It is important, however, that the auxiliary model be the same for all three models when the computations in Subsection 3.1 are carried out. Otherwise the normalizing constant of f would be required. One divides the predictive density for each model by the sum for the three models to get the probabilities for relative model assessment.

Note that what one is actually doing is comparing the three models $f(y|x, g_1(\theta_1))$, $f(y|x, g_2(\theta_2))$, $f(y|x, g_3(\theta_3))$, with respective priors $\pi_1(\theta_1)$, $\pi_2(\theta_2)$, $\pi_3(\theta_3)$. This is an important observation. Inference is actually being conducted with likelihoods $\prod_{t=1}^{n} f(y_t|x_{t-1}, g_1(\theta_1))$, $\prod_{t=1}^{n} f(y_t|x_{t-1}, g_2(\theta_2))$, $\prod_{t=1}^{n} f(y_t|x_{t-1}, g_3(\theta_3))$, not $\prod_{t=1}^{n} p_1(y_t|x_{t-1}, \theta_1)$, $\prod_{t=1}^{n} p_2(y_t|x_{t-1}, \theta_2)$, $\prod_{t=1}^{n} p_3(y_t|x_{t-1}, \theta_3)$.

If f encompasses all p_i, i.e., if (35) holds, then the former and the latter are the same. If not, the matter needs consideration. In Gallant and McCulloch's (2009) application they give two examples. In the first, the presence or absence of GARCH in the auxiliary model makes a dramatic difference to habit model parameter estimates. In the second, changing the thickness of the tails of the auxiliary model makes no difference. They argue on the basis of common sense and their examples that what is actually required is that the auxiliary model fit the observed data, not that it encompass p. That is why they use the term statistical model for f. However, their argument is not a proof. We examine this issue more closely in Section 5.

3.3 Absolute Model Assessment

We now shift our focus. The model of interest is the auxiliary model $f(\cdot|\cdot,\eta)$ and its parameter η. The role of the structural model $p(\cdot|\cdot,\theta)$ is to define the binding function $g(\theta)$ and the manifold

$$\mathcal{M} = \{\eta \in H \,:\, \eta = g(\theta),\, \theta \in \Theta\}\,. \tag{39}$$

The structural model can be viewed as a sharp prior on f that restricts the posterior distribution of η to lie on the manifold $\mathcal{M}$. As this prior is relaxed the posterior for η will move along a path toward the likelihood of the data under f. One can select waypoints κ_i along this path, view them as the discrete values of a parameter, assign them equal prior probability, and compute their posterior probability. If waypoints near $\mathcal{M}$ receive high posterior probability, then the data support the structural model. If waypoints far from $\mathcal{M}$ receive high posterior probability, then the data do not support the structural model. The ideas proceed as follows.

We add the additional assumption that the auxiliary model is identified (without a prior and in the frequentist sense) and has more parameters than the structural model. This assumption implies that $g^{-1}(\eta)$ exists on $\mathcal{M}$. If the structural model is identified (without a prior and in the frequentist sense), $g^{-1}(\eta)$ will map to a single point; if not, $g^{-1}(\eta)$ will be a set. With respect to the habit model, due to the discrete support of the proposal density q particular to the habit model, $g^{-1}(\eta)$ maps to a single point even though identification is problematic. In general, however, evaluating $\pi(g^{-1}(\eta))$ may involve computing the probability of a set rather than evaluating π at a single point.

We impose closeness to $\mathcal{M}$ by means of the prior

$$\pi_\kappa(\eta) \propto \pi\Big(g^{-1}(\eta^o)\Big)\, \exp\left(-\frac{1}{2}\,(\eta-\eta^o)'\,(\kappa\Sigma_\eta)^{-1}\,(\eta-\eta^o)\right) \tag{40}$$

where

$$\eta^o = \underset{\eta^o\in\mathcal{M}}{\operatorname{argmin}}\ (\eta-\eta^o)'\,(\Sigma_\eta)^{-1}\,(\eta-\eta^o)\,, \tag{41}$$

$\pi(\theta)$ is the prior for the structural model, and Σ_η is given by (38). It is easy and cheap to evaluate (41) once the computations described in Subsection 3.1 have been carried out because the binding function g is represented by pairs (θ,η) stored together with $\pi(\theta)$ at the

conclusion of Subsection 3.1. Store is traversed to find the pair (θ^o, η^o) such that η^o solves (41). Then $\pi(g^{-1}(\eta^o)) = \pi(\theta^o)$. The modifications are obvious if $g^{-1}(\eta^o)$ maps to a set. The pairs (θ, η) and scale Σ_η used to compute (41) and (40) are those for the θ-chain that draws from the prior because they are not tainted by data.

Choose three (for specificity) values κ_1, κ_2, and κ_3, ordered from small to large. Consider f under priors π_{κ_1}, π_{κ_2}, and π_{κ_3} to be three different models and compute the posterior probability for the three models with each having prior probability 1/3. That is, the pair $(f(\cdot|\cdot, \eta), \pi_\kappa(\eta))$ is considered to be a model and the posterior probability of each κ choice is proportional to $\int \prod_{t=1}^n f(y_t|x_{t-1}, \eta)\, \pi_\kappa(\eta)\, d\eta$. We use method f_5 of Gamerman and Lopes (2006, section 7.2.1) to compute $\int \prod_{t=1}^n f(y_t|x_{t-1}, \eta)\, \pi_\kappa(\eta)\, d\eta$ for κ_1, κ_2, κ_3, then normalize by the sum to get posterior probabilities.

If the posterior probability of model κ_1 is small, that is evidence against the structural model. Conversely, if it is large, that is evidence in favor of the structural model.

3.4 The Auxiliary Model

In the bivariate case the observed data are

$$y_t = (\text{consumption growth, stock returns})$$

for $t = 1, \ldots, n$. Lagged values of y_t are denoted as x_{t-1}. For auxiliary models f_0 through f_4, $x_{t-1} = y_{t-1}$. For auxiliary model f_5, $x_t = (y_{t-1}, y_{t-2})$.

The data are modeled as

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

where

$$\mu_{x_{t-1}} = b_0 + B x_{t-1}, \tag{42}$$

which is the location function of a vector autoregression, and R_{t-1} is the Cholesky factor of

$$\begin{aligned}
\Sigma_{x_{t-1}} &= R_0 R_0' && (43)\\
&\quad + Q \Sigma_{x_{t-2}} Q' && (44)\\
&\quad + P(y_{t-1} - \mu_{x_{t-2}})(y_{t-2} - \mu_{x_{t-2}})' P' && (45)\\
&\quad + \max[0, V(y_{t-1} - \mu_{x_{t-2}})] \max[0, V(y_{t-1} - \mu_{x_{t-2}})]'. && (46)
\end{aligned}$$

In our specification, R_0 is an upper triangular matrix, P and V are diagonal matrices, and Q is scalar; $\max(0, x)$ is applied elementwise. This specification is BIC preferred in simulations from the habit, long run risks, and prospect theory models at the parameter values shown in the first column of Tables 3, 5, and 7, respectively. In general P, Q, and V could be scalar, diagonal, or full matrices and there could be additional terms in higher order lags. This is the BEKK form of multivariate GARCH described in Engle and Kroner (1995) with an added leverage term (46). In computations, $\max(0, x)$ in (46) is replaced by a twice differentiable cubic spline approximation that plots slightly above $\max(0, x)$ over (0,0.1) and coincides elsewhere. Auxiliary model f_0 has term (43) only, f_1 has terms (43), (44), and (45), and f_2 through f_5 have all four terms.

The density $h(z)$ of z 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 (Gallant and Nychka, 1987). The density $h(z)$ is the normal when the degree of the Hermite polynomial is zero, which is the case for auxiliary models f_0 through f_2. For model f_3 the degree is four. For models f_4 and f_5 the degree is four but the constant term of the Hermite polynomial is a linear function of y_{t-1}. This has the effect of adding a nonlinear term to the location function (42) and the variance function (43). It also causes the higher moments of $h(z)$ to depend y_{t-1} as well.

The univariate auxiliary models are the same as the above but $\mu_{x_{t-1}}$ in (42) has dimension one and becomes the location function of a first order autoregression and $\Sigma_{x_{t-1}}$ in (43) has dimension one and becomes a GARCH(1,1) with a leverage term added.

3.5 Diagnostic Checks

The idea behind diagnostic checking is straightforward: If one has compared two structural models (p_1, π_1) and (p_2, π_2) using the same auxiliary model $f(\cdot|\eta)$ and the fit of (p_2, π_2) is preferred, then one can examine the posterior means (or modes) $\hat{\eta}_1$ and $\hat{\eta}_2$ of $f(\cdot|\eta)$ corresponding to the two fits to see which elements changed. The same is true for absolute model assessment. If one fits (f, π_{κ_1}) and (f, π_{κ_2}) and concludes that (f, π_{κ_1}) fails to fit the data, then one can examine the changes in the elements of the posterior means (or modes)

of $f(\cdot|\eta)$ corresponding to the two fits to see which elements changed.

The changes in the elements of $\hat{\eta}_1$ and $\hat{\eta}_2$ need to be normalized to facilitate meaningful comparison. Let $\hat{\eta}_{1i}$ and $\hat{\eta}_{2i}$ denote the respective ith elements of $\hat{\eta}_1$ and $\hat{\eta}_2$. Let $\hat{\sigma}_{2i}$ denote the ith posterior standard deviation of the second fit, i.e., the preferred fit. The normalization we suggest is

$$t_i = \frac{\eta_{1i} - \eta_{2i}}{\sigma_{2i}}. \tag{47}$$

The elements of the parameter η in models f_0 through f_5 of Table 1 are easy to interpret, which aids this exercise. Table 12 is an example.

There is a caveat. The t_i are often very informative but are subject to the same risk as the interpretation of t-statistics in a regression, namely, a failure to fit one characteristic of the data can show up not at the parameters that describe that characteristic but elsewhere due to correlation (colinearity). Nonetheless, despite this risk, inspection of the t_i is often the most informative diagnostic available. If one does change a structural model as suggested by the t_i, one can check to see if the modification was successful by means of absolute model assessment.

The methods proposed here are likelihood methods which means that at the conclusion of an estimation exercise a transition density that represents the data under the fitted model is available. The most useful are $f(y|x, g(\theta))$ with θ set to the posterior mode from a fit of (p, π) and $f(y|x, \eta)$ with η set to the posterior mode from a fit of (f, π_κ). One can apply standard diagnostics to these transition densities such as plotting conditional means against the data or comparing conditional volatility plots. Examples are Figures 9 through 12.

3.6 Forecasts

A forecast can be viewed as a functional $\Upsilon : f(\cdot|\cdot, \eta) \mapsto \upsilon$ of the auxiliary model that can be computed from $f(\cdot|\cdot, \eta)$ either analytically or by simulation. If $f(\cdot|\cdot, \eta)$ encompasses the structural model $p(\cdot|\cdot, \theta)$ then, due to the map $\eta = g(\theta)$, this forecast can also be viewed both as a forecast from the structural model and as function of θ. As such, it can computed at each draw in the θ-chain for the posterior and the posterior mean, mode, and standard deviation obtained. Similarly for draws from the prior. Details are in Gallant and McCulloch (2009). Examples are Figures 3, 5, and 8.

4 Habit, Long Run Risks, Prospect?

Table 8 presents the posterior probabilities for a relative model comparison for the univariate stock returns data. Table 9 is the same for an absolute model assessment. These two tables are easily summarized: All three models fit the univariate stock returns data reasonably well and none of them strongly dominates.

Table 8 about here

Table 9 about here

Results are startlingly different when we look at the bivariate consumption growth and stock returns data. Table 10 presents the posterior probabilities for a relative model comparison. The long run risks model is totally dominant over the 1930–2008 period whereas the habit model is totally dominant over the 1950–2008 period.

Table 10 about here

Table 11 presents the results for an absolute model assessment for the bivariate consumption growth and stock returns data. The auxiliary model is the encompassing model f_5. The habit model fails definitively in the 1930–2008 period and the prospect theory model fails definitively in the 1950–2008. The ordering of the posterior probabilities is as Table 10 would suggest but, aside from the habit model in the 1930–2008 period and the prospect theory model in the 1950–2008 period, results are not as stark as in Table 10.

Table 11 about here

It is of interest to determine why the habit persistence model fails to fit the bivariate data using the diagnostic checks described in Subsection 3.5. For this purpose it is more informative to use the simpler auxiliary model f_1 rather than the encompassing model f_5. Because the habit model has seven parameters and f_1 has twelve, this is a legitimate choice for the habit model. It would not be a legitimate choice for the long run risks model, which has thirteen parameters, and would be somewhat dubious for the prospect theory model which has eleven.

Table 12 presents the diagnostics for the habit model. In the table the fit of (f_1, π_κ) with $\kappa = 0.1$ is compared to the fit with $\kappa = 10$ for the bivariate data over the period 1930–2008 and over the period 1950–2008. It is clear from the table what the problems are. The habit model does not track consumption growth over the 1930–2008 period. The estimates of B_{11}, which is the feedback of consumption growth onto itself, and P_{11}, which is the feedback of consumption growth into its own volatility, are too small in absolute value. Also, the magnitude of the intercept term $b_{0,1}$ is incorrect. Of these problems, the failure to put enough conditional heteroskedasticity into the consumption growth process seems the most important ($t = -4.98$). This is consistent with the results of Gallant and McCulloch (2009). As pointed out in Subsection 3.5, one cannot say definitively whether these problems are actually due to the way consumption growth is specified in the habit model or are spillover effects from failures elsewhere. To determine this one would have to modify the habit model and use the absolute assessment procedure to see if the modification was successful.

Table 12 about here

We can confirm these interpretations visually. Figure 9 plots the conditional means for $\kappa = 0.1$ and $\kappa = 10$ against the data. The discrepancies between the fit for $\kappa = 10$, which is presumed to be the more correct fit, and the fit for $\kappa = 0.1$, which is the fit with the habit model imposed, are small. The situation changes rather dramatically in Figure 10, which plots conditional volatilities. The habit model cannot track the volatility in consumption growth or the conditional correlation between consumption growth and stock returns over 1930–1950.

Figure 9 about here

Figure 10 about here

In Table 12, the habit model does better over the more quiescent 1950–2008 period. It is still having the same problem with feedback from consumption growth to itself and is now having a problem with the feedback of stock returns to consumption growth. The problem with conditional heteroskedasticity is gone, but this is due to the quiescence of the data as

exemplified by the less restricted estimate of P_{11} in the fit with $\kappa = 10$, which is now much lower.

Plots over the quiescent period (not shown) look like Figures 9 and 10 from 1950 onward but with the dashed line superimposed upon the solid line. Also, the volatility plots are slightly smoother.

Figures 3, 5, and 8 examine the out-of-sample differences among the models and have been discussed previously (Section 2). Figures 11 and 12 examine in-sample differences over 1930–2008. The scaling in Figures 11 and 12 is the same as the scaling in Figures 9 and 10 to permit comparison. The solid line in Figures 11 and 12 is the long run risks model, which is the most correct of the three according to the relative model comparison in Table 10. In Figure 11 one sees that the conditional mean of the long run risks model (solid line) tracks consumption growth somewhat better than either the habit model (dashed line) or prospect theory model (dot-dash line). The differences in the conditional mean for stock returns in the lower panel are probably irrelevant because the volatility of stock returns is large. The differences in conditional volatility in Figure 12 are somewhat more dramatic.

Figure 11 about here

Figure 12 about here

In our view the bivariate results reported in this section can be summarized as follows. The strongest information in the data is the conditional mean of consumption growth and the conditional volatility of stock returns. The three models track this information reasonably well. The internals of these models affects how tracking the conditional mean of consumption growth and the conditional volatility of stock returns well spills over into the representation of the conditional volatility of consumption growth and the conditional correlation between consumption growth and stock returns. These two spillover effects are the main statistical differences among the three models.

5 Sensitivity Analysis

There is much experience with the data shown in Figure 1. That experience suggests that about the richest model one would be willing to fit to these data is a model with one-

lag VAR location, GARCH scale, and normal innovations. Recall, these data are annual, not quarterly, monthly, or daily where one might consider more complex specifications. The exact specification one gets using upward F-testing, BIC, AIC, etc. is sensitive to the sample period used. One can get slightly richer or coarser specifications. We think that it is fair to claim that the consensus view is that a one-lag VAR location, GARCH scale, and normal innovations is the richest model one ought to entertain. We denote this model by f_1. It is the second of the six shown in Table 1 that we shall consider in our sensitivity analysis. We use the BEKK form of multivariate GARCH (Engle and Kroner, 1995) because it is flexibly parameterized and allows a leverage effect to be included if desired. Analytic expressions for these auxiliary models are in Subsection 3.4 and code implementing them is part of the public domain distribution at `http://econ.duke.edu/webfiles/arg/gsm`.

As seen in Subsection 3.1, theory requires that the auxiliary model encompass the structural model for estimation under the structural model and for relative model comparison. A model that will encompass the three structural models that we consider has the following characteristics: a two-lag linear conditional mean function with a one-lag nonlinear conditional mean term added to it, a one-lag GARCH conditional variance function with a one-lag leverage term and a one-lag nonlinear conditional variance term added, and a flexible innovation distribution that permits fat tails and bumps. We denote this model by f_5. It is the last of the six in Table 1. It is not just one of the three structural models that requires this complexity, they all do.

Gallant and McCulloch (2009) found the same to be true with the habit model, except that they used the Bubnov-Galerkin method (Miranda and Fackler, 2002, p 152–3) to solve the habit model and used data from 1933–2001 with the years 1930–1932 used to prime recursions. They dismissed f_5 out of hand as absurd and worked with a VAR with normal innovations, which we term f_0 here, and f_1 with an R-GARCH variance specification instead of BEKK. R-GARCH is more stable numerically than BEKK but cannot allow for leverage. They did try a fat tailed innovation distribution and found that that did not change results.

The model f_0 comes closest to mimicking the results of calibration and GMM procedures as customarily implemented in macro/finance. The sufficient statistics for this model are the mean and variance of y_t and the first order autocorrelations. One is, effectively, finding

parameter values for a model that best match six moments for the bivariate consumption growth and stock returns series and best match three moments for the univariate stock returns series. Using f_0 and bivariate data from 1933–2001, Gallant and McCulloch (2009) matched Campbell and Cochrane's (1999) calibrations fairly closely as do we using f_0 over the period 1950–2008.

As discussed in Subsection 3.1 and in Gallant and McCulloch (2009), the logically correct view toward using f_1, which fits the data, instead of f_5, which encompasses the structural model, is that it is not the likelihood of the structural model that is being used. It is some other likelihood. Therefore it is not the structural models that are actually being estimated and compared.

Another point of view is the argument advanced by Gallant and McCulloch (2009) that using a sensible auxiliary model is akin to GMM estimation. One only asks that the structural models match certain features of the data and allows them to ignore others.

The logically correct way to permit the structural model to match certain features of the data and allow it to ignore others is to use auxiliary model f_5 and a prior of the form

$$\pi(\theta, \eta) = \pi(\theta)\pi(\eta) \tag{48}$$

where $\pi(\theta)$ is (12), (18), or (31) and $\pi(\eta)$ suppresses the unwanted features of f_5, namely leverage, nonlinearity, the second lag, and non-normal innovations. The functional form of f_5 makes it easy to construct such a $\pi(\eta)$ and to construct variants that impose preferences for a model without leverage, nonlinearity, a second lag, and non-normal innovations that are milder than outright suppression.

The logically correct approach will not work for the models considered here. Repeated attempts to impose (48) and milder variants have convinced us that there do not exist parameters θ that can be approached along a path from the published values for θ or other plausible values for θ that will suppress leverage, nonlinearity, the second lag, and non-normal innovations in simulations from the structural model $p(y|x, \theta)$ that use the model solution methods recommended by the proposers of the habit persistence, long run risks, and prospect theory models. There may be isolated points that may require different solution methods. If so, our computational methods cannot find them.

What to do? About all one can do is try a battery of auxiliary model specifications and see what happens.

We noted discrepancies between the work reported here and that reported in Gallant and McCulloch (2009). What turned out to cause them was partially the difference in solution methods but mostly the difference in sample periods. Accordingly, we check sensitivity to sample period as well. Looking at Figure 1 it is fairly obvious what periods to choose. The behavior of consumption growth between 1930 and 1950 is dramatically different than that for 1950 onwards while there is little difference in the behavior of stock returns anywhere. We consider two periods 1930–2008 and 1950–2008. We only use the data from 1925–1929 to prime BEKK recursions for the 1930–2008 analysis because it is of lesser quality than the data for 1930 onwards. We do not need to go back past 1930 to prime the BEKK recursions for the 1950–2008 analysis.

There is no logical requirement that the auxiliary model encompass the structural model for the purpose of absolute model assessment. In our application, the requirement that the auxiliary model have more parameters than the structural model compels the use of auxiliary model f_5 in the univariate case. We chose to use f_5 in the bivariate case as well. The results for the absolute model comparison using f_5 are shown in Tables 9 and 11 and were discussed in Section 4. To some extent a sensitivity analysis of absolute model assessment is irrelevant because one is free to choose an auxiliary model as judgment suggests as long as it has more parameters than the structural model. What can happen if one uses an auxiliary model that has fewer parameters than the structural model anyway is that nearly equal posterior probability gets assigned to all values of κ because for every relevant η in H there is nearly always an η in $\mathcal{M}$ that is close to it thereby causing π_κ not to depend on κ. Other than the commentary in this paragraph, we do not analyze the sensitivity of absolute model assessment in this section.

Table 13 displays the results for the relative comparisons for the univariate stock returns data over the period 1930–2008 and Table 14 is the same over the period 1950–2008. There is considerable sensitivity to specification of the auxiliary model in Table 13. Conclusions would be affected by the choice of auxiliary model. Throughout all specifications in Table 14 one would be indifferent between the habit model and the long run risks model. The preference

for the prospect theory model increases as the complexity of the auxiliary model increases.

Table 13 about here

Table 14 about here

Table 15 displays the results for the relative comparisons for the bivariate consumption growth and stock returns data over the period 1930–2008 and Table 16 is the same over the period 1950–2008. These results are stark. The likelihoods are far enough apart that the choice of auxiliary model is irrelevant. The choice of data period is not. The habit model simply cannot cope with the volatility of consumption growth over the period 1930–1950.

Table 15 about here

Table 16 about here

To summarize, there can be sensitivity to auxiliary model choice, as seen in Tables 13 and 14: the choice of auxiliary model does matter. Because one is not actually comparing structural models if the auxiliary model is not encompassing, it would seem that for relative model comparison it is best to use the encompassing auxiliary model, which is f_5.

We now consider the sensitivity of estimates to choice of the auxiliary model. Conceptually this entails constructing Tables 3, 5, and 7 for all our specifications. This is 72 tables; they are available at `http://econ.duke.edu/webfiles/arg/papers/appendix.pdf`. This is too much tabular information to digest. What we shall do instead is select certain lines from these tables and plot them by model over all specifications. The values selected are the risk aversion parameter, the equity premium, the volatility of stock returns, and, for the case of the bivariate data, the correlation between consumption growth and stock returns. These are Figures 13 through 16.

It is hard to make general statements after inspecting Figures 13 through 16. Sensitivity varies by value considered and by structural model. If one wants to mimic the calibrations of the macro/finance literature, then one would use the values plotted at 0 on the horizontal

axis in these plots. If one finds the argument advanced by Gallant and McCulloch (2009) that estimates from the auxiliary model that best fits the data be used, then one would use the values plotted at 1. Otherwise one would use the estimates plotted at 5, which is our choice for the results reported in Section 4. The defense of that choice is that it does make a difference and theory supports the use of auxiliary model f_5.

Figure 13 about here

Figure 14 about here

Figure 15 about here

Figure 16 about here

6 Conclusion

We used Bayesian statistical methods proposed by Gallant and McCulloch (2009) to compare the habit persistence asset pricing model of Campbell and Cochrane (2003), the long run risks model of Bansal and Yaron (2004), and the prospect theory model of Barberis, Huang, and Santos (2001). This comparison fills a void in the literature because there are, to our knowledge, no head-to-head comparisons of asset pricing models from macro/finance that strictly adhere to the principles of statistical science.

We undertook two types of comparisons, relative and absolute, over two sample periods, 1930–2008 and 1950–2008, using two series, univariate U.S. stock returns and bivariate U.S. consumption growth and stock returns. The prior for each model is that the real interest rate be within 1% of 0.896% with probability 0.95 together with a preference for model parameters that are near their published values. This prior appears to strike the right balance. It is tight enough to insure that MCMC chains mix well but loose enough to allow the data to be influential.

For the univariate series and for both sample periods, the models perform about the same in the relative comparison and fit the data reasonably well in the absolute assessment.

For the bivariate series, in the relative comparison the long run risks model overwhelmingly dominates over the 1930–2008 period, while the habit persistence model overwhelmingly dominates over the 1950–2008 period; in the absolute assessment, the habit model fails definitively in the 1930–2008 period and the prospect theory model fails definitively in the 1950–2008 period.

We undertook a diagnostic analysis to discover why the models differ when estimated from bivariate consumption growth and stock returns data. In our view the bivariate results can be summarized as follows. The strongest information in the data is the conditional mean of consumption growth and the conditional volatility of stock returns. The three models track this information reasonably well. The internals of these models affect how tracking the conditional mean of consumption growth and the conditional volatility of stock returns well spills over into the representation of the conditional volatility of consumption growth and the conditional correlation between consumption growth and stock returns. These two spillover effects are the main statistical differences among the three models.

The estimator proposed by Gallant and McCulloch (2009) is a simulation based estimator. Simulations from a structural model, which here is either the habit model, the long run risks model, or the prospect theory model, are used to evaluate a map $\eta = g(\theta)$ from the parameters θ of the structural model to the parameters η of an auxiliary model $f(y_t|x_{t-1}, \eta)$, where y_t is the observed data and x_{t-1} are predetermined variables. Thereafter, $\mathcal{L}(\theta) = \prod_{t=1}^{n} f(y_t|x_{t-1}, g(\theta))$ is used whenever a likelihood is required. Theory requires that the auxiliary model encompass the structural model. However, Gallant and McCulloch argue that one is better served by an auxiliary model that best represents the data rather than an auxiliary model that best represents simulations from the structural model. We undertook a sensitivity analysis and recomputed our results for a battery of six auxiliary models. The simplest produces estimates that mimic values obtained by calibration or GMM estimation as customarily employed in macro/finance. The next in order of complexity represents the data well. The last encompasses the three structural models considered. We find that results are sensitive to the choice of auxiliary models to some degree. Most importantly, results can differ between the model that best represents the data and the model that best represents the structural models. In view of this difference and the fact that theory supports the latter,

our conclusions are based on the encompassing auxiliary model.

The models show interesting differences in their forecasts over the 2009–2013 horizon. For these forecasts we used the bivariate data over the period 1950–2008 and the encompassing auxiliary model f_5. The habit model predicts an end to the current recession in 2009 and return to steady-state growth by 2010. Stock returns are predicted to be high in 2009 with a return to steady-state returns by 2013. The the long run risks model predicts an end to the current recession in 2010 and slow increase in the growth rate thereafter. Stock returns are predicted to be approximately at their steady-state values over the entire forecast period. The prospect theory model predicts steady-state growth throughout the forecast period. Stock returns are predicted to be double their steady-state value in 2009, reach steady-state by 2011, and remain at steady-state thereafter.

There is little substantive difference in these forecasts if one changes either the sample period from 1950–2008 to 1930–2008 or the auxiliary model from f_5 to the auxiliary model that best represents the data f_1. All that changes is that when the auxiliary model is f_1, stock returns for the prospect theory model reach steady-state in 2010 rather than 2011 and that a bump in stock returns for the habit model at 2011 is smoothed out.

7 References

Bansal, R., and A. Yaron. (2004). "Risks For the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance 59,* 1481–1509.

Bansal, R., A. R. Gallant, and G. Tauchen. (2007). "Rational Pessimism, Rational Exuberance, and Asset Pricing Models." *Review of Economic Studies,* 74, 1005–1033.

Bansal, R, D. Kiku. and A. Yaron (2006). "Risks for the Long Run: Estimation and Inference," Manuscript, Department of Economics, Duke University, Durham NC.

Barberis, N., M Huang and T. Santos (2001), "Prospect Theory and Asset Prices" *Quarterly Journal of Economics* 116, 1–54.

Campbell, John Y. (2003). "Consumption-based asset pricing," in: G.M. Constantinides & M. Harris & R. M. Stulz (eds.), *Handbook of the Economics of Finance, Volume 1*,

Elsevier, 803–887.

Campbell, J. Y., and J. Cochrane. (1999). "By Force of Habit: A Consumption-based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy 107,* 205–251.

Dejong, David N., Beth F. Ingram, Charles H. Whiteman (2000), "Keynesian Impulses versus Solow Residuals: Itentifying Sources of Business Cycle Fluctuations," *Journal of Applied Econometrics* 15, 311–329.

Del Negro, Marco, and Frank Schorfheide (2004), "Priors from General Equilibrium Models for VARS,' *International Economic Review* 45, 643–673.

Engle, R. F, and K. F. Kroner (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory* 11, 122–150.

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

Fletcher, R. (1987), *Practical Methods of Optimization, 2nd Edition,* Wiley, New York.

Gallant, A. Ronald, and Han Hong (2007), "A Statistical Inquiry into the Plausibility of Recursive Utility," *Journal of Financial Econometrics* 5, 523–590.

Gallant, A. R., and R. E. McCulloch. (2005). "On the Determination of General Statistical Models with Application to Asset Pricing." *Journal of the American Statistical Association* 104, 117–131.

Gallant, A. Ronald, and Douglas W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363–390.

Gallant, A. R. and G. Tauchen (1996) "Which Moments to Match?" *Econometric Theory* 12, 657–681.

Gallant, A. Ronald, and George Tauchen (2009), "Simulated Score Methods and Indirect Inference for Continuous-time Models," in Yacine Aït-Sahalia and Lars Peter Hansen, eds. (2009), *Handbook of Financial Econometrics,* Elsevier/North-Holland, Amsterdam.

Gamerman, D., and H. F. Lopes (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd Edition),* Chapman and Hall, Boca Raton, FL.

Gourieroux, C., A. Monfort, and E. Renault (1993), "Indirect Inference," *Journal of Applied Econometrics* 8, S85–S118.

Hansen, Lars Peter, and James J. Heckman (1996), "The Empirical Foundations of Calibration," *The Journal of Economic Perspectives* 10, 87–104.

Kiku, D. (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.

Kydland, Finn E., and Edward C. Prescott (1996), "The Computational Experiment: An Econometric Tool," *The Journal of Economic Perspectives* 10, 69–85.

Miranda, Maria J., and Paul L. Fackler (2002), *Applied Computational Economics and Finance*, MIT Press.

Mishkin, Frederick S. (1981) "The Real Rate of Interest: An Empirical Investigation," *Carnegie-Rochester Conference Series on Public Policy, The Cost and Consequences of Inflation* 15, 151–200.

Smith, A. A. (1993), "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics* 8, S63–S84.

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

Table 1. Auxiliary Models

	f_0	f_1	f_2	f_3	f_4	f_5
Mean	1 lag	1 lag	1 lag	1 lag	1 lag	2 lags
Variance	constant	garch	garch leverage	garch leverage	garch leverage	garch leverage
Errors	normal	normal	normal	flexible	flexible nonlinear	flexible nonlinear
Parms univar	3	5	6	10	11	12
Parms bivar	9	12	14	22	24	28

Bivariate GARCH variance matrices are of the BEKK form (Engle and Kroner, 1995) with one lag throughout. A nonlinear error density adds nonlinear terms that depend on one lag to the conditional mean and variance. When evaluated, data are centered and scaled and lags are attenuated by a spline transform. See Gallant and Tauchen (2009) for details. The functional form is displayed in Subsection 3.4. Parms univar is the number of parameters when the data are stock returns alone and parms bivar is the number of parameters when the data are consumption growth and stock returns. The habit persistence model has 7 parameters, the long run risks model has 13, and the prospect theory model has 11.

Table 2. Correlation Matrix of the Habit Model Prior

Parameter	g	σ	ρ	σ_w	ϕ	δ	γ
g	1.00	0.04	0.05	-0.01	-0.05	0.15	0.07
σ	0.04	1.00	0.04	-0.07	0.03	0.05	0.07
ρ	0.05	0.04	1.00	0.03	0.08	0.03	0.07
σ_w	-0.01	-0.07	0.03	1.00	-0.02	0.01	0.01
ϕ	-0.05	0.03	0.08	-0.02	1.00	0.45	0.32
δ	0.15	0.05	0.03	0.01	0.45	1.00	-0.29
γ	0.07	0.07	0.07	0.01	0.32	-0.29	1.00

Table 3. Prior and Posterior Habit Model Parameters

	Prior			Posterior		
Parameter	Mode	Mean	Std.Dev.	Mode	Mean	Std.Dev.
g	0.00157547	0.00156519	0.00008128	0.00166893	0.00159147	0.00007473
σ	0.00440979	0.00431169	0.00022113	0.00502777	0.00501054	0.00018533
ρ	0.20068359	0.20053348	0.01072491	0.19445801	0.19892873	0.00931413
σ_w	0.03228760	0.03247938	0.00169052	0.03193665	0.03175960	0.00138630
ϕ	0.98826599	0.98830499	0.00042475	0.98769760	0.98773761	0.00033629
δ	0.99046326	0.99041700	0.00043605	0.99033737	0.99033565	0.00044495
γ	2.04296875	2.04076156	0.08924751	1.97558594	1.96336336	0.07720679
r_f	0.97796400	1.07587200	0.13273052	1.02530400	0.96219600	0.12647089
$r_d - r_f$	6.04969200	5.98359600	0.07700698	6.26854800	6.23908800	0.07426341
σ_{r_d}	19.67246807	19.69228275	0.14078849	20.17062220	20.14121148	0.14442220

Parameter values are for the monthly frequency. Returns are annualized. Mode is the mode of the multivariate density. It actually occurs in the MCMC chain whereas the mean does not. In the data, $r_d - r_f = 5.59 - 0.89 = 4.7$ and $\sigma_{r_d} = 19.72$. The auxiliary model is f_5 as described in Table 1. The data are annual consumption growth and stock returns for the years 1930–2008.

Table 4. Correlation Matrix of the Long Run Risks Model Prior

Parameter	δ	γ	ψ	μ_c	ρ	ϕ_e	$\bar{\sigma}^2$	ν	σ_w	μ_d	ϕ_d	π_d	ϕ_u
δ	1.00	-0.13	-0.10	-0.07	-0.16	0.06	-0.03	0.00	0.10	-0.00	0.04	-0.10	0.07
γ	-0.13	1.00	-0.03	-0.04	0.07	0.05	0.11	-0.06	0.10	-0.01	0.00	0.00	-0.09
ψ	-0.10	-0.03	1.00	-0.04	-0.08	-0.06	0.10	-0.07	0.06	-0.05	-0.00	0.03	0.03
μ_c	-0.07	-0.04	-0.04	1.00	0.06	-0.02	0.08	-0.01	-0.06	0.09	0.03	0.01	0.05
ρ	-0.16	0.07	-0.08	0.06	1.00	0.08	-0.04	-0.04	-0.06	0.11	0.05	0.03	-0.05
ϕ_e	0.06	0.05	-0.06	-0.02	0.08	1.00	-0.05	-0.01	0.01	0.06	0.01	0.02	-0.06
$\bar{\sigma}^2$	-0.03	0.11	0.10	0.08	-0.04	-0.05	1.00	-0.02	-0.01	-0.04	-0.03	-0.06	0.06
ν	0.00	-0.06	-0.07	-0.01	-0.04	-0.01	-0.02	1.00	-0.04	0.01	0.00	0.02	0.05
σ_w	0.10	0.10	0.06	-0.06	-0.06	0.01	-0.01	-0.04	1.00	-0.07	0.07	0.03	0.06
μ_d	-0.00	-0.01	-0.05	0.09	0.11	0.06	-0.04	0.01	-0.07	1.00	0.03	0.11	0.01
ϕ_d	0.04	0.00	-0.00	0.03	0.05	0.01	-0.03	0.00	0.07	0.03	1.00	-0.05	0.14
π_d	-0.10	0.00	0.03	0.01	0.03	0.02	-0.06	0.02	0.03	0.11	-0.05	1.00	-0.05
ϕ_u	0.07	-0.09	0.03	0.05	-0.05	-0.06	0.06	0.05	0.06	0.01	0.14	-0.05	1.00

Table 5. Prior and Posterior Long Run Risks Model Parameters

	Prior			Posterior		
Parameter	Mode	Mean	Std.Dev.	Mode	Mean	Std.Dev.
δ	0.99961090	0.99934096	0.00031172	0.99964905	0.99943058	0.00029362
γ	9.89062500	10.07348625	0.48583545	9.92187500	10.00010750	0.50121255
ψ	1.49609375	1.49614344	0.07859747	1.53906250	1.50321312	0.07244585
μ_c	0.00148392	0.00148142	0.00007031	0.00151825	0.00149122	0.00007685
ρ	0.98413086	0.98408021	0.00468241	0.98284912	0.98435210	0.00320064
ϕ_e	0.03204346	0.03202031	0.00160150	0.03204346	0.03202844	0.00162241
$\bar{\sigma}^2$	0.00004041	0.00004124	0.00000196	0.00004160	0.00004061	0.00000196
ν	0.98730469	0.98738766	0.00441105	0.98199463	0.98223563	0.00299350
σ_w	0.00000168	0.00000170	0.00000009	0.00000169	0.00000170	0.00000008
μ_d	0.00120926	0.00119140	0.00006114	0.00121307	0.00120186	0.00006030
ϕ_d	2.78906250	2.80749125	0.14620180	2.88281250	2.82820500	0.15095447
π_d	4.07031250	4.11655125	0.20586470	4.17187500	4.15665625	0.19923412
ϕ_u	6.14062500	6.27596375	0.31996896	6.45312500	6.19978500	0.30424633
r_f	0.94398000	1.16133600	0.12177703	0.90874800	1.11896400	0.11709356
$r_d - r_f$	4.30737600	4.98738000	0.48844526	4.11223200	4.59213600	0.28433000
σ_{r_d}	18.28002188	18.85677597	0.17586080	19.07839616	18.58935179	0.13239826

Parameter values are for the monthly frequency. Returns are annualized. Mode is the mode of the multivariate density. It actually occurs in the MCMC chain whereas the mean does not. In the data, $r_d - r_f = 5.59 - 0.89 = 4.7$ and $\sigma_{r_d} = 19.72$. The auxiliary model is f_5 as described in Table 1. The data are annual consumption growth and stock returns for the years 1930–2008.

Table 6. Correlation Matrix of the Prospect Theory Model Prior

Parameter	g_C	g_D	σ_C	σ_D	ω	γ	ρ	λ	k	b_0	η
g_C	1.00	0.06	-0.00	0.02	0.03	-0.12	0.06	0.06	-0.11	-0.06	-0.06
g_D	0.06	1.00	0.02	0.06	-0.04	0.04	0.06	0.17	0.06	0.17	-0.02
σ_C	-0.00	0.02	1.00	0.01	0.05	-0.03	0.08	0.13	0.00	0.06	-0.03
σ_D	0.02	0.06	0.01	1.00	-0.03	-0.09	-0.14	0.13	0.06	0.11	0.10
ω	0.03	-0.04	0.05	-0.03	1.00	-0.02	0.06	-0.01	-0.01	-0.05	-0.07
γ	-0.12	0.04	-0.03	-0.09	-0.02	1.00	0.02	-0.03	-0.04	-0.03	-0.05
ρ	0.06	0.06	0.08	-0.14	0.06	0.02	1.00	0.06	-0.09	-0.14	0.02
λ	0.06	0.17	0.13	0.13	-0.01	-0.03	0.06	1.00	0.09	0.07	0.02
k	-0.11	0.06	0.00	0.06	-0.01	-0.04	-0.09	0.09	1.00	0.23	0.14
b_0	-0.06	0.17	0.06	0.11	-0.05	-0.03	-0.14	0.07	0.23	1.00	0.04
η	-0.06	-0.02	-0.03	0.10	-0.07	-0.05	0.02	0.02	0.14	0.04	1.00

Table 7. Prior and Posterior Prospect Theory Model Parameters

	Prior			Posterior		
Parameter	Mode	Mean	Std.Dev.	Mode	Mean	Std.Dev.
g_C	0.01828003	0.01792775	0.00093413	0.01846313	0.01795106	0.00095215
g_D	0.01870728	0.01833821	0.00095276	0.01849365	0.01845027	0.00097794
σ_C	0.03918457	0.03764040	0.00200690	0.03295898	0.03356905	0.00201110
σ_D	0.12231445	0.12023010	0.00611083	0.11962891	0.11738381	0.00597238
ω	0.14794922	0.15018164	0.00694094	0.14892578	0.15015283	0.00801015
γ	0.98632812	0.98511422	0.05145608	0.96484375	0.97603082	0.04958596
ρ	0.99972534	0.99783899	0.00163604	0.99969482	0.99783430	0.00202090
λ	2.17968750	2.24709750	0.11486810	2.23437500	2.18521953	0.11761822
k	9.82812500	9.86375625	0.53189914	9.90625000	9.84252984	0.53634137
b_0	2.00195312	2.00328703	0.10967111	1.89355469	1.93699477	0.12735310
η	0.91601562	0.89845969	0.04412695	0.85375977	0.85965642	0.02405305
r_f	1.75579200	1.91283600	0.05667617	1.76136000	1.91498400	0.06495191
$r_d - r_f$	5.92353600	5.49249600	0.19235810	4.88326800	4.78360800	0.12334973
σ_{r_d}	27.97748380	26.75881163	0.92424294	22.90177286	22.79236714	0.29273615

Parameter values are for the annual frequency. Mode is the mode of the multivariate density. It actually occurs in the MCMC chain whereas the mean does not. In the data, $r_d - r_f = 5.59 - 0.89 = 4.7$ and $\sigma_{r_d} = 19.72$. The auxiliary model is f_5 as described in Table 1. The data are annual consumption growth and stock returns for the years 1930–2008.

Table 8. Relative Model Comparison, Stock Returns

	Posterior Probabilities	
Model	1930–2008	1950–2008
Habit Persistence	0.28	0.44
Long Run Risks	0.48	0.42
Prospect Theory	0.24	0.14

The data are annual stock returns over the years shown. The auxiliary model is f_5 described in Table 1.

Table 9. Absolute Model Assessment, Stock Returns

	Posterior Probabilities					
	1930–2008			1950–2008		
Prior	hab	lrr	pro	hab	lrr	pro
$\kappa = 0.1$	0.29	0.36	0.10	0.40	0.39	0.29
$\kappa = 1.0$	0.30	0.26	0.30	0.38	0.35	0.34
$\kappa = 10.0$	0.41	0.38	0.60	0.22	0.26	0.37

The data are annual stock returns over the years shown. The auxiliary model is f_5 is described in Table 1. κ is the standard deviation of a prior that imposes the habit model (hab), the long run risks model (lrr), and the prospect theory model (pro), respectively, on the auxiliary model. The prior weakens as κ increases. Assessment for the long run risks model may overstate probabilities for small κ because the condition that the auxiliary model should have more parameters than the structural model is violated.

Table 10. Relative Model Comparison, Consumption Growth and Stock Returns

	Posterior Probabilities	
Model	1930–2008	1950–2008
Habit Persistence	0.00	1.00
Long Run Risks	1.00	0.00
Prospect Theory	0.00	0.00

The data are annual consumption growth and stock returns over the years shown. The auxiliary model is f_5 as described in Table 1.

Table 11. Absolute Model Assessment, Consumption Growth and Stock Returns

	Posterior Probabilities					
	1930–2008			1950–2008		
Prior	hab	lrr	pro	hab	lrr	pro
$\kappa = 0.1$	0.00	0.41	0.28	0.31	0.16	0.08
$\kappa = 1.0$	0.00	0.36	0.28	0.31	0.21	0.08
$\kappa = 10.0$	1.00	0.23	0.44	0.38	0.64	0.84

The data are annual consumption growth and stock returns over the years shown. The auxiliary model is f_5, which is described in Table 1. κ is the standard deviation of a prior that imposes the habit model (hab), the long run risks model (lrr), and the prospect theory model (pro), respectively, on the auxiliary model. The prior weakens as κ increases.

Table 12. Diagnostics for the Habit Persistence Model

	1930–2008			1950–2008		
Parameter	Mode $\kappa = 0.1$	Mode $\kappa = 10$	Diagnostic	Mode $\kappa = 0.1$	Mode $\kappa = 10$	Diagnostic
$b_{0,1}$	-0.08	-0.05	-1.30	-0.06	-0.05	-0.21
$b_{0,2}$	0.07	0.04	0.53	0.06	0.04	0.34
B_{11}	0.08	0.16	-1.62	0.09	0.15	-1.21
B_{21}	-0.16	-0.09	-0.94	-0.15	-0.22	0.64
B_{12}	0.29	0.32	-0.80	0.29	0.23	1.58
B_{22}	0.02	0.02	-0.10	0.02	0.00	0.35
$R_{0,11}$	-0.03	-0.01	-0.23	-0.03	-0.06	0.41
$R_{0,12}$	0.23	0.27	-0.85	0.23	0.22	0.29
$R_{0,22}$	0.21	0.21	-0.07	0.20	0.26	-0.74
P_{11}	-0.06	0.17	-4.98	-0.05	-0.02	-0.55
P_{22}	-0.21	-0.22	0.16	-0.21	-0.24	0.93
Q_{11}	0.91	0.91	-0.04	0.91	0.91	0.13

Shown are the posterior modes from fitting (f_1, π_κ) to the bivariate consumption growth and stock returns data over the periods and κ values shown together with the diagnostic checks described in Subsection 3.5.

Table 13. Posterior Probability, Relative Comparison, Stock Returns, 1930–2008

Model	f_0	f_1	f_2	f_3	f_4	f_5
Habit	0.47	0.71	0.28	0.36	0.28	0.28
LR Risks	0.49	0.25	0.57	0.34	0.45	0.48
Prospect	0.04	0.04	0.15	0.30	0.27	0.24

The data are annual stock returns 1930–2008. Auxiliary models f_0 through f_5 are described in Table 1.

Table 14. Posterior Probability, Relative Comparison, Stock Returns, 1950–2008

Model	f_0	f_1	f_2	f_3	f_4	f_5
Habit	0.51	0.49	0.44	0.42	0.46	0.44
LR Risks	0.47	0.42	0.51	0.49	0.45	0.42
Prospect	0.02	0.10	0.05	0.09	0.09	0.14

The data are annual stock returns 1950–2008. Auxiliary models f_0 through f_5 are described in Table 1.

Table 15. Posterior Probability, Relative Comparison, Consumption Growth and Stock Returns, 1930–2008

Model	f_0	f_1	f_2	f_3	f_4	f_5
Habit	0.00	0.00	0.00	0.00	0.00	0.00
LR Risks	1.00	1.00	1.00	1.00	1.00	1.00
Prospect	0.00	0.00	0.00	0.00	0.00	0.00

The data are annual stock returns and consumption growth 1930–2008. Auxiliary models f_0 through f_5 are described in Table 1.

Table 16. Posterior Probability, Relative Comparison, Consumption Growth and Stock Returns, 1950–2008

Model	f_0	f_1	f_2	f_3	f_4	f_5
Habit	1.00	1.00	1.00	1.00	1.00	1.00
LR Risks	0.00	0.00	0.00	0.00	0.00	0.00
Prospect	0.00	0.00	0.00	0.00	0.00	0.00

The data are annual stock returns and consumption growth 1950–2008. Auxiliary models f_0 through f_5 are described in Table 1.

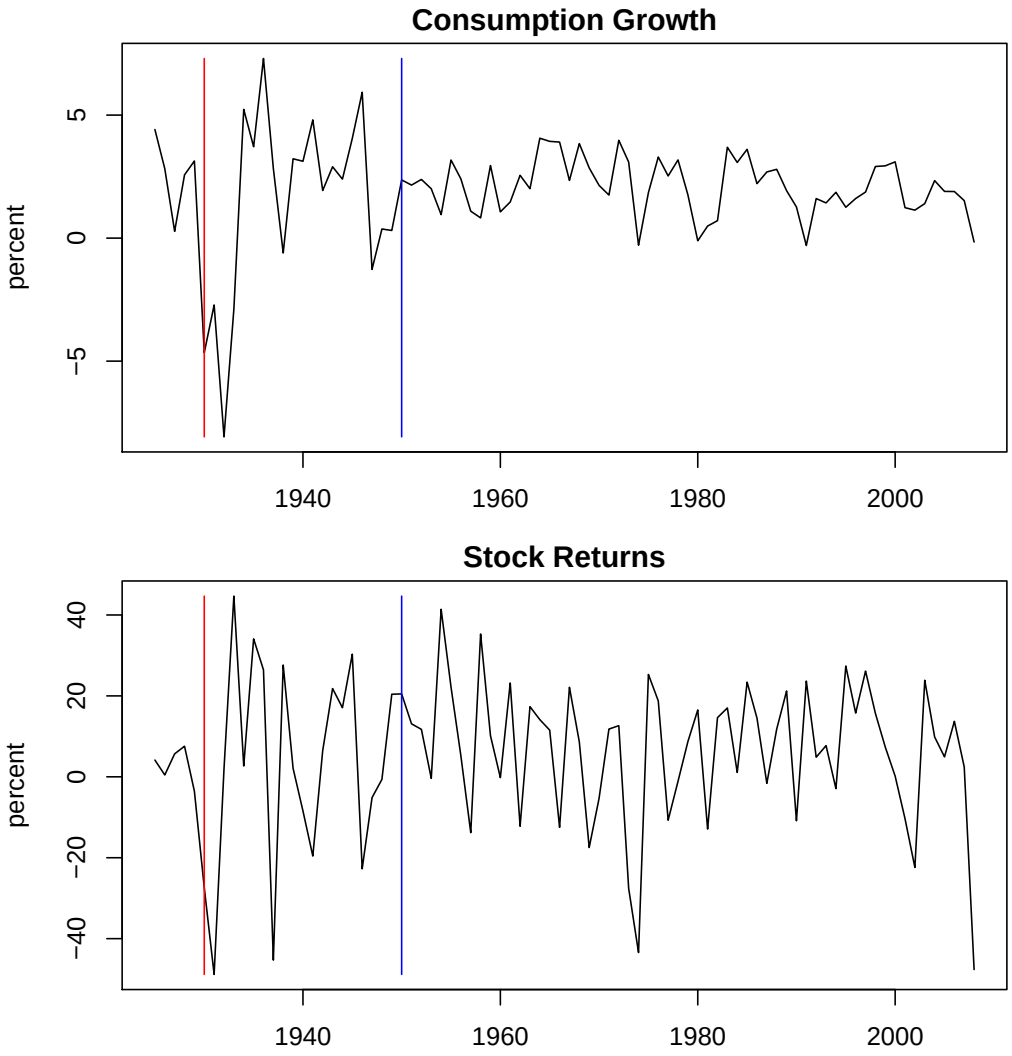


Figure 1. Real Stock Returns and Consumption Growth, 1925–2008 The left vertical line is at 1930 and the right at 1950. The data collection protocol is as described in Bansal, Gallant, and Tauchen (2007) for the period 1930–2008. The earlier data, which is only used to prime recursions, are the inflation adjusted Dow-Jones industrial average and a real U.S. consumption growth series kindly supplied by Robert Barro.

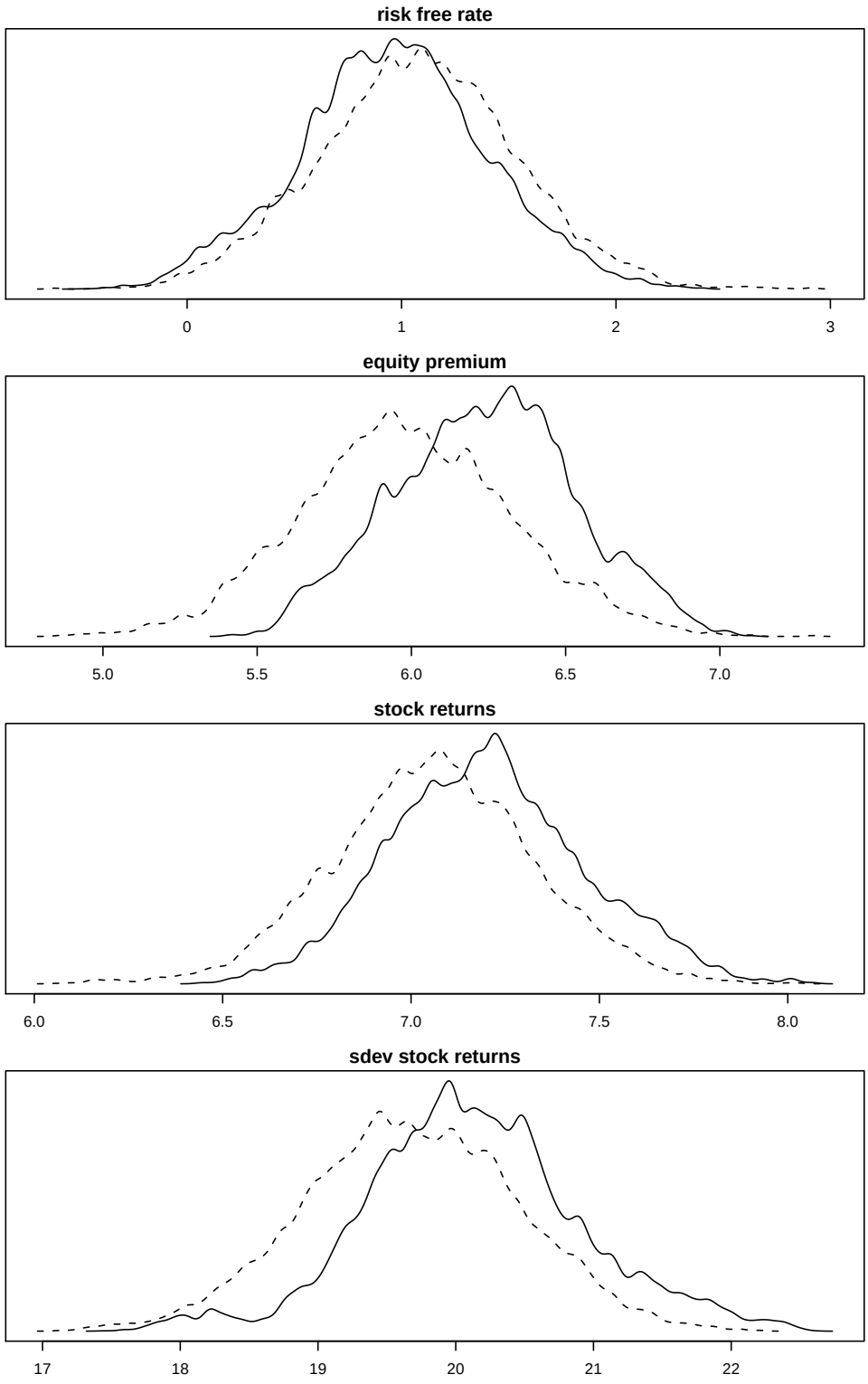


Figure 2. Prior and Posterior Density Estimates, Habit Persistence Model. The dashed line is the prior. The solid line is the posterior. Other details as in Table 3. Bandwidths are small to reduce smudging of isolated, peaked modes.

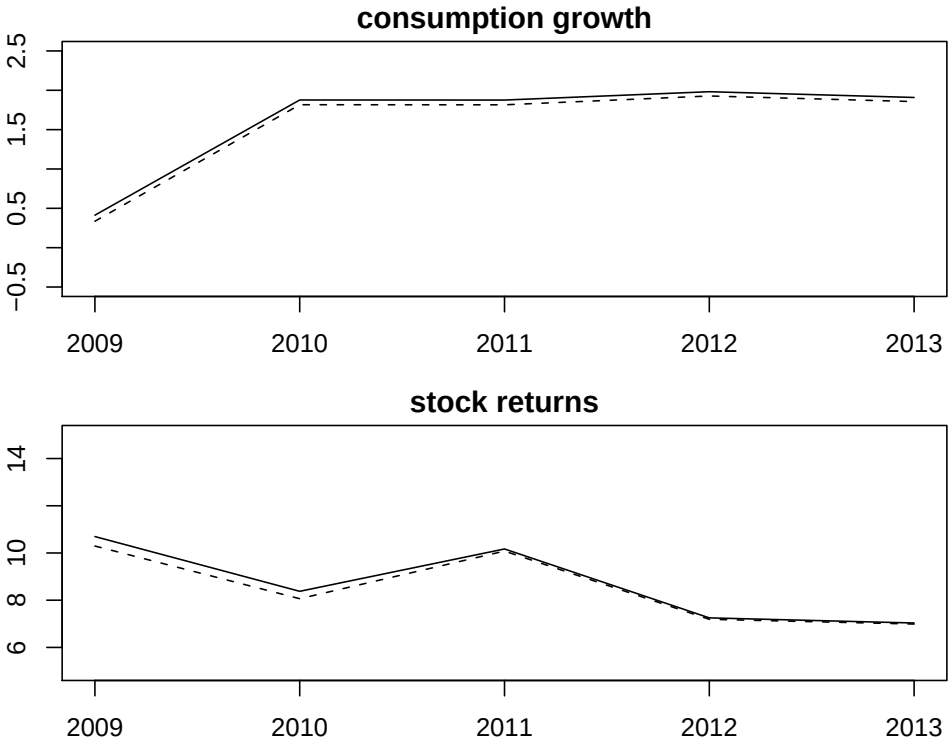


Figure 3. Prior and Posterior Forecasts, Habit Persistence Model. The dashed line is the prior. The solid line is the posterior. They are mean prior and posterior forecasts, respectively. Other details as in Table 3.

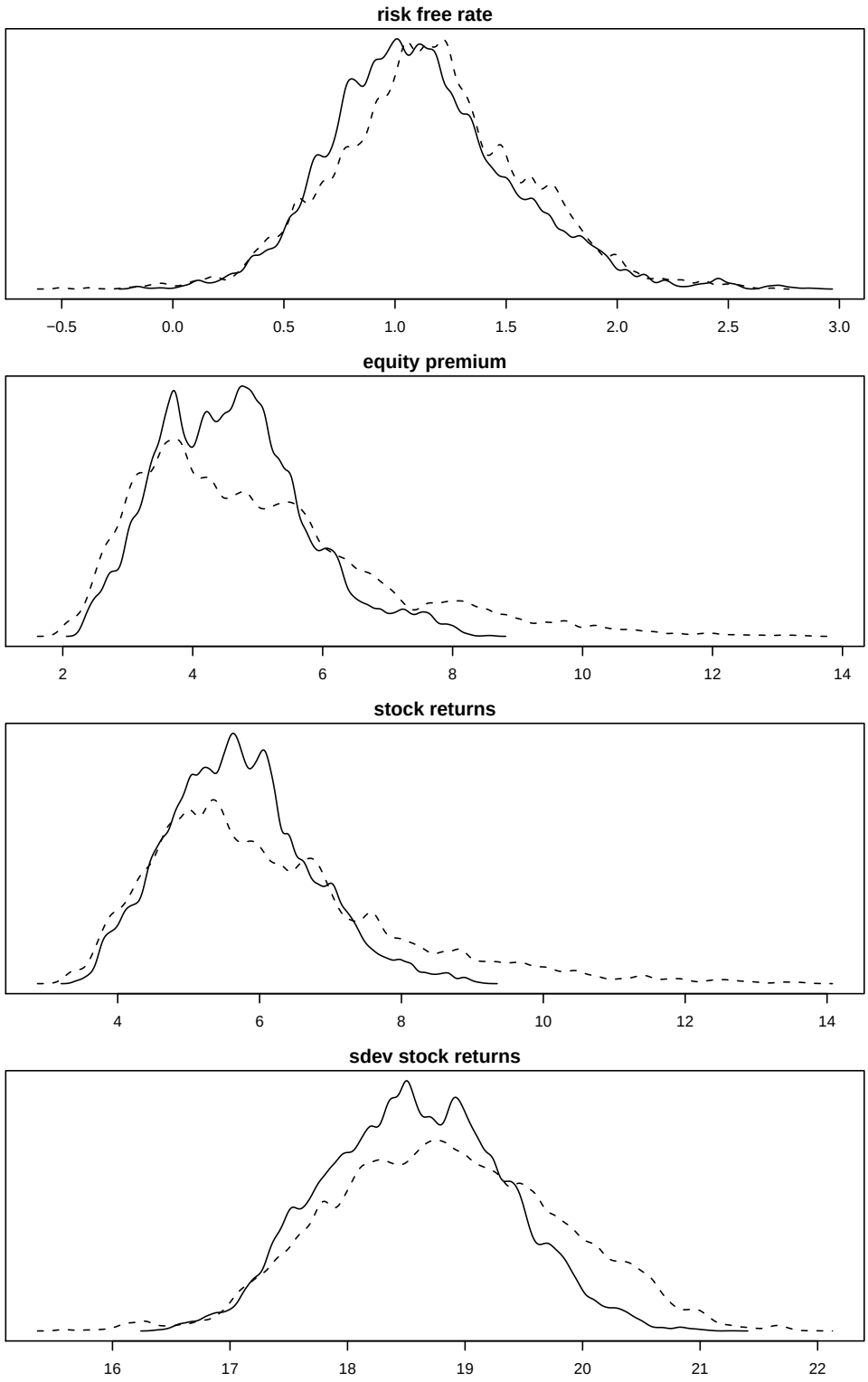


Figure 4. Prior and Posterior Density Estimates, Long Run Risks Model. The dashed line is the prior. The solid line is the posterior. Other details as in Table 5. Bandwidths are small to reduce smudging of isolated, peaked modes.

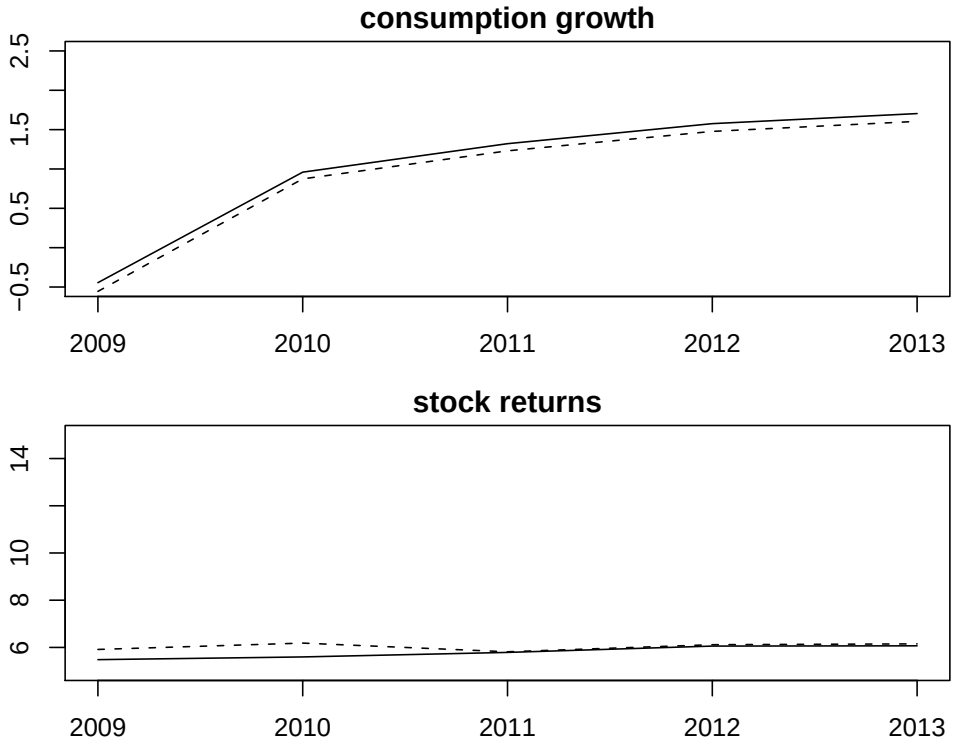


Figure 5. Prior and Posterior Forecasts, Long Run Risks Model. The dashed line is the prior. The solid line is the posterior. They are mean prior and posterior forecasts, respectively. Other details as in Table 5.

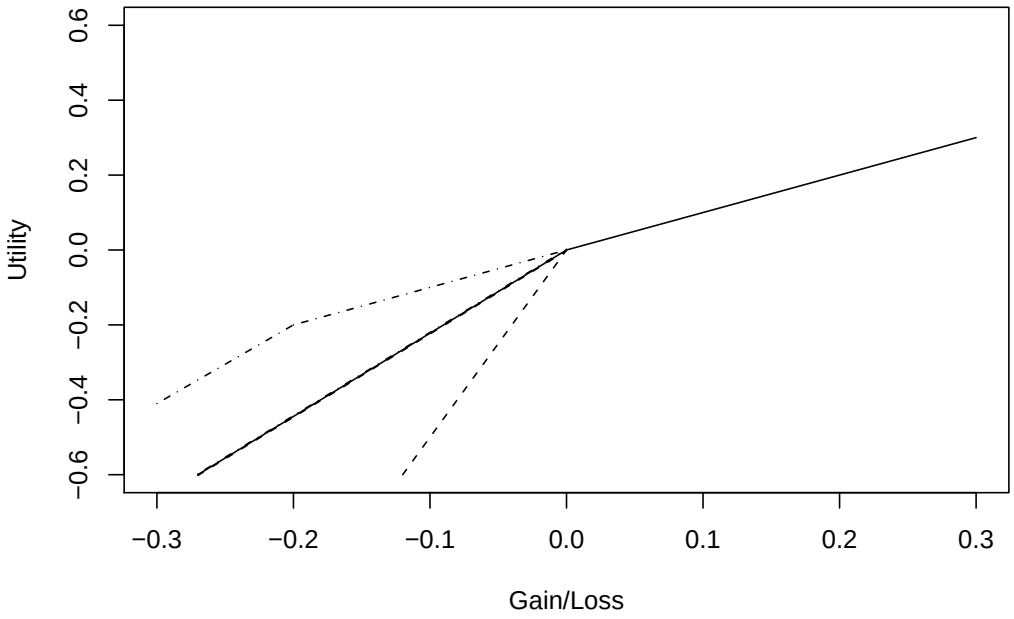


Figure 6. Utility of Gains and Losses in the Prospect Theory Model. The dot-dash line represents the case where the investor has prior gains ($z < 1$), the dashed line the case of prior losses ($z > 1$), and the solid line the case where the investor has neither prior gains nor losses ($z = 1$).

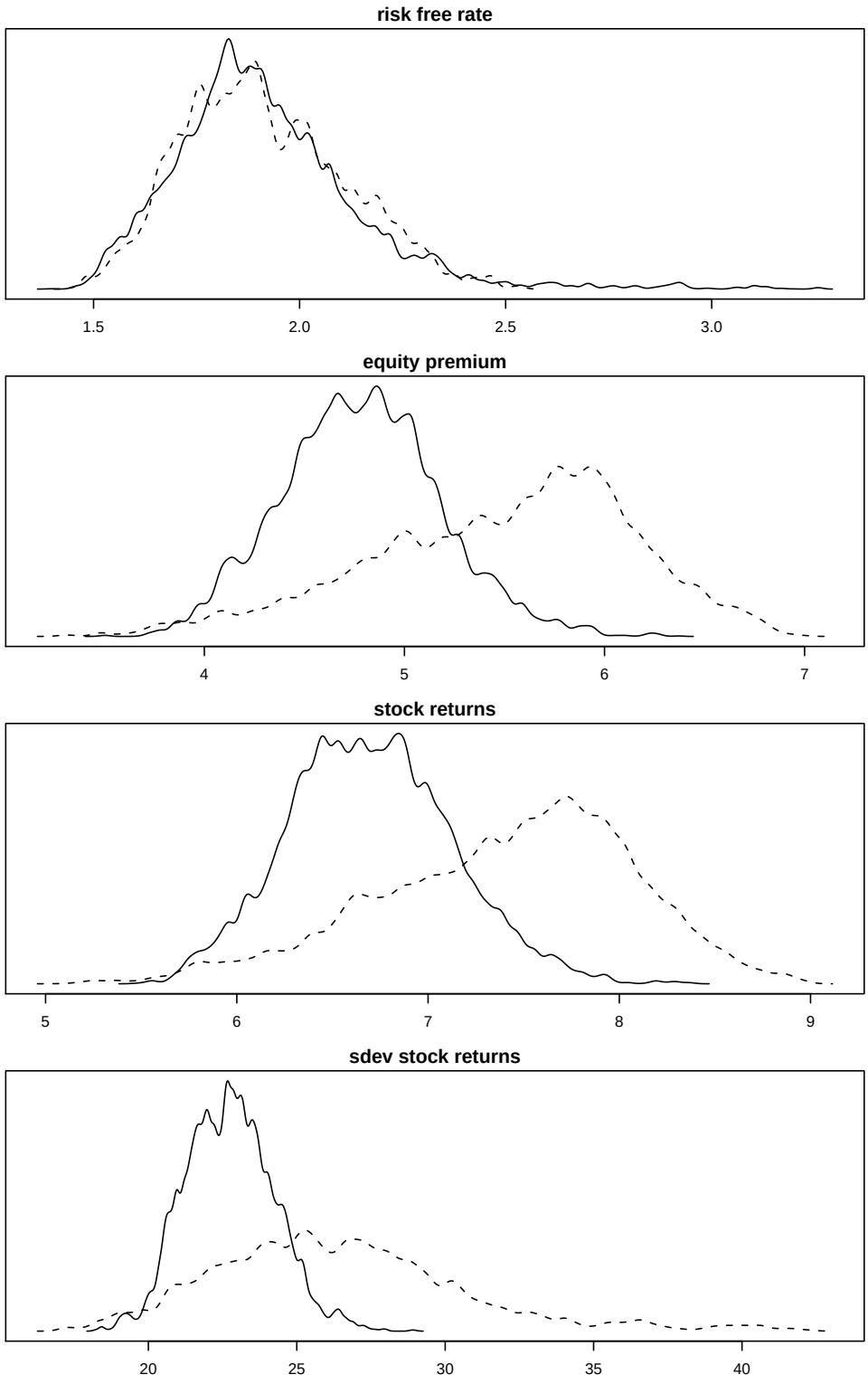


Figure 7. Prior and Posterior Density Estimates, Prospect Theory Model. The dashed line is the prior. The solid line is the posterior. Other details as in Table 7. Bandwidths are small to reduce smudging of isolated, peaked modes.

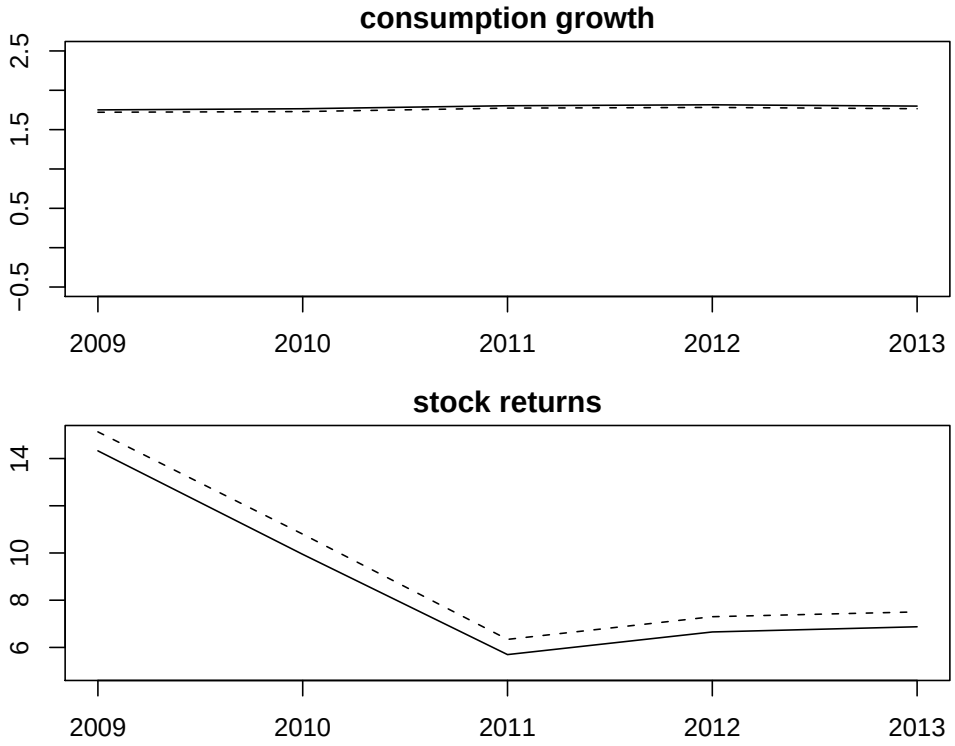


Figure 8. Prior and Posterior Forecasts, Prospect Theory Model. The dashed line is the prior. The solid line is the posterior. They are mean prior and posterior forecasts, respectively. Other details as in Table 7.

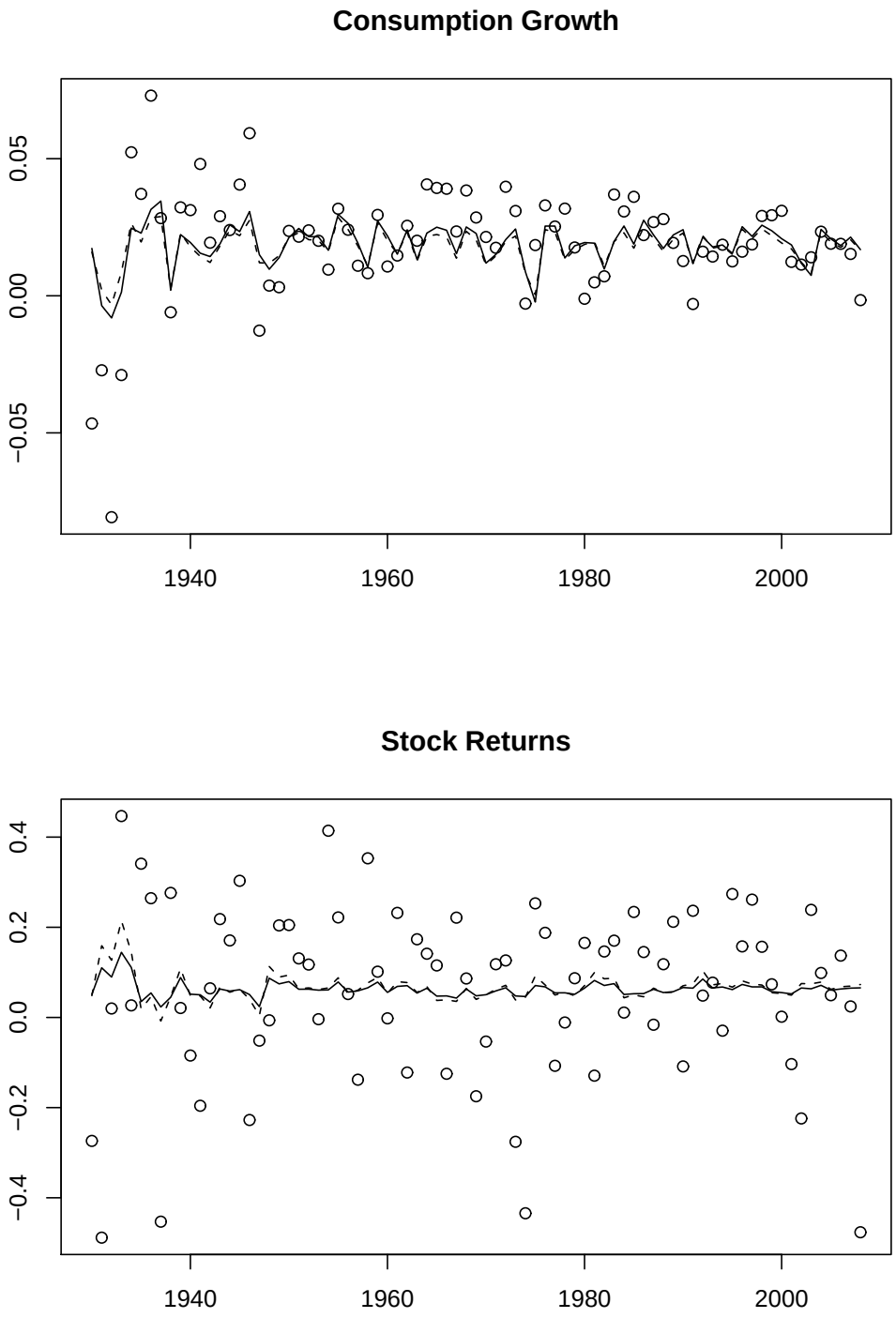


Figure 9. Conditional Mean of the Habit Persistence Model. The solid line is the conditional mean of auxiliary model f_1 with its parameters set to the posterior mode from fitting (f_1, π_κ) with $\kappa = 10$ to the bivariate consumption growth and stock returns data over the period 1930–2008. The dashed line is the same with $\kappa = 0.1$. κ is the standard deviation of a prior that imposes the habit persistence model on the auxiliary model f_1. The prior weakens as κ increases.

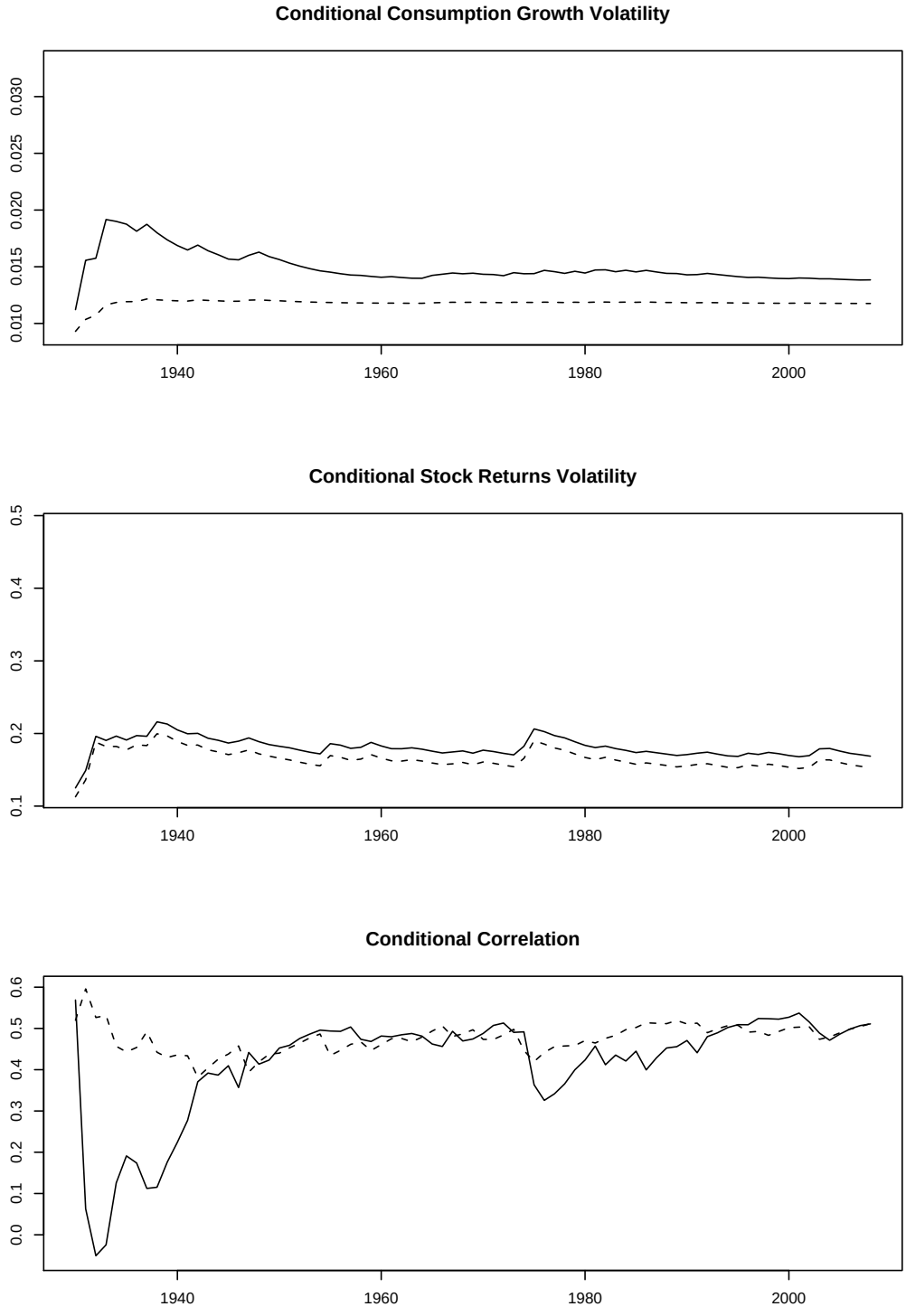
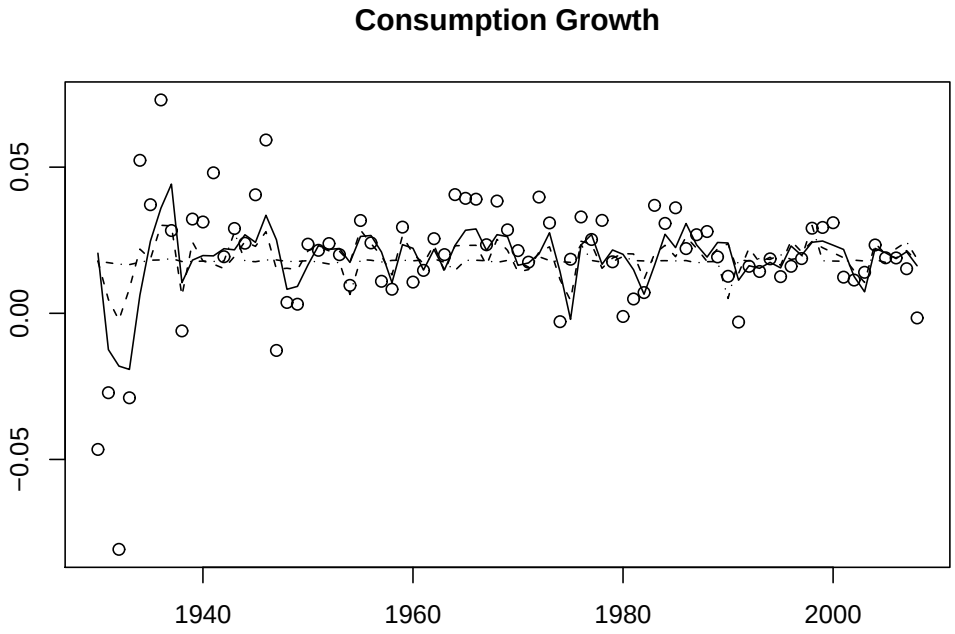


Figure 10. Conditional Volatility of the Habit Persistence Model. The solid line is the conditional volatility of auxiliary model f_1 with its parameters set to the posterior mode from fitting (f_1, π_κ) with $\kappa = 10$ to the bivariate consumption growth and stock returns data over the period 1930–2008. The dashed line is the same with $\kappa = 0.1$. κ is the standard deviation of a prior that imposes the habit persistence model on the auxiliary model f_1. The prior weakens as κ increases.



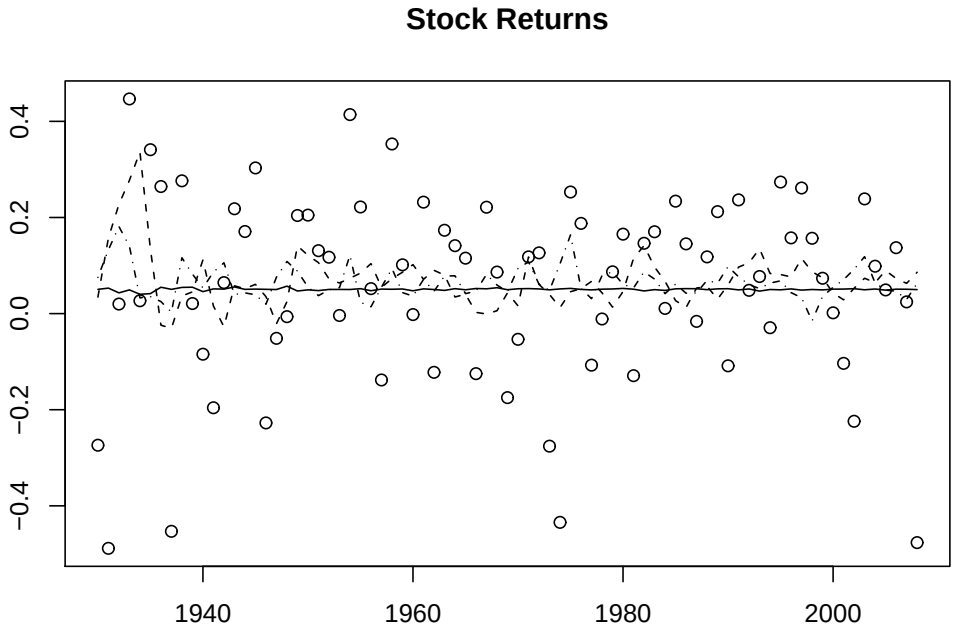


Figure 11. Conditional Means of the Three Models. The solid line is the conditional mean of the long run risks model with its parameters set to the posterior mode from fitting to the bivariate consumption growth and stock returns data over the period 1930–2008 using auxiliary model f_5. The dashed line is the same for the habit persistence model and the dot-dash line is the same for the prospect theory model.

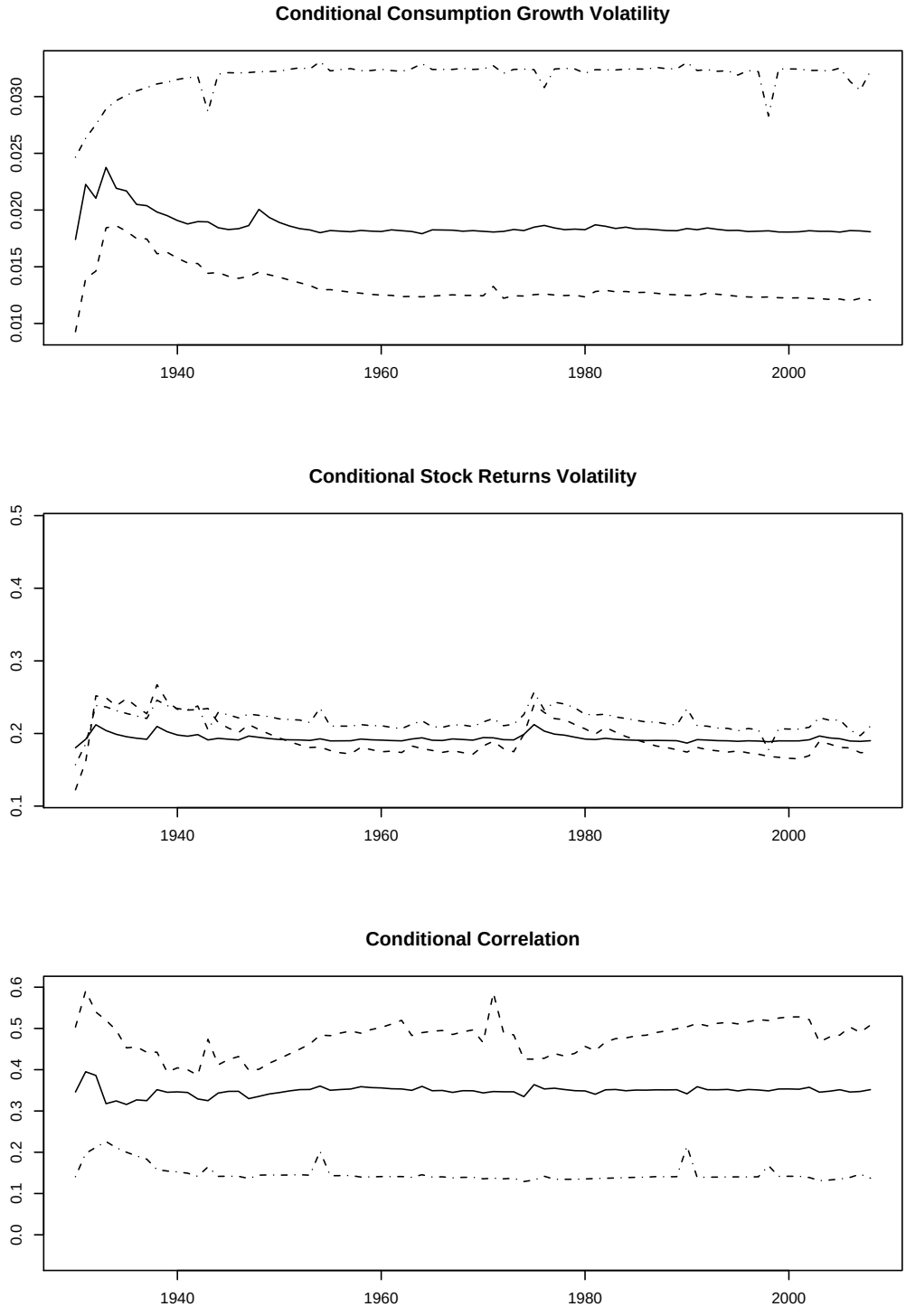


Figure 12. Conditional Volatility of the Three Models. The solid line is the conditional volatility of the long run risks model with its parameters set to the posterior mode from fitting to the bivariate consumption growth and stock returns data over the period 1930–2008 using auxiliary model f_5. The dashed line is the same for the habit persistence model and the dot-dash line is the same for the prospect theory model.

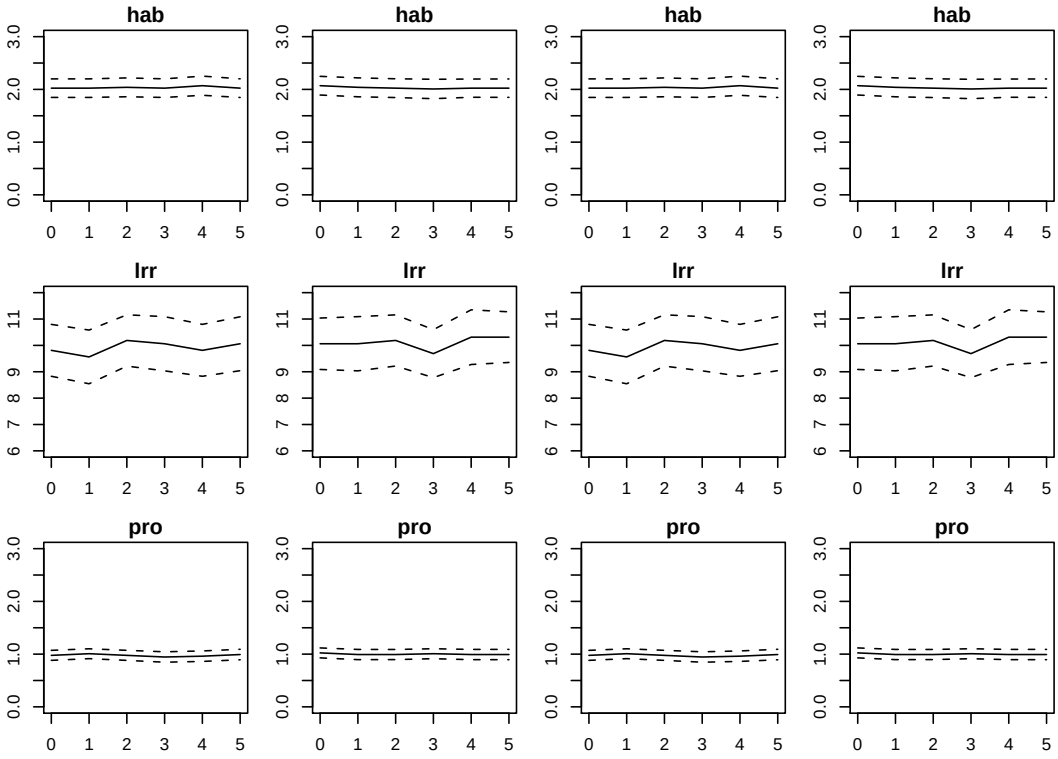


Figure 13. Sensitivity to Specification of the Risk Aversion Parameter. In each plot, the solid line is the posterior mean and the dashed lines are plus and minus 1.96 posterior standard deviations plotted against the auxiliary models f_0 through f_5. From the left, the first column is for the bivariate consumption growth and stock returns data from 1930–2008, the second for the bivariate data from 1950–2008, the third for the univariate stock returns data 1930–2008, and the fourth for the univariate data from 1950–2008.

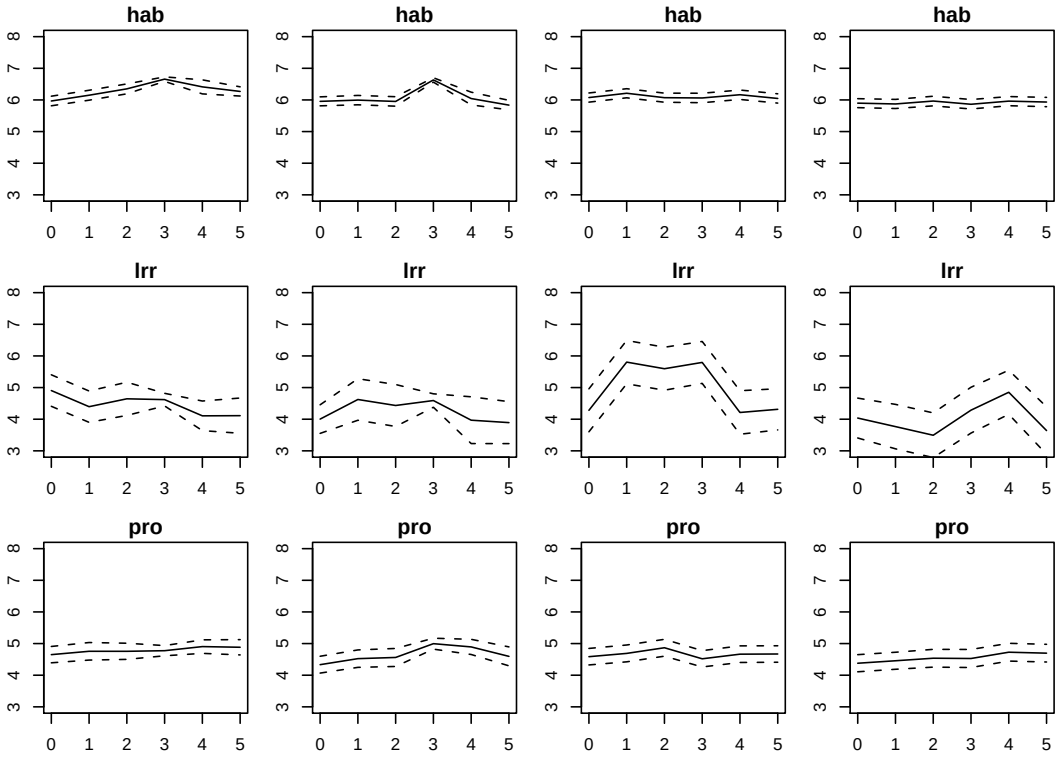


Figure 14. Sensitivity to Specification of the Equity Premium. In each plot, the solid line is the posterior mean and the dashed lines are plus and minus 1.96 posterior standard deviations plotted against the auxiliary models f_0 through f_5. From the left, the first column is for the bivariate consumption growth and stock returns data from 1930–2008, the second for the bivariate data from 1950–2008, the third for the univariate stock returns data 1930–2008, and the fourth for the univariate data from 1950–2008.

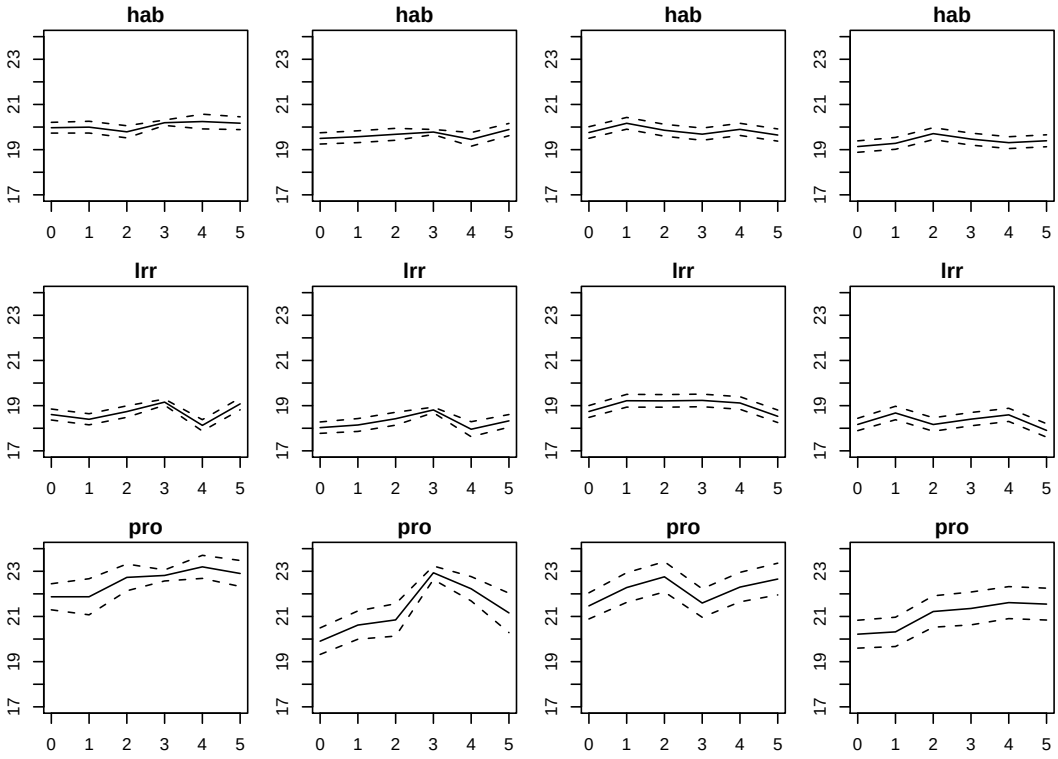


Figure 15. Sensitivity to Specification of Stock Returns Volatility. In each plot, the solid line is the posterior mean and the dashed lines are plus and minus 1.96 posterior standard deviations plotted against the auxiliary models f_0 through f_5. From the left, the first column is for the bivariate consumption growth and stock returns data from 1930–2008, the second for the bivariate data from 1950–2008, the third for the univariate stock returns data 1930–2008, and the fourth for the univariate data from 1950–2008.

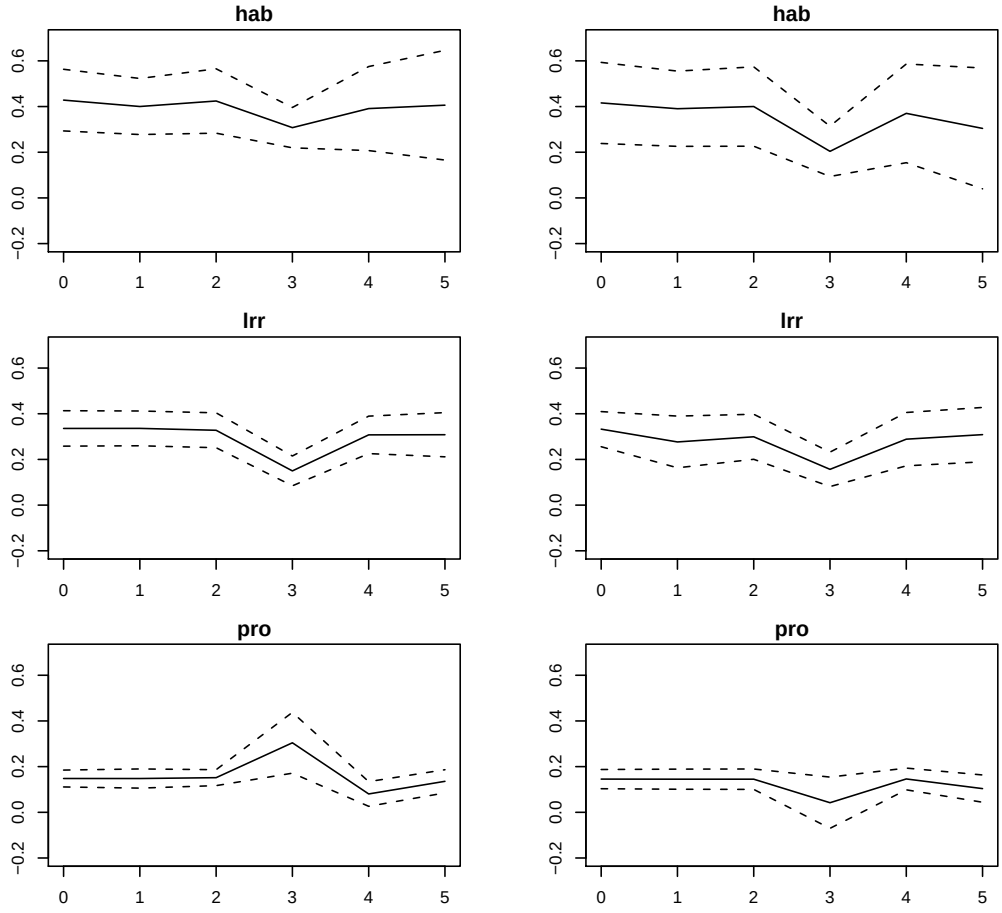


Figure 16. Sensitivity to Specification of the Correlation between Consumption Growth and Stock Returns. In each plot, the solid line is the posterior mean and the dashed lines are plus and minus 1.96 posterior standard deviations plotted against the auxiliary models f_0 through f_5. From the left, the first column is for the bivariate consumption growth and stock returns data from 1930–2008 and the second is for the bivariate data from 1950–2008.