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

Eric M. Aldrich	A. Ronald Gallant
Duke University and	Duke University and
Federal Reserve Bank	New York University
of Atlanta	

Forthcoming, *Journal of Financial Econometrics.*

Abstract

We use recently proposed 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 improve these Bayesian methods so that they can accommodate highly nonlinear models such as the three aforementioned. Our substantive results can be stated succinctly: If one believes that the extreme consumption fluctuations of 1930–1949 can recur, although they have not in the last sixty years even counting the current recession, then the long-run risks model is preferred. Otherwise, the habit model is preferred.

Keywords and Phrases: Statistical Tests, Habit, Long-Run Risks, Prospect Theory, Asset Pricing.

JEL Classification: E00, G12, C51, C52

Introduction

The goal of this article is to fill a void in the literature. There are, to our knowledge, no head-to-head, statistical (i.e. likelihood based or asymptotically equivalent) comparisons of asset pricing models from macro/finance. This paper fills the void. The asset pricing models considered are the habit persistence model of Campbell and Cochrane (1999), CC hereafter, the long-run risks model of Bansal and Yaron (2004), BY hereafter, and the prospect theory model of Barberis, Huang, and Santos (2001), BHS hereafter. There are two reasons for this choice: These three models are arguably the leading contenders and the authors describe their computational methods precisely enough to permit replication of their results.

The need for a statistical comparison of asset pricing models is underscored by the ongoing debate between advocates of the long-run risks and habit models. Beeler and Campbell (2009) claim that the long-run risks model is rejected by historical data on the basis of the predictability of excess returns, consumption growth, dividend growth, and their respective volatilities by the price to dividend ratio. Bansal, Kiku, and Yaron (2009) argue that the long-run risks model provides adequate predictability results when using a vector autoregression (VAR) based on consumption growth, price to dividend ratio, and the real risk-free rate. Bansal, Kiku and Yaron also argue that the habit model provides counterfactual predictability results for the price to dividend ratio when using lagged consumption growth as a regressor. Because the argument is based on the selective use of statistics by the advocates, it can be continued indefinitely without resolution. In contrast, the likelihood of model output contains all the information in these models. Therefore, likelihood based inference, in principle, resolves the debate definitively.

We know of only one other study that attempts a head-to-head, statistical comparison of asset pricing models: Bansal, Gallant, and Tauchen (2007), BGT hereafter. BGT compared the habit model to the long-run risks model using frequentist methods. Their methods could not distinguish between the 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 papers cited

above use annual data. BHS insist that annual is the only frequency that is appropriate to their model. Using more abundant higher frequency data to compare models is not an option. They were not designed to explain high frequency data.

Failing to achieve a definitive statistical result, BGT proceeded to compare the models using the more traditional methods of macro/finance, which consist of enumerating some moments, evaluating them both from the data and from a model simulation, and comparing, often without taking sampling variability into account. On the basis of such comparisons, BGT conclude that the long-run risks model is preferred. Of these comparisons, they relied mostly on the fact that the habit model provides counter factual predictability results for the price to dividend ratio when using lagged consumption growth as a regressor.

In addition to the fact that the BGT comparison, in the end, was not statistical, there are other concerns. BGT did not actually compare the models proposed by CC and BY. 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). Our view is that fairness dictates that one should use the model that was actually proposed by the originator in comparisons, not a modified model, and that one should use the same solution method that was proposed. To state our view succinctly, the model is the simulation algorithm proposed by the originator; it is not the mathematical equations that suggested the algorithm.

The data we use are annual, per capita, real, U.S., stock returns, consumption growth, and the price to dividend ratio from 1925–2008. The comparisons are for the periods 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. It turns out that the difference in consumption growth volatility dramatically influences results.

(Figure 1 about here)

Gallant and McCulloch (2009), GM hereafter, introduced a Bayesian method for fitting a model from a scientific discipline (scientific model for short) for which a likelihood is not readily available to sparse data. They synthesize a likelihood by means of an auxiliary

4

model and simulation from the scientific model. GM used the term "statistical model" for "auxiliary model." We do not because the consonance between "scientific model" and "statistical model" causes confusion and also because we call into question the premise on which GM chose the term "statistical model."

In the GM framework, the auxiliary model must nest the scientific model for the methodology to be logically correct. GM argue that it is better to use an auxiliary model that represents the data well, even if it does not nest the scientific model. 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 GM; f_5 is the nesting model for univariate (stock returns alone) and bivariate (consumption growth and stock returns) data.

We use the protocol set forth in GM to establish nesting, which, briefly, is as follows. The models shown in Table 1 are the beginning of a sequence (whose progression is described in Subsection 2.4) that, if continued, is dense for the space in which the scientific model must lie. Therefore, what needs to be done to find a nesting model is to select the correct truncation point. Here we use the Schwarz (1978) criterion to select it followed by some diagnostic checks that GM discuss.

Increasing the dimension of a multivariate time series often simplifies the conditioning structure. That happens here: f_0 is the nesting model for trivariate (stock returns, consumption growth, and price to dividend ratio) data. We check our f_0 results using models f_1 through f_3. Computations to fit a nonlinear model to data that does not identify it well are often unstable, as they are for models f_4 and f_5 when fitted to trivariate simulations. Therefore, we do not consider f_4 and f_5 for trivariate data. The prospect theory model puts its mass on a two dimensional subspace. This violates the GM regularity conditions. Therefore, we do not consider the prospect theory model for trivariate data.

(Table 1 about here)

We find that the computational methods that GM proposed are not sufficiently accurate to compare the habit, long-run risks, and prospect theory models. This is due to the fact that the auxiliary model f_5 that nests these three is more complex than the auxiliary models that GM considered. A contribution of this paper is a refinement of GM's methods that increases

accuracy to the point that auxiliary models as complex as f_5 can be used in applications.

1 Models Considered

The intuitive notions behind any consumption based asset pricing model are that agents receive wage, interest, and dividend income from which they purchase consumption. Agents seek to reallocate their consumption over time by trading shares of stock that pay a random dividend and bonds that pay interest with certainty. This reallocation is done for the purpose of insuring against spells of unemployment, providing for retirement, etc. Trading activity enters the model via the constraint that an agent's purchase of consumption, bonds, and stock cannot exceed wage, interest, and dividend income in any period. When applying the model to a national economy, consumption and dividends can be used as the driving processes instead of wages and dividends because wages can be recovered by subtracting dividends from consumption. (Someone must own the stock so the dividends must be received while for bonds someone pays interest and another receives so there are no net interest receipts.) Agents are endowed with a utility function that depends on the entire consumption process. The first order conditions of their utility maximization problem determine a map from present and past values of the driving processes to the present price of a stock and a bond. These models are simulated by first simulating the driving processes and then evaluating the map that determines stock and bond prices. For each model, we shall describe the driving processes and the utility function, leaving a description of the algorithm for computing the map to the cited literature.

Our prior, which is the same for all models, is

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

where $r_f^a = \lim_{n\to\infty}(1/n)\sum_{t=1}^{n} r_{ft}^a$ is the ergodic mean of the risk-free rate r_{ft}^a and the θ_i^* are the parameter values published by the proposer. 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. BGT argue that imposing the risk-free rate *a priori* will produce better estimates than using an *ex ante* risk-free rate series because any empirical *ex*

ante risk-free rate series is mostly noise due to the difficulty of determining *ex ante* inflation (Mishkin, 1981). The proposers used their judgment to determine the θ_i^*. They took data, some of which differed from ours, partially into account in forming their judgment but none could have used data more recent than 2003. Thus our prior is partially, if not completely, independent of our data. 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. It is loose enough to allow the data to be influential. Making it looser still causes numerical problems. For the purposes of this paper, it is pointless to consider a tighter prior.

Throughout this section, lower case denotes the logarithm of an upper case quantity; e.g., $c_t = \log(C_t)$, where C_t is consumption during time period t, and $d_t = \log(D_t)$, where D_t is dividends paid during period t. The exceptions are the geometric risk-free interest rate $r_{ft} = -\log P_{f,t-1}$ and the geometric stock return inclusive of dividends $r_{dt} = \log(P_{dt} + D_t) - \log P_{d,t-1}$, where $P_{f,t-1}$ is the price of a discount bond at the beginning of time period t that pays \$1 with certainty at the end of period t, $P_{d,t-1}$ is the price of a stock at the beginning of time period t, and P_{dt} its price at the end.

1.1   The Habit Persistence Model

The driving processes for the habit persistence model are

$$\text{Consumption: } c_t - c_{t-1} = g + v_t, \tag{2}$$

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

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

The utility function is

$$\mathcal{E}_0 \left(\sum_{t=0}^{\infty} \delta^t \frac{(S_t C_t)^{1-\gamma} - 1}{1 - \gamma} \right), \tag{3}$$

where habit persistence is implemented by two equations:

$$\text{Surplus ratio: } s_t - \bar{s} = \phi \left(s_{t-1} - \bar{s} \right) + \lambda(s_{t-1}) v_{t-1}, \tag{4}$$

$$\text{Sensitivity function: } \lambda(s) = \begin{cases} \frac{1}{\bar{s}} \sqrt{1 - 2(s - \bar{s})} - 1 & s \leq s_{\max} \\ 0 & s > s_{\max} \end{cases}. \tag{5}$$

γ is a measure of curvature, which scales attitudes toward risk, and δ is the agent's discount factor. $\mathcal{E}_t$ 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 as $\bar{S} = \sigma\sqrt{\gamma/(1-\phi)}$ and $s_{\max} = \bar{s} + (1 - \bar{S}^2)/2$. The variable $X_t = C_t(1 - S_t)$ is called external habit. By substituting $S_t C_t = C_t - X_t$ in (3) one can see that utility is extremely low when consumption is close to X_t for $\gamma > 1$. Habit X_t is determined by past consumption as is seen by noting that $v_{t-1} = \log(C_{t-1}/C_{t-2}) - g$ in (4). Given the habit model's parameters

$$\theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma), \tag{6}$$

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

$$C_t^a = \sum_{k=0}^{11} C_{12t-k}, \tag{7}$$

$$c_t^a = \log(C_t^a), \tag{8}$$

$$r_{dt}^a = \sum_{k=0}^{11} r_{d,12t-k}, \tag{9}$$

$$r_{ft}^a = \sum_{k=0}^{11} r_{f,12t-k}, \tag{10}$$

where N is the annual simulation size.

The prior is (1). For the habit model, the scale factor used for ϕ and δ is 0.001 rather than the 0.1 shown in (1) to overcome an identification problem. The MCMC chain will not mix when the scale factor for ϕ and δ in (1) 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 MCMC chains will mix. Because of the first term of the right hand side of (1), (1) is not an independence prior; in simulations from the prior no correlation between parameters is zero.

Measures of location and scale for the prior and posterior distributions of habit model parameters are shown in the top panel of Table 2. The prior and posterior densities of the risk-free rate, equity premium, equity returns, and the standard deviation of equity returns are plotted in the left column of Figure 2. Overall, Table 2 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.

(Table 2 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 model, computed as described in Subsection 2.6, is plotted in the left column of Figure 3. The habit model predicts an end to the current recession in 2009 and return to steady-state growth by 2010. This is dictated by (2), (7), and (8) which imply that annual consumption growth for the habit model is a first order autoregression with an autogression parameter of about 0.25 (Working, 1960). Stock returns are predicted to be high in 2009 with a return to steady-state returns by 2013.

(Figure 3 about here)

1.2   The Long-run Risks Model

The driving processes for the long-run risks model are

$$\text{Consumption:} \quad c_{t+1} - c_t = \mu_c + x_t + \sigma_t \eta_{t+1},$$

$$\text{Dividends:} \quad d_{t+1} - d_t = \mu_d + \phi_d x_t + \pi_d \sigma_t \eta_{t+1} + \phi_u \sigma_t u_{t+1},$$

$$\text{Long-Run Risks:} \quad x_{t+1} = \rho x_t + \phi_e \sigma_t e_{t+1},$$

$$\text{Stochastic Volatility:} \quad \sigma_{t+1}^2 = \bar{\sigma}^2 + \nu(\sigma_t^2 - \bar{\sigma}^2) + \sigma_w w_{t+1},$$

$$\text{Random Shocks:} \quad (\eta_t, e_t, w_t, u_t) \sim \text{NID}(0, I).$$

The long-run risks model derives its name from the random shifts in the location of consumption and dividends due to x_t. Note that long-run risks x_t and stochastic volatility σ_t^2 evolve autonomously. The utility function is

$$U_t = \left[(1 - \delta) C_t^{\frac{\psi-1}{\psi}} + \delta \left(\mathcal{E}_t U_{t+1}^{1-\gamma} \right)^{\frac{\psi-1}{\psi(1-\gamma)}} \right]^{\frac{\psi}{\psi-1}}. \tag{11}$$

γ and δ have the same interpretation as for the habit model; ψ, which summarizes preferences across time periods, is a separate parameter. Separation of attitudes toward risk and preferences accross time is the main advantage of (11). $\mathcal{E}_t$ is the conditional expectation with respect to x_t and σ_t, which are the state variables.

The long-run risks model is richly parametrized

$$\theta = (\delta, \gamma, \psi, \mu_c, \rho, \phi_e, \bar{\sigma}^2, \nu, \sigma_w, \mu_d, \phi_d, \pi_d, \phi_u). \qquad (12)$$

It is so richly parametrized that identification would have to come from the prior even were data abundant because half of the models in Table 1 have fewer parameters than θ. The time increment is one month. Aggregation of monthly $\{c_t, r_{dt}, r_{ft}\}_{t=1}^{12N}$ to the annual frequency $\{c_t^a, r_{dt}^a, r_{ft}^a\}_{t=1}^{N}$ is by means of (7) through (10).

The prior is (1). The autoregressive parameters ρ and η cause problems. The solution method proposed by BY degrades as ρ and η increase from their published values. Because 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. Our solution to this problem is to set the scale factor for ρ and ν in (1) to 0.01 rather than 0.1 and attenuate the tails for ρ and ν above 0.995.

Measures of location and scale for the prior and posterior distributions of the parameters of the long-run risks model are shown in the middle panel of Table 2. The prior and posterior densities of the risk-free rate, equity premium, equity returns, and the standard deviation of equity returns are plotted in middle column of Figure 2. As for the habit model, Table 2 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 mean posterior forecast for the long-run risks model is plotted in in the middle column of Figure 3. 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 at their steady-state values over the entire forecast period. A flat response of asset returns to a consumption growth and asset return shock is a structural characteristic of the long-run risks model. It is due to the fact that stochastic volatility is autonomous and therefore not affected by consumption growth and asset returns shocks. The stochastic volatility process is the factor that affects the risk premium in the long-run risks model.

1.3 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 prospect theory model distinguishes between aggregate consumption $\bar{C}_t$, which is not a choice variable, and the agent's consumption C_t, which is. 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

$$X_{t+1} = S_t(R_{t+1} - R_f) \tag{13}$$

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

$$z_{t+1} = \eta \left(z_t \frac{\bar{R}}{R_{t+1}} \right) + (1 - \eta) \tag{14}$$

denote the benchmark level, where $\bar{R}$ is chosen to make median $\{z_t\} = 1$. The prospect theory utility function is

$$\mathcal{E}_0 \left[\sum_{t=0}^{\infty} \left(\rho^t \frac{C_t^{1-\gamma} - 1}{1 - \gamma} + b_0 \bar{C}_t^{-\gamma} \rho^{t+1} S_t \, \hat{v}(R_{t+1}, z_t) \right) \right], \tag{15}$$

where

$$\hat{v}(R_{t+1}, z_t) = \begin{cases} R_{t+1} - R_f & z_t \le 1, R_{t+1} \ge z_t R_f \\ (z_t R_f - R_f) + \lambda(R_{t+1} - z_t R_f) & z_t \le 1, R_{t+1} < z_t R_f \\ R_{t+1} - R_f & z_t > 1, R_{t+1} \ge R_f \\ \lambda(z_t)(R_{t+1} - R_f) & z_t > 1, R_{t+1} < R_f \end{cases} \tag{16}$$

and

$$\lambda(z_t) = \lambda + k(z_t - 1). \tag{17}$$

$\mathcal{E}_t$ is the conditional expectation with respect to the benchmark level z_t, which is the state variable. In (15) the agent's discount factor is ρ and risk aversion parameter is γ. The

11

second term of (15) is the utility from gains or losses, where b_0 is a scale factor. From a plot of $\hat{v}(R_{t+1}, z_t)$, Figure 1 of BHS, one can see that 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 pain is delayed until the "house's money has been lost". The parameter η in (14) 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 delay. See BHS for the relation of the prospect theory utility function to the psychology literature.

Like the long-run risks model, the prospect theory model is richly parameterized:

$$\theta = (g_C, g_D, \sigma_C, \sigma_D, \omega, \gamma, \rho, \lambda, k, b_0, \eta) \tag{18}$$

Identification requires the prior for most of the auxiliary models considered. The time increment is one year: Simulate directly and set $c_t^a = \log(\bar{C}_t)$, $r_{dt}^a = r_{dt}$, and $r_{ft}^a = r_{ft}$.

Measures of location and scale for the prior and posterior distributions of the parameters of the prospect theory model are shown in the bottom panel of Table 2. The prior and posterior densities of the risk-free rate, equity premium, equity returns, and the standard deviation of equity returns are plotted in the right column of Figure 2. As previously, Table 2 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 mean posterior forecast for the prospect theory model is plotted in the right column of Figure 3. The the prospect theory model predicts steady-state growth throughout the forecast period. This is dictated by the fact that annual consumption growth for the prospect theory model is iid. Stock returns are predicted to be double their steady-state value in 2009, reach steady-state by 2011, and remain at steady-state thereafter.

2 Inference for General Scientific Models

We describe the Bayesian methods proposed by GM and the modifications that we found necessary. Public domain code implementing the method for the auxiliary models in Table 1

and a User's Guide are available at `http://econ.duke.edu/webfiles/arg/gsm`.

2.1 Estimation of Scientific Model Parameters

Let the transition density of the scientific model be denoted by

$$p(y_t|x_{t-1}, \theta), \quad \theta \in \Theta, \tag{19}$$

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 but that we can simulate data from $p(\cdot|\cdot, \theta)$ for given θ. We presume that simulations from the scientific model are ergodic. We assume that there is a transition density

$$f(y_t|x_{t-1}, \eta), \quad \eta \in H \tag{20}$$

and a map

$$g : \theta \mapsto \eta \tag{21}$$

such that

$$p(y_t|x_{t-1}, \theta) = f(y_t|x_{t-1}, g(\theta)) \qquad \theta \in \Theta. \tag{22}$$

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 and g is called the implied map. When (22) holds f is said to nest p. Whenever we need the likelihood $\prod_{t=1}^n p(y_t|x_{t-1}, \theta)$, we use

$$\mathcal{L}(\theta) = \prod_{t=1}^n f(y_t|x_{t-1}, g(\theta)), \tag{23}$$

where $\{y_t, x_{t-1}\}_{t=1}^n$ are the data and n is the sample size. After substituting $\mathcal{L}(\theta)$ for $\prod_{t=1}^n p(y_t|x_{t-1}, \theta)$, standard Bayesian MCMC methods become applicable. That is, we have a likelihood $\mathcal{L}(\theta)$ from (23) and a prior $\pi(\theta)$ from (1) and need nothing beyond that to implement Bayesian methods by means of MCMC. A good introduction to these methods is Gamerman and Lopes (2006).

The difficulty is computing the implied map g accurately enough that the accept/reject decision in an MCMC chain (step 5 in the algorithm below) is correct when f is a nonlinear model. We describe the algorithms that we have found effective next.

Given θ, $\eta = g(\theta)$ 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 the 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 implied map is computed as

$$g : \theta \mapsto \underset{\eta}{\arg\max} \sum_{t=1}^{N} \log f(\hat{y}_t \mid \hat{x}_{t-1}, \eta). \tag{24}$$

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.

GM run a Markov chain $\{\eta_t\}_{t=1}^{K}$ of length K to compute $\hat{\eta}$ that solves (24). 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 GM propose, the $\hat{\eta}$ that corresponds to the maximum of $\sum_{t=1}^{N} \log f(\hat{y}_t \mid \hat{x}_{t-1}, \eta)$ 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 a multivariate specification of generalized autoregressive conditional heteroskedasticity (GARCH) that Engle and Kroner (1995) call BEKK. Likelihoods incorporating BEKK are notoriously difficult to optimize. We use $\hat{\eta}$ as a starting value and maximize (24) using the BFGS algorithm (Fletcher, 1987, 26–40). This also is not a sufficiently accurate evaluation of $g(\theta)$. A second refinement is necessary. The second refinement is embedded within the MCMC chain $\{\theta_t\}_{t-1}^{R}$

14

of length R that is used to compute the posterior distribution of θ. It is called the θ-chain. (We use $R = 25000$ past the point transients have dissipated.) 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 (23).

The prior may require quantities computed from the simulation $\{\hat{y}_t, \hat{x}_{t-1}\}_{t-1}^N$ used to compute (23). Our prior requires r_f^a. The sequence $\{\hat{r}_{ft}^a\}_{t-1}^N$ is available from the simulation. 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 scientific 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, proportional to a normal. Two aspects of the proposal scheme are worth noting. The first is that the wider the separation between the points in the support of q the less accurately $g(\theta)$ needs to be computed for α at step 5 of the algorithm below to be correct. As an example, the long-run risks model is not sensitive to the risk aversion parameter γ so that values of γ could be separated as much as 1/4 without making any difference to the usefulness of the θ-chain with respect to inference regarding economics. A practical constraint is that the separation cannot be much more than a standard deviation of the proposal density or the chain will not move. Our separations are typically 1/8 of a standard deviation of the proposal density. In turn, the standard deviations of the proposal density are typically no more than the standard deviations in Table 2 and no less than one order of magnitude smaller. The second aspect worth noting is that the prior is putting mass on these discrete points in proportion to $\pi(\theta)$. Because we never need to normalize

$\pi(\theta)$ this does not matter. Similarly for the joint distribution $f(y|x, g(\theta))\pi(\theta)$ considered as a function of θ. However, $f(y|x, \eta)$ must be properly normalized as a function of y, at least to the extent that (24) 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. 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 optima 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

$$\Sigma_\eta = \frac{1}{L} S_\eta. \tag{25}$$

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.

We compute posterior probabilities using a method that 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 $\alpha = \min\left(1, \frac{\pi(\theta^*, \psi^*)\, q(\theta^*, \theta^o)}{\pi(\theta^o, \psi^o)\, q(\theta^o, \theta^*)}\right)$.

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 2.3 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 $\alpha = \min\left(1, \frac{\pi_\kappa(\eta^*)\, q(\eta^*, \eta^o)}{\pi_\kappa(\eta^o)\, q(\eta^o, \eta^*)}\right)$.

2.2 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 2.1.

One computes the marginal density, $\int \prod_{t=1}^{n} f(y_t|x_{t-1}, g(\theta))\, \pi(\theta)\, d\theta$, for the three scientific 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 2.1. It is important, however,

that the auxiliary model be the same for all three models when the computations in Sub-section 2.1 are carried out. Otherwise the normalizing constant of f would be required. One divides the marginal 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 nests all p_i, i.e., if (22) holds, then the former and the latter are the same. If not, the matter needs consideration. In GM's 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 nests p. That is why they use the term statistical model for f. However, their argument is not a proof. We examine this issue in Section 4.

The realization 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 $\pi_1(\theta_1)$, $\pi_2(\theta_2)$, $\pi_3(\theta_3)$, allows us to perform a change of measure to η-space and illustrate relative model comparison graphically. In η-space, the prior π_i for model i restricts η to a manifold $\mathcal{M}_i = \{\eta \in \mathcal{H} : \eta = g_i(\theta_i), \ \theta_i \in \Theta_i\}$ with each $\eta = g_i(\theta_i)$ in $\mathcal{M}_i$ receiving prior weight $\pi_i(\eta) = \pi_i(\theta_i)$ (recall that Θ_i is discrete). Think of this manifold as a line in η-space. This is shown graphically in Figure 4 for two hypothetical models. Although conceptually a line, what is plotted in Figure 4 has area with gray fill to represent the density of the prior along the line. A density weighted integral along the line would have approximately the same value as an integral over the area shown. Also shown in Figure 4 are the likelihood contours of a hypothetical auxiliary model $f(y|x, \eta)$. The marginal density for model i is $\int_{\mathcal{M}_i} \prod_{t=1}^{n} f(y_t|x_{t-1}, \eta) \, \pi_i(\eta) \, d\eta$, which is approximately the integral over the gray areas for the hypothetical models in Figure 4. As shown in the figure, the likelihood is larger over $\mathcal{M}_2$ than over $\mathcal{M}_1$, which implies the same for the integrals over $\mathcal{M}_2$ and $\mathcal{M}_1$. Thus, the second model has the higher marginal likelihood and is

18

therefore preferred.

(Figure 4 about here)

2.3 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 scientific model $p(\cdot|\cdot, \theta)$ is to define the implied map $g(\theta)$ and the manifold

$$\mathcal{M} = \{\eta \in H \: : \: \eta = g(\theta), \; \theta \in \Theta\}. \tag{26}$$

The scientific model can be viewed as a sharp prior on f that restricts the posterior distribution of η to lie on the manifold $\mathcal{M}$. Think of it as a line in η-space. If this prior is relaxed, the line becomes a region with volume in η-space. Relaxation can be indexed by a scale parameter κ. As κ increases, the size of the region increases and posterior for η will move along a path toward the likelihood of the data under f. Figure 5 is an illustration. 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 scientific model. If waypoints far from $\mathcal{M}$ receive high posterior probability, then the data do not support the scientific model. The idea is illustrated graphically in Figure 5 for the case when a model is rejected and in Figure 6 for the case when a model is accepted. The formal development proceeds as follows.

(Figure 5 about here)

(Figure 6 about here)

We add the additional assumption that the auxiliary model is identified and has more parameters than the scientific model. This assumption implies that $g^{-1}(\eta)$ exists on $\mathcal{M}$. If the scientific model is identified $g^{-1}(\eta)$ will map to a single point; if not, $g^{-1}(\eta)$ will be a set.

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

$$\pi_\kappa(\eta) \propto \pi\left(g^{-1}(\eta^o)\right) \exp\left(-\frac{1}{2}(\eta - \eta^o)'(\kappa\Sigma_\eta)^{-1}(\eta - \eta^o)\right) \tag{27}$$

where

$$\eta^o = \underset{\hat{\eta} \in \mathcal{M}}{\operatorname{argmin}} \; (\eta - \hat{\eta})'(\Sigma_\eta)^{-1}(\eta - \hat{\eta}), \tag{28}$$

19

$\pi(\theta)$ is the prior (1) for the scientific model, and Σ_η is given by (25). It is easy and cheap to evaluate (28) once the computations described in Subsection 2.1 have been carried out because the implied map g is represented by pairs (θ, η) stored together with $\pi(\theta)$ at the conclusion of Subsection 2.1 computations. Store is traversed to find the pair (θ^o, η^o) such that η^o solves (28). Then $\pi(g^{-1}(\eta^o)) = \pi(\theta^o)$. (If $g^{-1}(\eta^o)$ maps to a set, $\pi(g^{-1}(\eta^o))$ is the sum of $\pi(\theta)$ over that set. Recall that Θ is discrete and normalization is not required.) The pairs (θ, η) and scale Σ_η used to compute (28) and (27) 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 the marginal likelihood $\int \prod_{t=1}^n f(y_t|x_{t-1}, \eta) \, \pi_\kappa(\eta) \, d\eta$.

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

2.4 The Auxiliary Model

The observed data are y_t for $t = 1, \ldots, n$. We first discuss the case where y_t is multivariate. 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, $x_{t-1} = (y_{t-1}, y_{t-2})$.

The data are modeled as

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

where

$$\mu_{x_{t-1}} = b_0 + B x_{t-1}, \tag{29}$$

which is the location function of a VAR, and $R_{x_{t-1}}$ is the Cholesky factor of

$$\Sigma_{x_{t-1}} = R_0 R_0' \tag{30}$$

$$+ Q \Sigma_{x_{t-2}} Q' \tag{31}$$

$$+ P(y_{t-1} - \mu_{x_{t-2}})(y_{t-1} - \mu_{x_{t-2}})' P' \tag{32}$$

$$+ \max[0, V(y_{t-1} - \mu_{x_{t-2}})] \max[0, V(y_{t-1} - \mu_{x_{t-2}})]'. \tag{33}$$

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 is the BEKK form of multivariate GARCH described in Engle and Kroner (1995) with an added leverage term (33). In computations, $\max(0, x)$ in (33) 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 (30) only, f_1 has terms (30), (31), and (32), and f_2 through f_5 have all four terms.

The density $h(z)$ of the iid z_t is the square of a Hermite polynomial times a normal density, the idea being that the class of such h is dense in Hellenger norm and can therefore approximate a density to within arbitrary accuracy in Kullback-Leibler distance (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 (29) and the variance function (30). It also causes the higher moments of $h(z)$ to depend on y_{t-1} as well.

The univariate auxiliary models are the same as the above but $\mu_{x_{t-1}}$ in (29) has dimension one and becomes the location function of a first order autoregression and $\Sigma_{x_{t-1}}$ in (30) has dimension one and becomes a GARCH(1,1) with a leverage term added.

2.5   Diagnostic Checks

The idea behind diagnostic checking is straightforward: If one has compared two scientific 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

$$t_i = \frac{\eta_{1i} - \eta_{2i}}{\sigma_{2i}}. \tag{34}$$

Table 4 is an example.

There is a caveat. The t_i are often 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 in 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.

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 comparative plots as in Figure 7.

2.6 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 $f(\cdot|\cdot, \eta)$ nests the scientific 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 scientific model and as function of θ. As such, it can be 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. An example is Figure 3.

3 Habit, Long-Run Risks, Prospect?

Table 3 presents the posterior probabilities for a relative model comparison and an absolute model assessment of the three asset pricing models fitted to the three data series over the two sample periods.

(Table 3 about here)

For the univariate series, all three models fit reasonably well and none of them strongly dominates. Our interest in the univariate case is due to the fact that this conclusion depends

on the choice of auxiliary model. We explore this issue in Section 4.

Our interest in the trivariate series stems from fact that BGT, Beeler and Campbell (2008), and Bansal, Kiku, and Yaron (2009) argue that the main difference between the habit model and the long-run risks model occurs with respect to predictability regressions that involve the price to dividend ratio series.

It is difficult to properly adjust dividend payouts for stock repurchases and other distortions caused by tax policy so as to make the measured data resemble the theoretical construct of asset pricing models. It is apparent from Figure 1 that the price to dividend series does not look like the realization of a stationary process. An informal check on the ability of a model to explain dividends is to simulate it for ten thousand years and see if the best match (with respect to a variance weighted mean squared error metric) of any contiguous segment of the simulation to the data looks like Figure 1. It does not. For all models and for both sample periods, the best match cannot produce the hump seen in the third panel of Figure 1.

As mentioned earlier, because the prospect theory model puts its (conditional) mass on a two-dimensional subspace, it cannot be fit to the trivariate series. However, the informal check just described is possible and the prospect theory model fails more definitively under it than either the habit or long-run risks model.

With respect to the remaining two models, in the relative model comparison presented in Table 3, the long-run risks model dominants over the 1930–2008 period whereas the habit model dominants over the 1950–2008 period. Both models fail to fit the data in the absolute model assessment. Interestingly, the failure of the habit persistence model in the 1950–2008 period is not as stark as for the long-run risks model. These results are for the nesting model f_0.

The fundamental objective of asset pricing models is to explain the relationship between asset prices and consumption. From this perspective, dividends can be treated as an internal construct that need not have any corresponding observable counterpart. Therefore, the bivariate results presented in Table 3 are the most important substantively.

As seen in Table 3, in the relative model comparison, the long-run risks model is dominant over the 1930–2008 period whereas the habit model is dominant over the 1950–2008 period. In the absolute model assessment, the habit model fails in the 1930–2008 period and the

prospect theory model fails in the 1950–2008. These results are for the nesting model f_5.

It is of interest to determine why the habit persistence model fails to fit the bivariate series using the diagnostic checks described in Subsection 2.5. For this purpose it is more informative to use the simpler auxiliary model f_1 rather than the nesting model f_5. Because the habit model has seven parameters and f_1 has twelve, this is a legitimate choice for the habit model.

Table 4 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 series over the period 1930–2008 and over the period 1950–2008. It is clear from the table what the problem is. The largest diagnostic, $t = -4.98$, indicates that P_{11}, which is the feedback of consumption growth into its own volatility, is too small in absolute value. The habit model fails to put enough conditional heteroskedasticity into the consumption growth process. The problem disappears in the 1950–2008 period because, as seen by comparing the entries for $\kappa = 10$ across the row for P_{11}, the conditional volatility in the data drops substantially.

(Table 4 about here)

Figure 7 plots the conditional volatility of the three models over the 1930–2008 period. The solid line in Figure 7 is the long-run risks model, which is the most correct of the three according to the relative model comparison in Table 3. The shaded region shows ± 1.96 posterior standard deviations about this line. All three models track the conditional volatility of stock returns about the same after taking the standard deviations into account. Where they differ is in how they track the conditional volatility of consumption growth and the conditional correlation between consumption growth and stock returns. The same plot for the 1950–2008 period (not shown) looks qualitatively similar: agreement in the middle panel and disagreement in the other two.

(Figure 7 about here)

Plots of the conditional mean of consumption growth (not shown) indicate that the habit model and long-run risks model agree over the 1930–2008 period except over the years 1930–1940. They both track the data moderately well. Because, as was seen in Figure 3, the conditional mean of consumption growth for the prospect theory model must plot as a straight line, it does not track well. All models disagree in plots (not shown) of the

24

conditional mean of stock returns, with long-run risks plotting as a straight line; all track the data poorly.

4   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. The exact specification one gets using standard model selection procedures, such as upward F-testing, is sensitive to the sample period used. One can get slightly richer or coarser specifications. It is fair to say that the consensus view is that a one-lag VAR location, GARCH scale, and normal innovations is the richest model one ought to entertain, which is model f_1 of Table 1.

We do not discuss the trivariate series in this section other than to remark that for it, f_0 is the nesting model, results for f_0 are reported in Table 3, conclusions do not change under auxiliary models f_1 through f_3, and computations for models f_4 through f_5 cannot be undertaken because simulations do not identify them (BFGS becomes unstable). Similarly, we do not discuss results for the for the bivariate consumption growth and stock returns data because the results shown in Table 3 are the same for all models in Table 1. Finally, we do not discuss absolute model assessment because there is no logical requirement that the auxiliary model nest the scientific model for the purpose of absolute model assessment. All that is logically required is that the auxiliary model have more parameters than the scientific model. Other than that, one is free to choose an auxiliary model as judgment suggests.

For the univariate (and bivariate) series, a model that will nest the three scientific 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. GM found the same to be true for the habit model, except that they 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 f_0 and f_1. They did verify that a fat tailed innovation distribution did not change results.

Using model f_0 most closely corresponds to calibration as customarily implemented in the macro/finance literature. The sufficient statistics for f_0 are the mean and variance of y_t and the first order autocorrelations. One is, effectively, finding parameter values that best match three moments for univariate data, nine for bivariate, and eighteen for trivariate.

As discussed in Subsection 2.1 and in GM, the logically correct view toward using f_1, which fits the data, instead of f_5, which nests the scientific model, is that it is not the likelihood of the scientific model that is being used. It is some other likelihood. Therefore it is not the scientific models that are actually being estimated and compared. Another point of view is the argument advanced by GM that using a sensible auxiliary model is akin to method of moments estimation. One only asks that the scientific models match certain features of the data and allows them to ignore others. What to do? About all one can do is try a battery of auxiliary model specifications and see what happens.

Table 5 displays the results for the relative comparisons for the univariate stock returns data over the periods 1930–2008 and 1950–2008. There is considerable sensitivity to specification of the auxiliary model over the 1930–2008 period. Conclusions are affected by the choice of auxiliary model. Our view is that, because there can be sensitivity to auxiliary model choice, and because one is not actually comparing scientific models if the auxiliary model is not nesting, it is best to use the nesting auxiliary model in general, which is f_5 in this instance.

(Table 5 about here)

5   Conclusion

We used Bayesian statistical methods proposed by GM to compare the habit persistence asset pricing model of CC, the long-run risks model of BY, and the prospect theory model of BHS. This comparison fills a void in the literature.

We undertook two types of comparisons, relative and absolute, over two sample periods, 1930–2008 and 1950–2008, using three series, trivariate (consumption growth, stock returns, and the price to dividend ratio), bivariate (consumption growth and stock returns), and univariate (stock returns). The prior for each model is that the ergodic mean of the real interest rate be 0.896 within ± 1 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 dominates over the 1930–2008 period, while the habit persistence model dominates over the 1950–2008 period; in the absolute assessment, the habit model fails in the 1930–2008 period and the prospect theory model fails in the 1950–2008 period.

For the trivariate series, in the relative comparison the long-run risks model dominates over the 1930–2008 period, while the habit persistence model dominates over the 1950–2008 period; in the absolute assessment, both the habit model and the long-run risks model fail in both periods. The prospect theory model cannot be fitted to a trivariate series because it puts its (conditional) mass on a two-dimensional subspace.

The estimator proposed by GM is a simulation based estimator. Simulations from a scientific model, which here is either the habit model, the long-run risks model, or the prospect theory model, are used to determine a map $\eta = g(\theta)$ from the parameters θ of the scientific 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 lags. 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 nest the scientific model. GM argue that one is better served by an auxiliary model that represents the data well. We undertook a sensitivity analysis and recomputed our results for six auxiliary models ordered by complexity. The first produces estimates that mimic values obtained by methods customarily employed in macro/finance. The second represents the data. The sixth nests the three scientific models considered. We find that results can be sensitive to the choice of auxiliary models. Most importantly, results can differ between the model that represents the data well and the model that nests the scientific model. In view of this difference and the fact that theory supports the latter, our view is that the nesting auxiliary model ought to be used. Our substantive conclusions are based on the nesting model.

We found that the computational methods that GM proposed are not sufficiently accurate to compare the habit, long-run risks, and prospect theory models. A contribution of this paper is a refinement of GM's methods that allows comparision of these three models.

6 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 2009. "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *National Bureau of Economic Research,* W15504

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

Beeler, J., and J. Y. Campbell. 2008. "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," *National Bureau of Economic Research,* W14788

Campbell, J. 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.

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

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

Gallant, A. R., and R. E. McCulloch. 2009. "On the Determination of General Statistical Models with Application to Asset Pricing." *Journal of the American Statistical Association 104,* 117–131.

Gallant, A. R., and D. W. Nychka. 1987. "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica 55,* 363–390.

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

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.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics 6,* 461–464.

Working, Holbrook. 1960. "Note on the Correlation of First Differences of Averages in a Random Chain," *Econometrica 28,* 916–918.

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
Parms trivar	18	22	25	37		

Multivariate 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 2.4. Parms is the number of parameters, which depends on the dimension of the data: univariate, bivariate, or trivariate. The habit persistence model has 7 parameters, the long-run risks model has 13, and the prospect theory model has 11.

Table 2. Prior and Posterior Model Parameters

Parameter	Prior			Posterior		
	Mode	Mean	Std.Dev.	Mode	Mean	Std.Dev.
Habit Model						
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
Long-Run Risks Model						
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
Prospect Theory Model						
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 monthly frequency for the habit and long-run risks models and for the annual frequency for the prospect theory model. Mode is that of the multivariate density. Returns are geometric, annualized, and expressed as a percent for all models. In the data, returns are $r_d - r_f = 5.59 - 0.89 = 4.7$ 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, see Figure 1.

Table 3. Posterior Probabilities

Relative Model Comparison

| | Sample Period | | | | | |
| | 1930–2008 | | | 1950–2008 | | |
Series	hab	lrr	pro	hab	lrr	pro
Trivariate	0.00	1.00		1.00	0.00	
Bivariate	0.00	1.00	0.00	1.00	0.00	0.00
Univariate	0.28	0.48	0.24	0.44	0.42	0.14

Absolute Model Assessment

| | | Sample Period | | | | | |
| | | 1930–2008 | | | 1950–2008 | | |
Series	Prior	hab	lrr	pro	hab	lrr	pro
Trivariate	$\kappa = 0.1$	0.00	0.00		0.00	0.00	
	$\kappa = 1$	0.00	0.00		0.33	0.00	
	$\kappa = 10$	1.00	1.00		0.67	1.00	
Bivariate	$\kappa = 0.1$	0.00	0.41	0.28	0.31	0.16	0.08
	$\kappa = 1$	0.00	0.36	0.28	0.31	0.21	0.08
	$\kappa = 10$	1.00	0.23	0.44	0.38	0.64	0.84
Univariate	$\kappa = 0.1$	0.29	0.36	0.10	0.40	0.39	0.29
	$\kappa = 1$	0.30	0.26	0.30	0.38	0.35	0.34
	$\kappa = 10$	0.41	0.38	0.60	0.22	0.26	0.37

Shown are posterior probabilities for a relative comparison and an absolute assessment of three asset pricing models fitted to three data series over two sample periods. The trivariate series is annual consumption growth, stock returns, and the price dividend ratio over the years shown. The bivariate series is consumption growth and stock returns. The univariate series is stock returns alone. Variables not in a series are treated as latent in model fits. For the univariate and bivariate series. the auxiliary model is f_5, which is described in Table 1; for the trivariate series it is f_0. 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 4. Diagnostics for the Habit Persistence Model

	1930–2008			1950–2008		
Parameter	Mode $\kappa = 0.1$	Mode $\kappa = 10$	Diag-nostic	Mode $\kappa = 0.1$	Mode $\kappa = 10$	Diag-nostic
$b_{0,1}$	-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 2.5.

Table 5. Posterior Probability, Relative Comparison, Stock Returns

Model	f_0	f_1	f_2	f_3	f_4	f_5
			1930–2008			
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
			1950–2008			
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 over years shown. Auxiliary models f_0 through f_5 are described in Table 1.
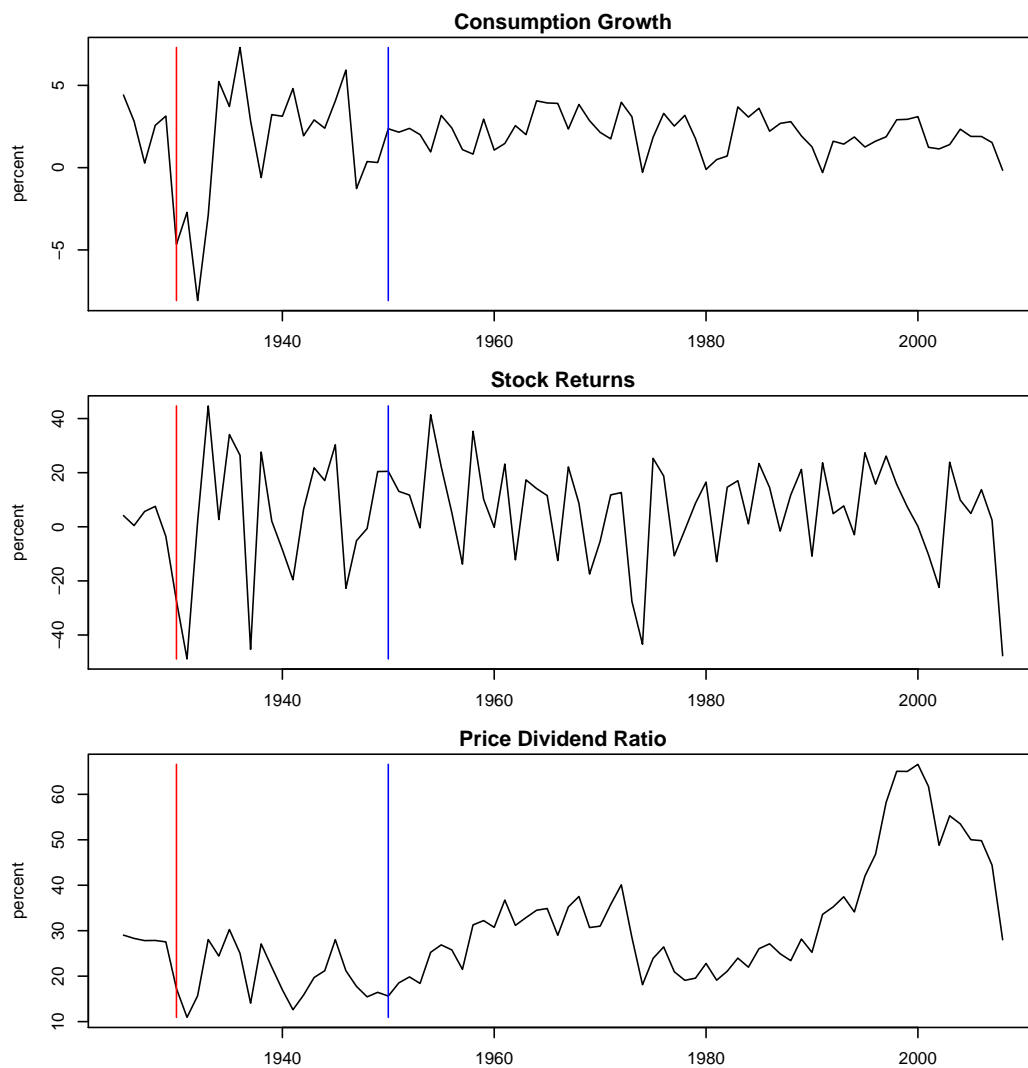
Figure 1. Consumption Growth, Stock Returns, and Price to Dividend Ratio, 1925–2008 The left vertical line is at 1930 and the right at 1950. The data collection protocol is as described in Bansal, Gallant, and Tauchen (2007) for the period 1930–2008. The data from 1925–1929, which are only used to prime recursions, are the inflation adjusted Dow-Jones industrial average, a real U.S. consumption growth series kindly supplied by Robert Barro, and from back-casted real price and dividend levels.
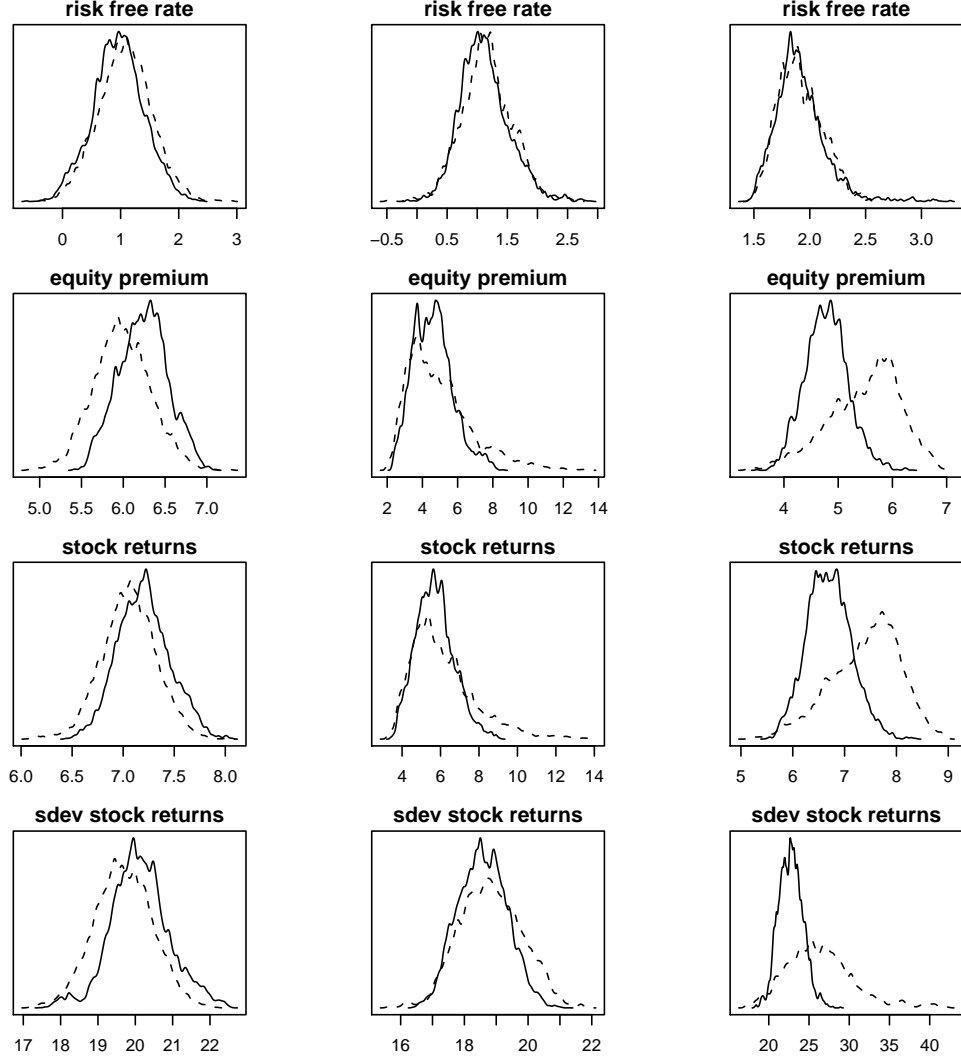
34

Figure 2. Prior and Posterior Density Estimates. The dashed line is the prior. The solid line is the posterior. Left column is for the habit model, middle for the long-run risks model, right for the prospect theory model. Returns are geometric, annualized, and expressed as a percent. Other details as in Table 2. Bandwidths are small to reduce smudging of isolated, peaked modes.
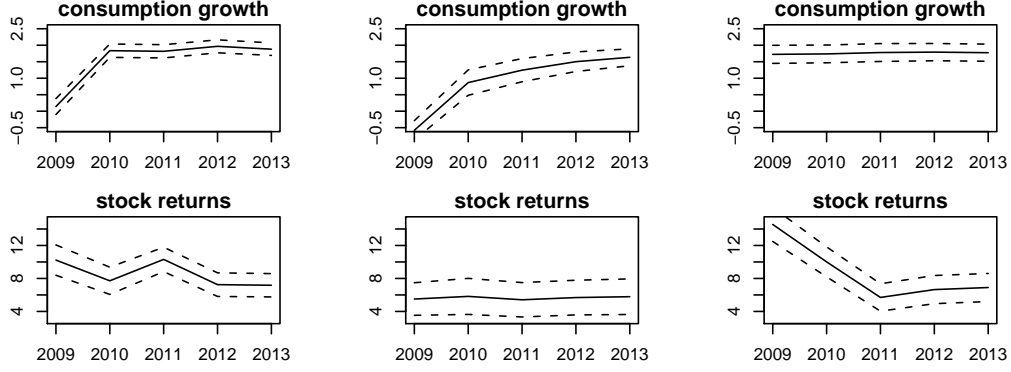
Figure 3. Prior and Posterior Forecasts. The solid line is the mean of the posterior annualized and expressed as a percent. The dashed lines are ± 1.96 posterior standard deviations. The left column is for the habit model, middle for the long run risks model, and left for the prospect theory model. Other details as in Table 2.
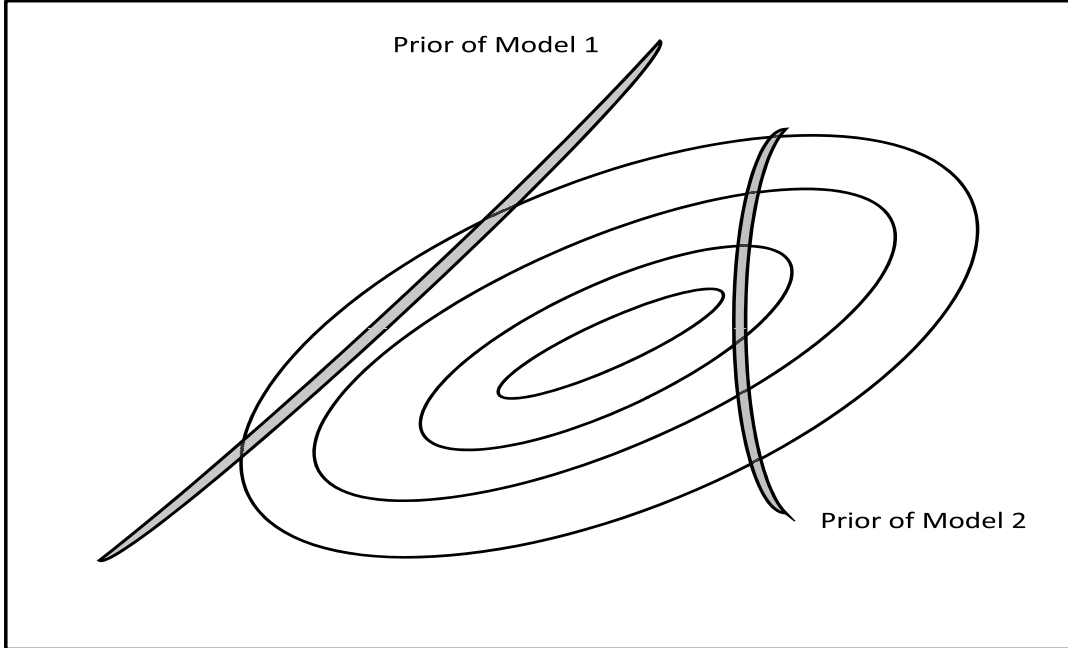

Figure 4. Relative Model Comparison. Shown is relative model comparison under a change of variables of integration $\theta \mapsto \eta$. The contours show the likelihood of the auxiliary model $f(\cdot \mid \eta)$. The curved lines, labeled "Prior of Model i", show the manifolds $\mathcal{M}_i = \{\eta \in H : \eta = g_i(\theta_i),\ \theta_i \in \Theta_i\}$ for Model 1, $i = 1$, and Model 2, $i = 2$. Thickness, represented by gray fill, is proportional to the priors π_1 and π_2. Posterior probabilities are proportional to the integral of the likelihood over the manifold weighted by the prior. Model 2 is preferred because the likelihood values on its manifold are larger than the likelihood values on the manifold of Model 1.
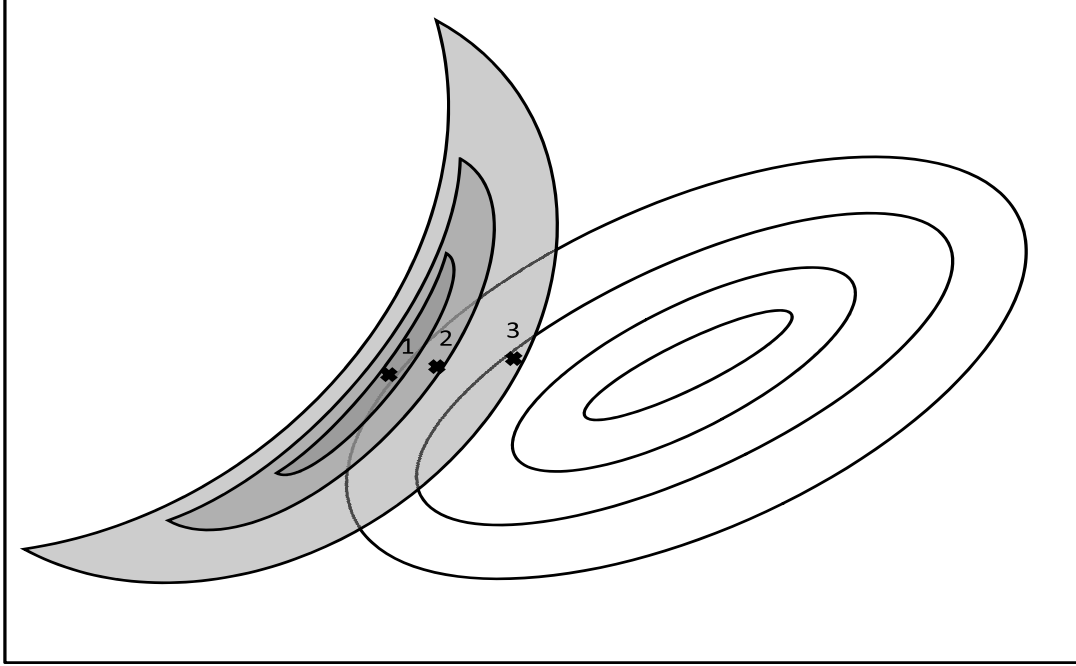
Figure 5. Absolute Model Assessment – Reject. The contours show the likelihood of the auxiliary model $f(\cdot \mid \eta)$. The shaded areas show the prior (27) for scale parameters $\kappa_1 < \kappa_2 < \kappa_3$. The smallest region corresponds to κ_1 and the largest to κ_3. The crosses show the mode of the posterior under κ_1, κ_2, κ_3. The posterior probabilities for absolute model assessment are proportional to the integrals of the likelihood over the respective shaded areas. The model is rejected because the likelihood, hence the integral, is larger over the κ_3-prior than over the κ_1-prior.
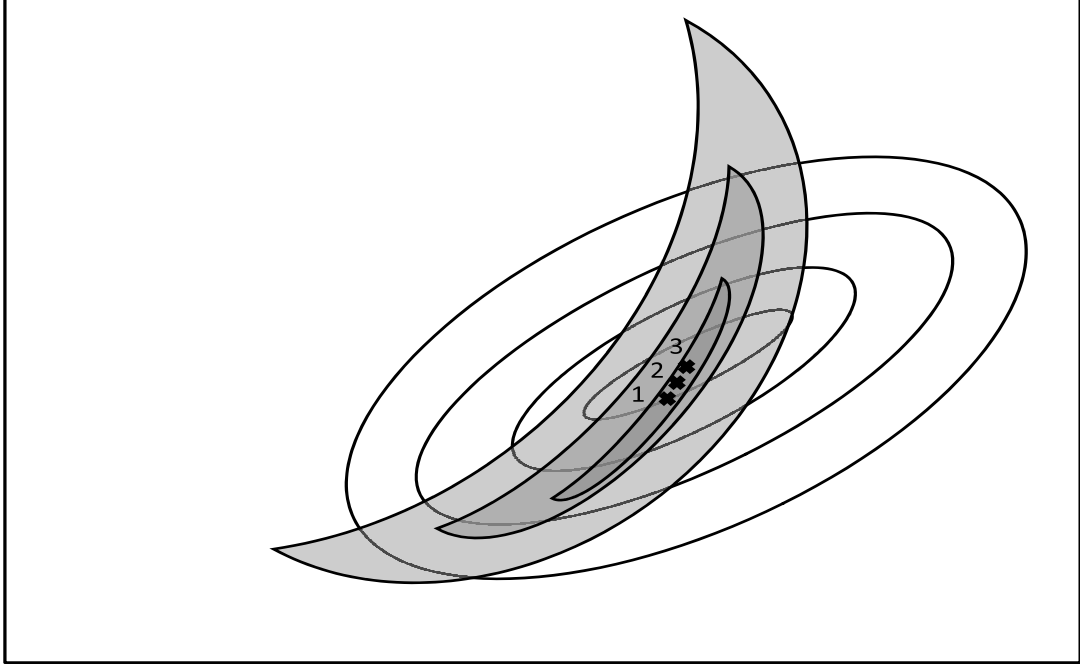
Figure 6. Absolute Model Assessment − Accept. The contours show the likelihood of the auxiliary model $f(\cdot\,|\,\eta)$. The shaded areas show the prior (27) for scale parameters $\kappa_1 < \kappa_2 < \kappa_3$. The smallest region corresponds to κ_1 and the largest to κ_3. The crosses show the mode of the posterior under κ_1, κ_2, κ_3. The posterior probabilities for absolute model assessment are proportional to the integrals of the likelihood over the respective shaded areas. The model is accepted because the likelihood, hence the integral, is larger over the κ_1-prior than over the κ_3-prior.
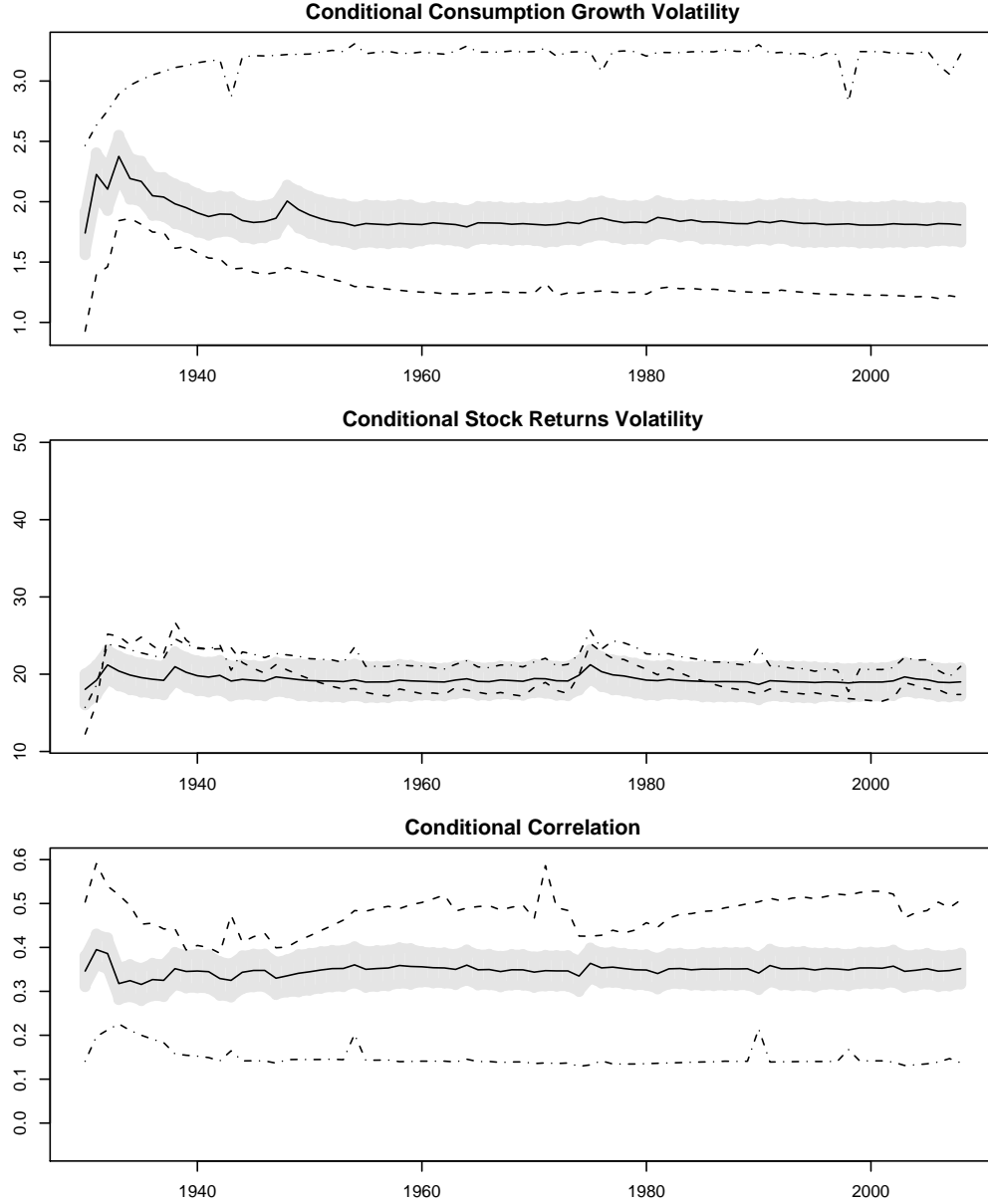
Figure 7. Conditional Volatility of the Three Models. The solid line is the conditional volatility of the data expressed as a percent for the long-run risks model with its parameters set to the posterior mode from fitting to the bivariate consumption growth and stock returns data over the period 1930–2008 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. The shaded area shows ± 1.96 posterior standard deviations about the conditional volatility of the long-run risks model.