

On the Determination of General Scientific Models
with Application to Asset Pricing ¹

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

Robert E. McCulloch
Graduate School of Business
University of Chicago
Chicago IL 60637-6040 USA

November 2003
This draft June 2005

Abstract

We consider a consumption based asset pricing model that uses habit persistence to overcome the known statistical inadequacies of the classical consumption based asset pricing model. We find that the habit model fits reasonably well and agrees with results reported in the literature if conditional heterogeneity is suppressed by a sharp prior but that it does not fit nor do results agree if conditional heterogeneity, well known to be present in financial market data, is allowed to manifest itself. We also find that it is the preference parameters of the model that are most affected by the presence or absence of conditional heterogeneity, especially the risk aversion parameter. The habit model exhibits exhibits four characteristics that are often present in models developed from scientific considerations: (1) a likelihood is not available; (2) prior information is available; (3) a portion of the prior information is expressed in terms of functionals of the model that cannot be converted into an analytic prior on model parameters; (4) the model can be simulated. The underpinning of our approach is that, in addition, (5) a parametric statistical model for the data, determined without reference to the scientific model, is known. In general one can expect to be able to determine a model that satisfies (5) because very richly parameterized statistical models are easily accommodated. We develop a computationally intensive, generally applicable, Bayesian strategy for estimation and inference for scientific models that meet this description together with methods for assessing model adequacy. An important adjunct to the method is that a map from the parameters of the scientific model to functionals of the scientific and statistical models becomes available. This map is a powerful tool for understanding the properties of the scientific model.

Keywords: Scientific models, simulation, Bayes, MCMC, estimation, inference, asset pricing.

1 Introduction

This article is motivated by an asset pricing application, namely the habit persistence asset pricing model, that has characteristics in common with all modern general equilibrium models. And this article contributes to that empirical literature. But many models derived from scientific considerations exhibit the same characteristics that make statistical analysis of the habit model challenging. These characteristics are (1) a likelihood is not available; (2) prior information is available; (3) a portion of the prior information is expressed in terms of functionals of the model that cannot be converted into an analytic prior on model parameters; (4) the model can be simulated. Additional examples of such models are the SEIR model from epidemiology (Olsen and Schaffer, 1990) and compartment models from pharmacokinetics (Mallet, Mentré, Steimer, and Lokiec, 1988).

We outline a general approach for the analysis of such models and then apply it to the habit model.

In some instances other methods are available. For instance, if the only cause of difficulties is a small number of latent variables, then a data augmentation approach will likely be applicable and be less computationally intensive than the methods proposed here. We are not concerned with such models. Our concern is with models such as our application where there seems to be little else available other than what we propose here. This becomes doubly true when data are sparse, as in our application, so that serious use of prior information becomes essential. Our proposals are especially helpful if, as in our application, some prior information may be expressed only in terms of functionals of the model. The methods proposed here generate ancillary information that can help to interpret the scientific model in terms of its statistical properties and to cast model inadequacies into sharp relief.

Our approach depends on an assumption that (5) an adequate statistical model for the data is available. Because richly parameterized statistical models are admissible in this connection and the statistical model only has to be fit to large simulations from the scientific model, (5) can nearly always be satisfied. Briefly, our proposal is as follows: Given (5), we can construct a map from the parameters of the scientific model to those of the statistical model such that a point in the parameter space of the scientific model and its image under

the map both correspond to the same data generating process. Typically the parameters of the statistical model will live in a higher dimensional space than that of the scientific model. The scientific model may therefore be viewed as a prior on the statistical model that has support entirely on the manifold that is the image of the map. Scientific prior information will then generate preferences on the manifold. The methodology developed here allows the scientific prior information to be expressed either directly on the parameters of the scientific model or on functionals of the scientific model that can be evaluated via simulation.

The discovery of the mapping from the parameters of the scientific model to those of the statistical model, which is an intermediate step of the methods proposed here, is often itself of scientific interest. For instance, the statistical model must, perforce, be expressed entirely in terms of observables whereas scientific models often contain unobservables. Having a mapping from the subset of the parameters that control the unobservable features of the scientific model to the parameters of a statistical model consisting entirely of observables can be extremely helpful in understanding the observable consequences of changes in a model's unobservable internal structure. The utility of this approach can be extended by using the same methods to find the map from the parameters of the scientific model to functionals of both the scientific and statistical models.

A Bayesian approach suggests itself for problems that exhibit the five characteristics just listed because the methodology gracefully accepts prior information into the analysis and, for dynamic models, does not require growth conditions on model output or data that are often counterfactual. Moreover, the estimates of parameter uncertainty are credible. That is, the asymptotics on which frequentist methods rely are often grossly inaccurate for the class of problems considered here (Duffee and Stanton, 2005) and bootstrap methods are so computationally burdensome as to be infeasible. Nonetheless, comparison, when possible, is of interest and in the application we contrast with results from a frequentist method (Gallant and Tauchen, 1996) that relies on asymptotics and that is in general use for estimating asset pricing models. As will be seen, sparse data forces a simplification on the frequentist approach that precludes discovery of our key findings.

Although we know of nothing in print, several people, notably Anthony A. Smith, Jr., Yale University, and Alan E. Gelfand, Duke University, having seen this work presented, told

us that they have had similar thoughts along the lines of the use of a statistical model to synthesize a likelihood as we do here, but either did not attempt or did not succeed in making them practicable. Our implementation relies on modern object oriented programming methods, modern data structures, and a discretization at a critical point in the computations. Bringing these elements to bear on the problem seems to be both novel to this work and essential to success. There is a related frequentist literature that is logically distinct from the ideas here but does make use of an auxiliary model as an adjunct to estimation and inference; see Smith (1993), Gouriéroux, Monfort, and Renault (1993), and Gallant and Tauchen (1996). We believe that our proposals for model assessment are new although there is a related literature wherein a prior is imposed on a statistical model by numerical methods that are akin to appending a simulation from the scientific model to the data in the style of Theil and Goldberger (1961); see Del Negro and Schorfheide (2004) and the references therein. In contrast to this literature, our proposals allow the imposition of priors that are more general and that fit logically within the classical Bayesian paradigm.

2 Scientific and Statistical Models

We shall use the notational conventions of time series analysis because most models of the sort considered here are dynamic. This is in no way essential because the results apply equally well to other data structures with a few obvious changes to notation.

Let the transition density of the scientific model be denoted as

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

where $x = (y_{t-1}, \dots, y_{t-L})$ if Markovian and $x = (y_{t-1}, \dots, y_1)$ if not. Our basic assumption is that there is no direct information about $p(\cdot|\cdot, \theta)$. All that we can do is simulate data from the model for given θ . If the model produces ergodic output, then a single long simulation for each setting of θ suffices for our purposes. If not, then many independently simulated replicates of the data are used.

The scientific model is built using subject matter knowledge. Thus, we expect that real prior information is available. This prior information may be expressible either in terms of elements of θ or in terms of characteristics ψ of the process. An element of ψ might be an

unconditional moment of a latent variable or the unconditional moment of the solution of a system of nonlinear conditional moment equations. In general, ψ can be regarded as a point in the range of a vector of functionals $\Psi : p(\cdot|\cdot, \theta) \mapsto \psi$ that is computable from a simulation. We will capture both of these types of information.

A key motivation for implementing a Bayesian approach to the problem is the importance of using prior information. This can be critical when data are sparse. When data are sparse, prior information can be used to fill in model features about which the data says little but the literature says much thereby enabling extraction of features about which the data are informative.

To compute a posterior, we need a likelihood. Since we do not have access to $p(\cdot|\cdot, \theta)$, there is no direct way to compute the likelihood. Our approach is to find a parametric family of distributions that is capable of representing the process $\{y_t\}$. Specifically, we assume that there is a transition density $f(y_t|x_{t-1}, \eta)$ and that there is a map $g : \theta \mapsto \eta$ such that

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

and, importantly, that the form of $f(\cdot|\cdot, \eta)$ is known. When we need a likelihood based on the unknown $p(\cdot|\cdot, \theta)$, we substitute $f(\cdot|\cdot, \eta)$ with the appropriate η . In most applications the dimension of η will be larger than that of θ thus restricting η to a manifold. There is no loss of generality in assuming this to be the case because this manifold \mathcal{M} can be the entire parameter space H of $f(\cdot|\cdot, \eta)$.

The model $f(\cdot|\cdot, \eta)$ is a statistical description of the observed data that is entirely divorced from scientific considerations. Thus we call it the statistical model. Usually this model will be known from the literature. In other cases it must be determined as part of the analysis. As richly parameterized flexible functional forms are admissible, success in finding an acceptable statistical model can be anticipated. It is to be emphasized that we only use the statistical model to fit large simulations from the scientific model (Section 3) or when augmented by a strong prior dictated by the scientific model (Section 4) so that the fact that the data may be too sparse to support it is not a consideration.

In this connection we call attention to the SNP model (Gallant and Nychka, 1987) which is a flexible functional form that is expressly designed to be a convenient adjunct to maximum

likelihood estimation. An example of its use in a dynamic situation is Chernov, Gallant, Ghysels, and Tauchen (2003); cross sectional examples are Leon, Tsiatis, and Davidian (2003) and Song, Davidian, and Tsiatis (2002).

In the formal logical structure adopted here, the analysis is carried out in η -space which is to say that the scientific model and its prior are viewed as placing a prior on η via the map $\eta = g(\theta)$ with the support of the prior contained in the manifold \mathcal{M} . (We are indebted to James Berger, Duke University, for this interpretation.) Of course, one is not thereby precluded from imposing additional prior information directly on η . Because it may at times be useful to be able to express this prior information in terms of functionals of the statistical model that can only be computed from a simulation, we introduce the notation $\Upsilon : f(\cdot|\cdot, \eta) \mapsto v$ to denote such a vector of functionals. Some elements of ψ for the scientific model and v for the statistical model can be the same, e.g. the moment of an observable. Logically, elements of ψ and v could also be functions of the parameters θ or η . Note also that when we impose prior information expressed in terms of η and v we are also indirectly imposing prior information on the parameters θ of the scientific model by changing the relative weighting of points on the manifold \mathcal{M} .

Figure 1 about here

The situation just described can be summarized graphically. Consider the simplest example: θ has dimension one, η has dimension two, the scientific model is $p(y|\theta) = n(y; \theta, \theta^2)$, and the statistical model is $f(y|\eta) = n(y; \eta_1, \eta_2)$, where $n(y; \mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2 . The mapping of the parameter θ of the scientific model to the parameters (η_1, η_2) of the statistical model is $g : \theta \mapsto (\theta, \theta^2)$. Figure 1 displays this example in η -space. We will call this the tinker toy example hereafter.

In each panel of Figure 1 the solid line is the image of the map $\eta = g(\theta)$. This is the manifold of $\eta = (\eta_1, \eta_2)$ such that $\eta_2 = \eta_1^2$. Fifty observations were simulated from the scientific model with $\theta = 2$ ($\eta_1 = 2, \eta_2 = 4$). The contours of the likelihood for these fifty observations under the statistical model are shown as dotted lines in the two panels on the left. In the two panels on the right, the dotted contours display the likelihood from fifty observations simulated from the statistical model with $\eta = (2.8, 4)$. Thus, on the left the

scientific model is supported by the data as evidenced by the manifold passing near the mode of the likelihood whereas on the right it is not as evidenced by the manifold missing the region of high likelihood.

The posterior for η under the scientific model, computed using methods described in Section 3, is also confined to the manifold and also weights points on the manifold. Draws from the posterior are shown in Figure 1 as dots. Since all draws must lie on the manifold, we "jittered" the dots by adding a bit of noise so that the dots may be seen thereby making the portions of the manifold with high posterior probability appear to be thicker in the plot. As with the prior, the posterior is expressed as a distribution over Θ that is transferred to η -space via the map $\eta = g(\theta)$. An advantage of this representation is that it provides posteriors on both the parameters θ of the scientific model and on the parameters η of the statistical model.

We may not want to impose the belief that the scientific model holds exactly. We can capture this idea by constructing a prior that expresses a preference for η that are close to the manifold. Our prior construction uses a single parameter that we call κ to control prior beliefs about how close η should be to the manifold (Section 4). The smaller κ is, the more prior weight is placed on η close to the manifold.

This prior is shown in Figure 1 by the shading, with darker shading indicating more prior weight on the corresponding η . The top two panels have a small κ so that the prior weight is on η close to the manifold. The bottom two panels have a larger κ so that η farther from the manifold get more weight. The solid contours in the panels of Figure 1 display the posteriors computed using the prior indicated by the shading and the data indicated by the dotted contours. In the left panels, where the scientific model is true, relaxing the prior (increasing κ) changes the shape of the posterior but there is no dramatic shift in its location. In the right two panels, relaxing the prior allows the posterior to move away from the manifold in search of the likelihood.

Figure 2 about here

Our κ indexed prior gives us a way to assess the scientific model. If the scientific model is correct, increasing κ should cause the posterior to spread out but not shift dramatically.

If the scientific model is incorrect, increasing κ should allow the posterior to shift away from the manifold. In high dimensional situations this shift can be seen by looking at interpretable posterior marginals of η or, more generally, posterior marginals of functionals Υ of the statistical model. Reasonable values of κ may be elicited by looking at interpretable prior marginals of Υ for various κ .

These ideas can be simply illustrated in the tinker toy example by considering the marginal posterior of the coefficient of variation $\Upsilon : f(\cdot|\cdot, \eta) \mapsto v = \frac{\eta_1}{\sqrt{\eta_2}}$. Under the tinker toy scientific model the coefficient of variation is one with probability one. The four panels of Figure 2 correspond to the four panels of Figure 1. In Figure 2 the dotted curves are the densities of the prior marginal of v while the solid curves are the posterior densities. In the left two panels we see that as the prior is relaxed (κ increases) the posterior spreads out but does not shift away from the true value which is one (the value consistent with the scientific model). In the right two panels, the posterior shifts as κ increases.

At this point, the main conceptual ideas that we shall propose have been set forth. The devil is in the details, to which we now proceed. The reader who would rather see our substantive results first can skip to Section 5.

3 Bayesian Estimation of Scientific Models

We have two cases to consider: The first is when the scientific model is forced upon the statistical model and η is on the manifold \mathcal{M} (c.f. Figure 1). The second is when η is merely attracted to the manifold. We will consider the first case in this section and the second in Section 4. In the first case, it is both equivalent and more convenient to describe the computations in terms of the posterior for θ .

3.1 Computing the Map

Given θ , how do we find the corresponding η ? How do we uncover the map $g : \theta \mapsto \eta$ that satisfies

$$p(y|x, \theta) = f(y|x, \eta) \quad \forall (y, x) \tag{1}$$

when all that we can do is simulate data $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$ according to the transition density $p(y_t|x_{t-1}, \theta)$? The basic idea is twofold: one notes that, as defined, the map minimizes the Kullback-Liebler divergence between the models $p(y_t|x_{t-1}, \theta)$ and $f(y_t|x_{t-1}, \eta)$ and that one can use simulation to integrate with respect to the scientific model $p(y_t|x_{t-1}, \theta)$. Intuitively this means that we can choose N , the simulation size, so large that the simulated data gives us all the information we need about the nature of the process at the given θ and then find the corresponding η by maximizing the likelihood of the simulated data under the statistical model $f(\cdot|\cdot, \eta)$. We find the η which gives us the same kind of data as θ . More formally, we are finding the η that puts the Kullback-Liebler divergence

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

to zero by solving a minimization problem and are noting that the term

$$\iint \log p(y|x, \theta) p(y|x, \theta) dy p(x|\theta) dx$$

does not have to be computed to solve the problem. We approximate the integral that does have to be computed in the usual way:

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

(Or by $\frac{1}{R} \sum_{r=1}^R \frac{1}{n} \sum_{t=1}^n \log f(\hat{y}_{t,r}|\hat{x}_{t-1,r}, \eta)$ if not ergodic. We assume ergodicity hereafter; if not, the requisite modifications are obvious.) Thus, upon dropping the division by N , the map is computed as

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

Our algorithm will incorporate a simple approach for computing this mle.

Computing the map is an equation solving problem and there are other approaches. One might, for example, choose some test functions such as $I(y_{i-1} < y \leq y_i, x_{j-1} < x \leq x_j)$ $i = 1, \dots, I, j = 1, \dots, J$, and solve the set of nonlinear equations

$$\frac{1}{N} \sum_{t=1}^N I(y_{i-1} < y \leq y_i, x_{j-1} < x \leq x_j) = \int_{x_{j-1}}^{x_j} \int_{y_{i-1}}^{y_i} f(y|x, \eta) dy f(x|\eta) dx$$

for η . The approach that we adopt has these advantages: It provides the scaling for the prior used for model assessment in a simple way. The code can be reused in Section 4 to

compute the posterior for η statistical model subject to this prior. The method requires only one integration.

Under the assumption that (1) holds, how one solves for η for given θ does not matter as long as it is done with reasonable accuracy. If one is unsure of the validity of the assumption in an application, it can be checked for specific values of θ by (1) fitting the statistical model to a scientific model simulation, (2) simulating from the fitted statistical model, and (3) checking to see if the empirical distributions of the two simulations match.

3.2 A Metropolis Algorithm for θ

We will use the Metropolis algorithm to compute the posterior distribution of θ . This algorithm will have to accommodate various forms of prior information and a computational scheme for obtaining the mle defining the map g .

The Metropolis algorithm is an iterative scheme generating a sequence of θ values according to a Markov chain whose stationary distribution is the posterior. Since for every θ , we shall need the corresponding $\eta = g(\theta)$, we will generate a sequence of (θ, η) pairs. (Thus providing the posterior both in θ -space and in η -space.) As in all Bayesian analyses, we must specify our prior and likelihood. For our Metropolis chain we must also specify a Markov chain in θ used to propose new values.

Let \mathcal{L} denote the likelihood of our observed data. Our basic assumption is that to compute this likelihood we may use f evaluated at the appropriate η :

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

where (y_t, x_{t-1}) denotes the observed data and n the sample size. Let π denote the prior distribution. As discussed in Section 2 above, this prior can depend on the parameters $\theta \in \Theta$ of the scientific model $p(y|x, \theta)$, the value ψ taken on by the functionals $\Psi : p(\cdot|\cdot, \theta) \mapsto \psi$, the parameters $\eta \in H$ of the statistical model $f(y|x, \eta)$, and the value v taken on by the functionals $\Upsilon : f(\cdot|\cdot, \eta) \mapsto v$. Accordingly, this prior has argument (θ, ψ, η, v) . Let q denote our Metropolis proposal. For a given θ , $q(\theta, \theta^*)$ defines a distribution of potential new values θ^* .

Given a current θ^o and the corresponding $\eta^o = g(\theta^o)$ we obtain the next pair of (θ', η')

values 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 the transition density $p(y_t|x_{t-1}, \theta^*)$.
3. Compute $\eta^* = g(\theta^*)$ and ψ^* from the simulation $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$ and v^* from η^* .
4. Let $\alpha = \min\left(1, \frac{\mathcal{L}[g(\theta^*)]\pi(\theta^*, \psi^*, \eta^*, v^*)q(\theta^*, \theta^o)}{\mathcal{L}[g(\theta^o)]\pi(\theta^o, \psi^o, \eta^o, v^o)q(\theta^o, \theta^*)}\right)$.
5. With probability α , $(\theta', \eta') = (\theta^*, \eta^*)$, otherwise we repeat, i.e. $(\theta', \eta') = (\theta^o, \eta^o)$.

Steps 1, 4, and 5 are just the standard Metropolis algorithm. Steps 2 and 3 are essential features of our approach.

In order to complete the specification of our algorithm we need to choose a q . We shall also propose a particular approach to the computation of η in Step 3 that accommodates all sources of prior information. These are the next topics.

3.3 Choice of θ Proposal

To specify our algorithm we must choose a proposal transition density q for θ . To compute the likelihood at a proposed θ , the scientific model must be simulated. For a sophisticated scientific model, this simulation may involve significant computation. Moreover, there could well be a call to a nonlinear optimizer or nonlinear equation solver that needs starting values involved in this simulation. This motivates us to consider proposing small changes in θ so that computational results from the old θ may be used in doing the computations for the proposed θ . In particular, if θ is not changed too much, results from the previous computation can be used as starting values for the new one. The cost of this strategy is in dependence in the Markov Chain. If we limit ourselves to small changes, it may take us a while to navigate from one place in the parameter space to another.

We start by discretizing θ because, as seen later, discretization permits significant improvements in computational efficiency. For the i^{th} component of θ we choose $a_i < b_i$, and s_i . We then let θ_i take on the values $a_i + js_i$ where j ranges from 1 to g_i which is equal to the integer part of $(b_i - a_i)/s_i$. Thus, θ_i takes values between a_i and b_i with step sizes s_i .

To propose a new θ we first randomly choose a component to change, with each component having the same chance of being chosen. If the i^{th} component is chosen, there is some j such that the current $\theta_i = a_i + js_i$. We choose a set of distributions $q_i(j, k)$ on $\{1, 2, \dots, g_i\}$ where i is the θ component, j is the current grid position of that component, and k denotes the random new grid position to be drawn. We draw $k \sim q_i(j, \cdot)$ and let θ^* be obtained from θ by changing the i^{th} component from $a_i + js_i$ to $a_i + ks_i$.

To specify the $q_i(j, k)$ we choose a σ_i for the i^{th} component of θ and let

$$q_i(j, k) \propto \begin{cases} \exp(-\frac{1}{2\sigma_i^2}(k - j)^2) & k \neq j \\ 0 & \text{else} \end{cases}$$

The choice of σ_i determines the number of s_i that we tend to move. We assign 0 probability to proposing that we stay put since there is no point in proposing that we go to where we are.

To run the θ Metropolis chain, we have to choose a starting value for each θ_i . The choice of a_i and b_i is not critical; a_i and b_i can be set so that the intervals (a_i, b_i) cover the support of the posterior by a wide margin without noticeably degrading performance. The choice of s_i is crucial. We will move away from the starting value in increments of size s_i . The combination of the choice of s_i and σ_i determines the size of the changes that q proposes. The choice of s_i determines the accuracy of our inference. When we choose s_i we are saying that, as a practical matter, we only need to know θ_i in terms of s_i units. Two θ 's that differ in component i by less than s_i are virtually the same as a practical matter. Since computation is expensive, we should not waste resources by determining θ on a finer scale than we actually care about.

3.4 Computing the MLE of η with the Simulated Data

Step 3 of the Metropolis algorithm presented in Section 3.2 is the computation of the mle of η under the statistical model given the large simulated data set.

Since the $f(\cdot|\cdot, \eta)$ family is generally chosen to be flexible and high dimensional, this likelihood can be complicated. However, the simulated data set is large and we have a good starting value. In the notation of Section 3.2, the current η^o should be a good starting value in the search for η^* . This assumes that θ^* is not too different from θ as discussed in

Section 3.3. In order to keep our analytical requirements to a minimum, we would like our method to require the computation of the objective

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

and nothing else.

Given these considerations, to find the mle we run a Markov chain for η using the simulated data. With the large sample size, the Markov chain will quickly move from the η to values close to η^* . We use a normal random walk Metropolis within Gibbs approach. That is, we first subdivide the η vector into subvectors. In the manner of a Gibbs sampler, we cycle through the subvectors one at a time. For subvector η_i , we use the normal proposal

$$q(\eta_i, \eta_i^*) \sim N(\eta_i, \Sigma_i)$$

in a standard random walk Metropolis algorithm. Effectively, this is a simulated annealing optimization algorithm where N is the temperature parameter because N is what controls the peakedness of the likelihood.

The advantage of a simulated annealing strategy over a derivative based hill climbing method is that analytic derivatives are not required and one has better control over the computational burden. That is, one can easily control the length of an MCMC chain but it is hard to achieve any control over the number of iterations of the internal line search algorithm of a nonlinear optimizer and one can only bound the number of putative full steps. As nearly the entire computational cost of the proposed methodology is concentrated here in the computation of η , control is essential. A side benefit is that the chain for η also provides the scaling for the model assessment strategy proposed in Section 4.

The computation of η requires start values, as does any nonlinear optimization. If θ is only moved slightly between iterates of in an MCMC chain to compute the posterior for θ , then the last computed value of η will be a good start for the next. This consideration becomes doubly important when the scientific model $p(y|x, \theta)$ contains an embedded nonlinear computation requiring start values as does the habit model (Section 5.2). These requirements argue for a random walk proposal density that makes small moves in connection with the MCMC chain used to get the posterior distribution of θ .

We choose a fixed number of steps to run this chain, and keep the visited η which has the highest likelihood under the simulated data. In our experience, it is relatively straightforward to choose (i) a simulation sample size which is large enough to ensure that the map is adequately recovered by the mle and (ii) a number of steps to iterate the Markov chain in η that will ensure the chain has finished moving away from the starting value of η . We might also note that putting θ on a grid is a considerable help here because it reduces the accuracy to which η^* needs to be computed.

This is a computationally costly part of our overall procedure. Since the simulation sample size N is large, each computation of the likelihood for the η chain can take a long time. Nonetheless, we have found that, because of the large N , this part of the procedure is remarkably stable, even though the statistical model may actually be difficult to estimate on data samples of the size n that we actually observe.

The main reason for placing θ on a grid is that a significant reduction in computational time can be achieved. With θ on a grid, it takes only a modest amount of memory to store all previously computed values of $(\eta, \mathcal{L}(\eta), \pi(\theta, \psi, \eta, \nu))$ in a binary tree indexed by θ . When θ is revisited, both the fact that it is a revisit and the information required for Step 4 of the Metropolis algorithm for θ (Subsection 3.2) can be quickly obtained by traversing the tree. The two costliest Steps 2 and 3 are thereby eliminated. We have found the C++ associative map to be an exceptionally convenient implementation of such a tree. By storing previous results in a tree and looking them up, the θ chain runs faster as it becomes longer.

We might also remark that the object oriented features of the C++ language are exceedingly helpful. The selective ignorance that object oriented programming permits and enforcement of interfaces that object oriented programming provides allow a clean division of tasks among project participants thus permitting each to focus on the particular task at hand without having to be concerned with details elsewhere in the code. The polymorphic features of the language are also helpful in allowing the scientific and statistical models to be abstractions defined by an interface.

3.5 An Illustration of the MLE Computation

Figure 3 displays the results of ten runs of an η chain (from Section 5). Every run is clearly visible in the figure as a segment of 200 iterations. In the notation of Section 3.2, each segment displays the results of a Markov chain in η , started at η^o , using the data simulated from the scientific model at θ^* and the likelihood of the statistical model coupled with a flat prior on η . On the vertical axis, the log-likelihood is plotted. Within the second segment, we can see the likelihood quickly increase as the η value moves toward the mle. The segments level off at different likelihoods because they represent the likelihoods of different simulated data sets. Because of the large size of each simulated data set, $N = 50,000$ in this instance, the posterior is very tight around the mle and the chain quickly moves to a new level. In this case we could probably have used 100 iterations rather than 200 since (unlike most MCMC applications) we don't need to run it any longer after we have found the neighborhood of the mode.

Figure 3 about here

3.6 Prior Information

Now we consider the possibility that there may be important prior information about features of the scientific model which are not readily expressed in terms of θ . This information is available within the scientific model at the time at which the simulated data $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$ is computed and will often involve variables that are not elements of y_t . Typically, elements of ψ are statistics that are computable from an augmented simulation that includes both the elements of y_t and these additional latent variables. Formally ψ can be regarded as the value taken on by a vector of functionals of the scientific model, as noted previously. Also, as discussed previously, the prior can depend on η and v of the statistical model. Usually the joint prior will factor as

$$\pi(\theta, \psi, \eta, v) \propto \pi(\theta, \psi)\pi(\eta, v).$$

We will presume that this is so to allow us to keep the focus on $\pi(\theta, \psi)$ in the remaining discussion; the modifications are obvious and minimal if not.

For computational reasons, we partition the prior $\pi(\theta, \psi)$ into support restrictions that can be computed from knowledge of θ alone and those that require, in addition, knowledge of ψ . These are $\pi_1(\theta)$ and $\pi_1(\theta, \psi)$, respectively, and are zero-one indicator functions where one indicates that the support condition is satisfied. The prior can now be factored as

$$\pi(\theta, \psi) \propto \pi_1(\theta)\pi_1(\theta, \psi)\pi_2(\theta, \psi)$$

The idea is that $\pi_1(\theta)$ is cheap to compute and when zero the simulation $\{\hat{y}_t, \hat{x}_{t-1}\}_{t=1}^N$ need not, or perhaps even cannot, be run. If $\pi_1(\theta) = 0$, then the α of Step 4 in Section 3.2 is trivial to calculate and the proposed θ can be immediately rejected. In many applications $\pi_1(\theta, \psi)$ will indicate simulation success or failure. (This is usually because some analytically intractable support condition on θ is violated so that attempting to simulate is the only practicable method for checking it.) In the case of failure, computation cannot proceed and the proposed θ is rejected at Step 4. If both $\pi_1(\theta)$ and $\pi_1(\theta, \psi)$ are positive, then computations can proceed. The Metropolis algorithm is a wonderfully simple way to compute a posterior in the presence of priors with support restrictions and, therefore, well suited to problems of the sort under consideration here.

4 Inference Off the Manifold: Model Assessment

We now consider relaxation of the prior imposed on the statistical model by the scientific model. As η is no longer on the manifold \mathcal{M} (c.f. Figure 1) but merely attracted to it, we must now describe the computations in terms of the posterior for η .

Recall that the scientific model $p(y|x, \theta)$ together with the prior $\pi(\theta, \psi)$ are viewed as a prior on the statistical model $f(y|x, \eta)$, that may, itself, have prior $\pi(\eta, \nu)$. The information $\pi(\theta, \psi)$ consists of two pieces: (1) η is restricted to lie on the manifold

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

and (2) $\pi(\theta, \psi)$ weights some points on \mathcal{M} more heavily than others.

What one would like to do is see how results change as this prior is relaxed. However, once we have moved off the manifold \mathcal{M} we can no longer view results from the perspective of the scientific model $p(y|x, \theta)$ and must view them from the perspective of the statistical

model $f(y|x, \eta)$ because the scientific model loses meaning off the manifold. For instance, off the manifold ψ can no longer be computed. Therefore, seeing how results change must be taken to mean seeing how some functional of the statistical model changes. For convenience we presume that this functional is an element of Υ defined in Section 2. This functional might map to an element of η , if such has scientific meaning, or a statistic computable from a simulation such as the return over a hundred year planning horizon as in Figure 7.

Relaxation of the prior will be formulated in terms of a weighted distance of η from the manifold \mathcal{M} . Recall that θ has been restricted to a grid

$$\Theta_G = \{\theta_j \in \Theta : j = 1, \dots, G\}$$

Therefore we can cheaply compute the distance from η to the manifold as

$$d(\eta, \mathcal{M}) = \min_{j=1, \dots, G} [\eta - g(\theta_j)]' \Sigma_\eta^{-1} [\eta - g(\theta_j)],$$

where Σ_η is a scaling matrix. Later we shall also require the index \hat{j} at which the minimum occurs.

We propose that Σ_η be computed as follows: Initialize to zero. Whenever the Metropolis-Hastings chain for computing the mle of η described in Subsection 3.4 must be run, update

$$\Sigma_\eta \leftarrow \Sigma_\eta + (\eta_1 - \eta_2)(\eta_1 - \eta_2)'$$

where η_1 is a point on the chain immediately after transients have died out and η_2 is the last point on the chain. This method of scaling the distance measure is reasonable because it puts η on the scale of the posterior: Distance is being measured in units of standard deviation.

At each point in Θ_G the value of $\pi(\theta, \psi)$ is known, so $\pi(\theta_j, \psi_j)$ is easily evaluated at $j = \hat{j}$ and we can define our prior on H as

$$\pi_\kappa(\eta, \nu) \propto \pi(\theta_{\hat{j}}, \psi_{\hat{j}}) \exp\left(-\frac{d(\eta, \mathcal{M})}{2\kappa}\right) \pi(\eta, \nu). \quad (2)$$

The first multiplicative term of $\pi_\kappa(\eta, \nu)$ represents prior information along the manifold while the second term represents our prior about closeness to the scientific model. The prior becomes more diffuse and the scientific model less influential as the scale factor κ increases.

One should note that using posterior draws to compute the image \mathcal{M} of the map $g(\cdot)$ does make use of the data to determine $\pi_\kappa(\eta, \nu)$. We do not regard this as a problem because

the η draws will contain enough extreme values to make sure that the extent of \mathcal{M} is large enough. If one is particularly worried about this, one could run the theta chain with a smaller amount of data to make sure that \mathcal{M} is over explored.

The computation of the posterior distribution of η using the statistical model $f(y|x, \eta)$ and prior $\pi_\kappa(\eta, v)$ can be accomplished by a routine application of the Metropolis algorithm because $\pi_\kappa(\eta, v)$ is easily computable and an analytic expression for $f(y|x, \eta)$ is available.

Our proposal is that the scientific model be assessed by plotting a suitable measure of the location and scale of the posterior distribution of v against κ or, better, sequential density plots as in Figure 2. What one expects to see, for a well fitting scientific model, is that the location measure does not move by a scientifically meaningful amount as κ increases, which indicates that the model fits, and that the scale measure increases, which indicates that the scientific model has empirical content. We discuss the habit persistence asset pricing model next and apply our proposed methods. What we shall see (Figure 7) is that as κ increases the scale of two functionals of interest increases, indicating empirical content, but that location also shifts, providing evidence against the model.

5 Habit Persistence Asset Pricing Model

In this section we shall apply the proposed methods to the habit persistence asset pricing model of Campbell and Cochrane (1999). Although it is widely viewed as a behavioral model and it is the result of an admitted attempt to reverse engineer away the statistical inadequacies of the classical consumption based asset pricing model (Lucas, 1978), the habit model actually can be justified from plausible micro-foundations (Guisen, 2003).

The habit model exhibits all the characteristics discussed in Section 1: (1) the likelihood is not available; (2) prior information on model parameters is available; (3) prior information in the form of restrictions on model functionals is available; (4) the model can be simulated; (5) a generally accepted statistical model for its data is available.

In the remainder of this section, we describe the data, introduce the habit model, and apply the methods that we have proposed. Our main substantive conclusion is that the habit model is not supported by the data, as seen in Figure 7. We also find, as seen in Table 1, that the reason some studies conclude otherwise is that they use estimation methods that rely on

a counterfactual assumption that the conditional variance of the data is homogeneous. From the implied map we discover in Figure 6 that a sharply delineated subset of the habit model's parameters control the observable conditional heterogeneity and they are the parameters that describe the individual's preferences, particularly the parameter that most directly affects risk aversion.

5.1 The Data

The data are annual, U.S., per capita, non-durables and services consumption C_t^a and the price P_{dt}^a of a value weighted portfolio comprised of all stocks listed on the New York and American stock exchanges from 1929 through 2001 collected at the last trading day of the year, adjusted for inflation, and converted to a per capita basis. Data sources and collection protocol are described in Bansal, Gallant, and Tauchen (2004). These are converted to an annual consumption growth series and stock returns series using $\Delta c_t^a = \log(C_t^a) - \log(C_{t-1}^a)$ for the former and $r_{dt}^a = \log(P_{dt}^a) - \log(P_{d,t-1}^a)$ for the later. We use annual data rather than monthly or quarterly data because seasonality issues are thereby eliminated and because annual consumption data are generally regarded as being of higher quality and go back farther in time than higher frequency consumption data. In our previous vector notation, we have

$$y_t = \begin{pmatrix} \Delta c_t^a \\ r_{dt}^a \end{pmatrix} \quad t = 1, \dots, n = 72$$

We now consider how to simulate similar data from the habit model. As in Campbell and Cochrane (1999), we will run the model at a monthly frequency and then aggregate to get annual data. This is a relatively common procedure in this literature because it is more reasonable to assume that budgeting decisions are made monthly rather than annually whereas the data to be matched are annual for the reasons given above.

5.2 Model Description

The intuitive notions behind any consumption based asset pricing model are that agents receive wage income and dividend income from which they purchase consumption. Agents seek to smooth consumption over time by trading shares of stock that pay a random dividend

and bonds that pay interest with certainty. The driving processes of such a model are the wage process and dividends or, equivalently, consumption and dividends because 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 bond receipts. Agents are endowed with a utility function that depends on the entire consumption process. The parameters of this function determine their preferences. Standard specifications and parameter settings generate a preference for smooth rather than rough consumption streams. The first order conditions of their utility maximization problem determine the prices at which they are willing to trade securities. We shall describe the driving processes, the utility function, and the first order conditions of the habit model, in that order.

The driving processes of the habit model are real consumption growth and real dividend growth

$$\begin{aligned}c_t - c_{t-1} &= g + v_t \\d_t - d_{t-1} &= g + w_t\end{aligned}$$

The convention throughout this section is that lower case denotes logarithms of upper case quantities; e.g. $c_t = \log(C_t)$, $d_t = \log(D_t)$. As mentioned above, the time increment is one month.

The errors (v_t, w_t) of the driving processes are normal with mean zero and variance

$$\text{Var} \begin{pmatrix} v_t \\ w_t \end{pmatrix} = RR',$$

where R is upper triangular with nonzero elements r_{11} , r_{12} , and r_{22} . At times it is more convenient to express the variance matrix in terms of $\sigma^2 = \text{Var}(v_t)$, $\sigma_w^2 = \text{Var}(w_t)$, and $\rho = \text{corr}(v_t, w_t)$.

Upon exponentiation to get C_t and summing over adjacent blocks of twelve, the consumption process does correspond conceptually to the data series C_t^a described above. On the other hand, the dividend process does not correspond conceptually to observable data primarily because what can be observed is strongly influenced by tax policy causing, e.g., corporations to shift dividend payments into or out of stock repurchases. Therefore, d_t is

to be regarded as a latent variable. It remains to consider how the returns process r_{dt}^a is simulated.

The habit model asserts that all agents in the economy are endowed with the same utility function

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

where X_t is habit and δ the time discount factor. Risk aversion depends on X and is much higher than γ for plausible X . Habit is determined by the surplus ratio $S_t = (C_t - X_t)/C_t$. S_t is the state variable of the model and evolves as $s_t - \bar{s} = \phi(s_{t-1} - \bar{s}) + \lambda(s_{t-1})v_{t-1}$, where $\lambda(s) = \frac{1}{\bar{s}}\sqrt{1 - 2(s - \bar{s})} - 1$ if $s_t \leq s_{\max}$ and zero else. Recalling the upper and lower case convention, \bar{s} and s_{\max} are computed from model parameters $\theta = (g, r_{11}, r_{12}, r_{22}, \phi, \delta, \gamma)$ as $\bar{S} = \sqrt{(r_{11}^2 + r_{12}^2)\gamma/\phi}$ and $s_{\max} = \bar{s}(1 - \bar{S}^2)/2$,

Rather than compute the stock price directly one usually works in terms of the price dividend ratio P_{dt}/D_t because the price dividend ratio is stationary whereas prices are not. Agents are presumed to be so numerous that each can solve their own utility maximization problem without regard to the actions of the others. Under this assumption, the price dividend ratio satisfies $P_{dt}/D_t = V_\theta(S_t)$ where $V_\theta(\cdot)$ solves the integral equation

$$0 = \mathcal{E}_t \left\{ V_\theta(S_t) - \delta \left(\frac{S_{t+1}C_{t+1}}{S_tC_t} \right)^{-\gamma} \left(\frac{D_{t+1}}{D_t} \right) [1 + V_\theta(S_{t+1})] \right\} \quad (3)$$

and \mathcal{E}_t denotes conditional expectation given S_t . These are the first order conditions of each agent's optimization problem.

Using instrumental variables the conditional integral equation (3) can be converted to a set of unconditional integral equations. These unconditional integrals can be computed by averaging over a long simulation of (C_t, D_t, S_t) . The function $V_\theta(S_t)$ has a series representation that, when substituted into these integral equations, yields a system of nonlinear equations whose unknowns are the coefficients of the series expansion. The details are in Bansal, Gallant, and Tauchen (2004). Only one aspect of this computation is important here: The coefficients of $V_\theta(S_t)$ are determined by a nonlinear equation solver (Newton's method) that needs starting values. Further, if two adjacent points θ_i and θ_{i+1} in the MCMC chain for the posterior of θ are close together then the coefficients of $V_{\theta_i}(S_t)$ make good starting values for computing those of $V_{\theta_{i+1}}(S_t)$.

What one does, then, for θ given, is simulate consumption, dividends, and compute the surplus ratio to get the realization $\{\hat{C}_t, \hat{D}_t, \hat{S}_t\}_{t=1}^{12N}$. (The upper limit is $12N$ to allow aggregation of adjacent blocks of twelve to generate a simulations of size N at the annual frequency.) Next determine $V_\theta(\cdot)$ by solving a system of nonlinear equations that require $\{\hat{C}_t, \hat{D}_t, \hat{S}_t\}_{t=1}^{12N}$ for their computation. Then compute stock price as $\hat{P}_{dt} = \hat{D}_t V_\theta(\hat{S}_t)$. Geometric stock returns at the monthly frequency are obtained from $\hat{r}_{dt} = \log(\hat{P}_{dt}) - \log(\hat{P}_{d,t-1})$. Similar computations provide the risk free interest rate \hat{r}_{ft} (i.e. the geometric return on a one month bond), which is used to express a portion of our prior information in Subsection 5.3. Therefore, given model parameters $\theta = (g, r_{11}, r_{12}, r_{22}, \phi, \delta, \gamma)$, we can simulate, in particular, consumption, dividends, stock returns, bond returns and from these compute functionals of interest.

To convert from the monthly frequency to the annual frequency, we aggregate $\{\hat{C}_t, \hat{r}_{dt}\}_{t=1}^{12N}$ by summing consecutive blocks of twelve to get the annual series $\{\hat{C}_t^a, \hat{r}_{dt}^a\}_{t=1}^N$. The simulated series of observables at the annual frequency is $\{\hat{y}_t\}_{t=1}^N$ with $\hat{y}_t = (\log \hat{C}_t^a - \log \hat{C}_{t-1}^a, \hat{r}_{dt}^a)$.

To summarize, the scientific model has seven parameters. Four of the parameters, namely g, r_{11}, r_{12} , and r_{22} , control the driving process (C_t, D_t) . Three, namely ϕ, δ , and γ describe preferences with γ being termed the risk aversion parameter. The simulation \hat{C}_t^a of the observable process C_t^a can be obtained directly by aggregating a simulation of the driving process. To simulate the observable returns process r_{dt}^a , we have to generate a long simulation of the driving process which is used to compute the integrals in a system of equations that are derived from the model's first order conditions. The resulting nonlinear system is solved by Newton's method. Once solved, monthly returns can be computed and then aggregated to get the simulation \hat{r}_{dt}^a .

5.3 Prior Information

Prior information arises from multiple sources. Some consists of support conditions such as non-explosive restrictions on autoregressive parameters. As above, let the support conditions be represented by $\pi_1(\theta)$, where $\pi_1(\theta) = 1$ indicates that θ is in the support and 0 that it is not; $\pi_1(\theta)$ is next to costless to compute. A solution $V_\theta(\cdot)$ to (3) does not exist for all parameter values θ that satisfy support conditions. It is effectively impossible to determine

analytically the additional conditions that must be imposed on θ to guarantee existence. Therefore the only practical way to determine if a solution exists for given θ is to simulate $\{\hat{C}_t, \hat{D}_t, \hat{S}_t\}$ and attempt to solve (3). Let the support conditions that indicate for which θ equations (3) can be solved be represented by $\pi_1(\theta, \psi)$, where $\pi_1(\theta, \psi) = 1$ indicates that θ is in the support and 0 that it is not. $\pi_1(\theta, \psi) = 0$ is extremely costly to compute because the nonlinear equation solver runs longest when it fails.

Of more substance is that the ex-ante risk free rate r_f^a , which is the interest rate for the bonds of the model, is not directly observable because its computation from observed interest rates requires estimation of the representative agent's anticipated inflation (Mishkin, 1981). Because the evidence indicates that the risk free rate is low with a variance that is much smaller than the variance of inflation (Campbell, 2002), an attempt to generate a risk free rate series over our sample period would produce a series that would consist almost entirely of measurement error. It seems better to accept the values for the mean ($\mathcal{E}(r_f^a) = 0.89$) and standard deviation ($\text{Var}(r_f^a) \doteq 1$) that Campbell determined from several long historical time series and adopt a normal prior that implies $P(|\mathcal{E}(r_f^a) - 0.89\%| < 1\%) = 0.95$. In this computation $\mathcal{E}(r_f^a)$ is computed by aggregating the monthly (geometric) risk free rate r_{ft} in the habit model simulation to the annual frequency r_{ft}^a and then averaging to get $\mathcal{E}(r_f^a)$. This is complicated prior information that affects every element of θ ; it is essential to model identification and would be very hard to impose in any other way. In addition we impose normal priors that imply $P(|\rho - 0.2| < 0.1) = 0.95$ and $P(|\phi - 0.9884| < 0.01) = 0.95$. (Because θ is restricted to a grid, our priors are actually proportional to a histogram of a normal. We take this as understood and just call them normal.) Denote the product of these three normal densities by $\pi_2(\theta, \psi)$; $\pi_2(\theta, \psi)$ is moderately costly to compute.

The restrictions $P(|\rho - 0.2| < 0.1) = 0.95$ and $P(|\phi - 0.9884| < 0.01) = 0.95$ help with identification. A consequence of treating dividends as latent is that ρ , which is the correlation between monthly consumption growth and dividend growth, is so poorly determined by the observables, which are annual consumption growth and stock returns, that prior input is necessary. We impose $\rho \doteq 0.2$ following Campbell and Cochrane (1999). The restriction $\phi \doteq 0.9884$ also follows Campbell and Cochrane. The parameter ϕ is an autoregressive parameter and we find that it is necessary to prevent the MCMC chain from putting ϕ too

close to one. Experimentation indicates that the scaling $P(|\rho - 0.2| < 0.1) = 0.95$ and $P(|\phi - 0.9884| < 0.01) = 0.95$ is the mildest that permits good performance of the MCMC chain for θ .

5.4 The Statistical Model

The statistical model is a bivariate GARCH system with normal innovations:

$$\begin{pmatrix} c_t^a - c_{t-1}^a \\ r_{dt}^a \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} c_{t-1}^a - c_{t-2}^a - b_1 \\ r_{d,t-1}^a - b_2 \end{pmatrix} + \begin{pmatrix} r_{11,t-1} & r_{12,t-1} \\ 0 & r_{22,t-1} \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix}$$

$$\begin{pmatrix} r_{11,t-1} \\ r_{12,t-1} \\ r_{22,t-1} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} + \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \\ \rho_{31} & \rho_{32} \end{pmatrix} \begin{pmatrix} |z_{1,t-2}| \\ |z_{2,t-2}| \end{pmatrix} + \tau \begin{pmatrix} r_{11,t-2} - \rho_1 \\ r_{12,t-2} - \rho_2 \\ r_{22,t-2} - \rho_3 \end{pmatrix} \quad (4)$$

where (z_{1t}, z_{2t}) are independent standard normal random variables. The statistical model $f(y_t|x_{t-1}, \eta)$ has sixteen parameters

$$\eta = (b_1, b_2, b_{11}, b_{21}, b_{12}, b_{22}, \rho_1, \rho_2, \rho_3, \rho_{11}, \rho_{21}, \rho_{31}, \rho_{12}, \rho_{22}, \rho_{32}, \tau)$$

This model is the leading term of the SNP expansion (Gallant and Tauchen, 1996) that is widely used in applications involving multivariate time series data from financial markets and is known to fit such data well (Ahn, Dittmar, and Gallant, 2002; Chernov, Gallant, Ghysels, and Tauchen, 2003)

The analysis that follows indicates that our statistical model is rich enough to extract the information in the data that relates to the statistical adequacy of the habit model. We remark that we are not limited here by the number of observations as we would be when employing frequentist methods such as EMM. See, e.g., the EMM application in Bansal, Gallant, and Tauchen (2004) where the richest model that can actually be fit to a four dimensional series that includes our two variables is a diagonal VAR with homogenous conditional variance. (They must use four variables to get identification and are thus compelled to use dividend data despite its dubious quality.) The relevant sample size for us is the simulation size N , which is entirely under our control.

Without restricting τ in (4), even our two dimensional system is troublesome to fit by Bayesian MCMC methods when the number of observations drops to the sample size n . When iterating η indirectly via iteration of the θ MCMC chain as described in Section 3, the habit model and its prior constrain η to lie in a plausible region of the manifold \mathcal{M} thereby restricting τ to plausible values, i.e. to values with $|\tau| < 1$. However, when we iterate the η chain directly to explore model adequacy as described in Section 4 and allow movements far from the manifold by setting κ to a large value, then τ can move to explosive values and stick. Therefore, when off the manifold \mathcal{M} , we impose the normal prior $P(|\tau - 0.6| < 0.1568) = 0.95$, which is the term $\pi(\eta, v)$ in equation (2).

5.5 Model Estimation: Scientific Prior Imposed

We ran the θ chain for 800,000 iterations discarding every 8 leaving 100,000. The chain was started near the mode of the posterior density to avoid problems with transients. (The mode was determined from several initial runs of size 100,000.) Visual inspection of time series plots of the draws (not shown) indicated that this strategy for eliminating transients was successful.

Table 1 about here

Descriptors of the Bayesian posterior distribution of $\theta = (g, r_{11}, r_{21}, r_{22}, \phi, \delta, \gamma)$ are shown in the top half of Table 1. Also shown are statistics describing the frequentist (EMM) sampling distribution for $(g, \phi, \delta, \gamma)$ and some additional parameters that determine the distribution of the errors of the driving process under an assumption that dividend growth and consumption growth are cointegrated. (They are defined in the table legend.) The scaling for the descriptors in the top half of the table corresponds to the model as defined in Subsection 5.2. For convenience in interpretation, descriptors of the posterior distribution of the standard deviation of consumption growth σ , the standard deviation of dividend growth σ_w , and the correlation between them ρ are computed from those for (r_{11}, r_{21}, r_{22}) and, together with the descriptors for $(g, \phi, \delta, \gamma)$, are annualized and displayed in the bottom half of the table. Similarly for the statistics that describe the frequentist sampling distribution. Further down are descriptors for $(\mathcal{E}(r_f^a), \sqrt{\text{Var}(r_f^a)}, \mathcal{E}(r_d^a), \sqrt{\text{Var}(r_d^a)})$.

In the columns of Table 1 that are labeled mode, the values shown are the mode of the multivariate posterior for $\theta = (g, r_{11}, r_{21}, r_{22}, \phi, \delta, \gamma)$, not the mode of each marginal. This is because the values of the multivariate mode do appear in the MCMC chain for θ and therefore do correspond to an actual simulation whereas the vector of marginal modes might not and might not even satisfy $\pi_1(\theta, \psi) = 1$. For some purposes, e.g. computing $\mathcal{E}(r_f^a)$, this distinction is important. Similarly the posterior mean might not be in the chain. (As with any mean, the posterior means are both joint and marginal.) The values for $(\mathcal{E}(r_f^a), \sqrt{\text{Var}(r_f^a)}, \mathcal{E}(r_d^a), \sqrt{\text{Var}(r_d^a)})$ in the columns labeled mode are computed from the values for $(g, r_{11}, r_{21}, r_{22}, \phi, \delta, \gamma)$ that are shown above them in the upper half of the table; they are not marginal modes. On the other hand, those in the columns labeled mean are the marginal means of the posterior distribution of $(\mathcal{E}(r_f^a), \sqrt{\text{Var}(r_f^a)}, \mathcal{E}(r_d^a), \sqrt{\text{Var}(r_d^a)})$. All standard deviations in the columns headed Bayes are computed from marginal posteriors. The frequentist (EMM) values optimize a criterion function and therefore do correspond to a simulation; the values shown for $(\mathcal{E}(r_f^a), \sqrt{\text{Var}(r_f^a)}, \mathcal{E}(r_d^a), \sqrt{\text{Var}(r_d^a)})$ are computed from that simulation.

As with any discrete time dynamic model, annualized parameter values for the habit model do not have the property that simulations obtained by running the model at the annual frequency using the annualized parameter values will have the same distribution as simulations obtained by aggregating a monthly simulation. For this reason, the values at the monthly frequency in the upper half of of Table 1 should be regarded as the definitive estimates of model parameters. The annualized values in the lower half of the table are to be regarded as only an aid in their interpretation.

Figure 4 about here

Figure 5 about here

The Bayes-GARCH columns of Table 1 report computations that use the GARCH statistical model. Their most striking feature is the large deviation of the Bayes-GARCH estimate of mean stock returns $\mathcal{E}(r_d^a) \doteq 11\%$ from the value computed from the data of 6.02% and from the EMM estimate of 6.54%. The reason for this is that the EMM estimator as implemented

in Bansal, Gallant, and Tauchen (2004) completely ignores the conditional heterogeneity in the data. The auxiliary model used by Bansal, Gallant, and Tauchen, whose impact on the EMM estimation procedure is analogous to the impact of the statistical model on the Bayes procedure proposed in this article, is a VAR with homogeneous conditional variance. If we too use a VAR with homogeneous conditional variance to implement our Bayes estimator, the results in the column labeled Bayes-VAR are obtained, bringing the Bayes estimate of $\mathcal{E}(r_d^a)$ into agreement with the data and with the EMM estimates. The Bayes-VAR estimates are also in close agreement with the values reported in Campbell and Cochrane (1999) that were determined by informally matching to a set moments that do not identify model features influenced by conditional heterogeneity.

The reasons that Bansal, Gallant, and Tauchen had to suppress GARCH in their auxiliary model are twofold. Observable data rather than prior information are used to achieve identification despite concerns for its quality thereby creating the need for a four dimensional series. The data ($n = 72$) are too sparse to support a GARCH specification in four dimensions as discussed earlier in Subsection 5.4. We, on the other hand, have no such difficulties because we are fitting to large simulations, not to data.

Conceptually, substituting a VAR statistical model for the GARCH statistical model in our Bayesian analysis amounts to the imposition of a sharp prior that zeros out all but the constant term of the GARCH variance function (4). It is of interest to view the shift that this sharp prior imposes on marginal posteriors. These are shown as Figure 4 for $\theta = (g, r_{11}, r_{21}, r_{22}, \phi, \delta, \gamma)$ and as Figure 5 for $(\mathcal{E}(r_f), \sqrt{\text{Var}(r_f)}, \mathcal{E}(r_d), \sqrt{\text{Var}(r_d)})$. Elimination of the requirement that the habit model confront the conditional heterogeneity in the data results in a dramatic left shift in the marginal posterior of $\mathcal{E}(r_d)$ and a dramatic variance reduction in the posterior for $\sqrt{\text{Var}(r_d)}$. This, following the logic of Section 4, can be regarded as strong evidence against the VAR specification of the statistical model.

Figure 6 about here

It is also of interest to note that it is the preference parameters (ϕ, δ, γ) of the scientific model that control the GARCH parameters $(\rho_{11}, \rho_{21}, \rho_{31}, \rho_{12}, \rho_{22}, \rho_{32}, \tau)$ of the statistical model with γ having by far the most influence. This can be discovered by inspecting the

map $g : \theta \mapsto \eta$ but is best seen by inspecting plots of the conditional correlation between annual stock returns and consumption growth for the year 2002 against the parameters $\theta = (g, r_{11}, r_{21}, r_{22}, \phi, \delta, \gamma)$ of the scientific model as are shown in Figure 6.

The conditional correlation is a functional of the statistical model $f(\cdot|\cdot, \eta)$ determined by the map $\eta = g(\theta)$ and computed by applying the GARCH recursion (4) to the data. The last variance matrix given by the recursion is the conditional variance for the year 2002. The dots show the conditional correlation when the statistical model is GARCH and the circles when VAR. The difference between these two curves is the GARCH effect. In each panel, a θ_i is varied and the remaining θ held fixed at their posterior means. The solid vertical line is the posterior mean of θ_i when the statistical model is GARCH and the dashed line when VAR. The point where the solid line crosses the dots gives the conditional correlation when $\eta = g(\theta)$ is evaluated at the posterior mean under GARCH; similarly VAR for the dashed line and circles.

5.6 Model Assessment: Scientific Prior Relaxed

We now apply the methodology described in Section 4 to the habit model.

The functionals of interest are the mapping Υ_1 of $f(\cdot|\cdot, \eta)$ to the mean return on the stock portfolio over the period 2002–2102 and the mapping Υ_2 to the conditional correlation between the return on the stock portfolio and consumption growth for the year 2002. The conditioning event for both functionals is the 73 years of observed data. Υ_1 is computed from a realization obtained by simulating the GARCH model over the period 2002–2102; Υ_2 is obtained from the variance matrix for the year 2002 computed as described at the end of Subsection 5.5. Both depend on the data and η ; Υ_1 also depends on an initial seed that was the same for each η .

Figure 7 about here

Using the methods proposed in Section 4, we computed MCMC chains for three values of κ (1,20,100). The stationary distribution of each chain is the posterior for η under prior κ . We then evaluated the two functionals (Υ_1, Υ_2) at each η in the chain and plotted their densities in Figure 7. κ increases as we go down the rows. The left hand panels show prior

and posterior densities for the mean return and the right hand panels display the same for the correlation. The top right hand panel and the middle right hand panel are analogous to the two right hand panels of Figure 2. For small κ , we have a tight prior and the prior and posterior are similar (compare upper left hand panel of Figure 7 to Figure 5).

For larger κ , the posterior shifts. We are all familiar with the meaning of a correlation and can agree that in the three right hand plots the posterior distribution of the correlation shifts dramatically as κ increases while the prior remains reasonable. As κ increases we place less weight on η close to the scientific model and the posterior shifts. This gives clear evidence against the scientific model in a simple, interpretable way. (The correlation also has substantive meaning because a high positive correlation implies that risk averse investors will require high expected returns to induce them to invest.)

The mean stock return, unlike a correlation, has a purely substantive meaning without having an independent statistical meaning. Nonetheless, coping with retirement plan options and the laws governing bequests has made the notion of a mean return over a long planning horizon meaningful to most of us. For the mean return, we also see that as κ increases, the distribution shifts. The posterior mean of the mean return over the hundred year horizon is 0.1 in the top left panel and 0.08 in the bottom left so that while the shift may not appear as dramatic as for the correlation functional it is substantively large. Also, the standard deviation increases from 0.0127 to 0.0172. In the context of the asset pricing literature, this is strong substantive evidence against the scientific model.

Figure 7 focuses on the marginal priors and posteriors of two particular functionals of interest. Of course, as κ changes, the entire sixteen dimensional prior and posterior of η is changing. To get a sense of how the posterior is moving, we consider each κ as a choice of model and compute the posterior probability for the different models using the Newton and Raftery (1994) method.

Table 2 about here

The simple version of the Newton-Raftery method which we have employed uses the harmonic mean of the likelihoods of the η visited by each MCMC chain (for each κ). This method is known to have poor numerical properties and the values may have substantial

error. We take the values in Table 2 as a simple summary of the basic fact that as we loosen up the prior and allow the posterior to move away from the scientific manifold, it is able to find much higher likelihoods. This is fairly strong statistical evidence against the habit model but many of the contributors to the asset pricing literature attach little weight to general statistical tests of this nature and are far more likely to be persuaded by Figure 7 than by Table 2.

6 Conclusion

We considered a consumption based asset pricing model that uses habit persistence to overcome the known statistical inadequacies of the classical consumption based asset pricing model. We found that the habit model fits reasonably well and agrees with results reported in the literature if conditional heterogeneity is suppressed by a sharp prior but that it does not fit nor do results agree if conditional heterogeneity is allowed to manifest itself. We also found that it is the preference parameters of the model that are most affected by the presence or absence of conditional heterogeneity, especially the risk aversion parameter.

To obtain these results we proposed and implemented a general purpose Bayesian methodology for the analysis of complex models from the sciences. It relies on the ability to simulate from the scientific model, upon the availability of substantive prior information, and upon the willingness to use that prior information. Analysis is carried out by means of a richly parameterized statistical model $f(\cdot, \eta)$ that is viewed as being the correct description of the distribution of the data. The scientific model is viewed as imposing a severe prior $\pi_\kappa(\eta)$ on the statistical model and the proposed methodology directly implements this view. The correctness of the scientific model is assessed by relaxing $\pi_\kappa(\eta)$ and assessing the posteriors of scientifically meaningful functionals of $f(\cdot, \eta)$. If location does not change more than scientifically meaningful magnitudes as $\pi_\kappa(\eta)$ is relaxed, then the model is supported. If scale increases as $\pi_\kappa(\eta)$ is relaxed, then the model has empirical content.

In empirical investigations, scientists often use methods that are more like a subjective calibration of models wherein model parameters are adjusted so that statistics from model simulations subjectively match statistics computed from data. One might say that what is proposed here is akin to calibration in spirit but rather than being totally ad hoc is conducted

within a coherent philosophical framework.

References

- Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant (2002), “Quadratic Term Structure Models: Theory and Evidence,” *The Review of Financial Studies* 15, 243–288.
- Bansal, Ravi, A. Ronald Gallant, and George Tauchen (2004), “Rational Pessimism, Rational Exuberance, and Markets for Macro Risks,” Working paper, Fuqua School of Business, Duke University.
- Campbell, John Y. (2002), “Consumption Based Asset Pricing,” Manuscript, Department of Economics, Harvard, Cambridge MA.
- Campbell, John Y., and John Cochrane (1999) “By Force of Habit: A Consumption-based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy* 107, 205–251.
- Chernov, Mikhail, A. Ronald Gallant, Eric Ghysels, and George Tauchen (2003), “Alternative Models for Stock Price Dynamics,” *Journal of Econometrics* 116, 225–257.
- Del Negro, Marco, and Frank Schorfheide (2004), “Priors from General Equilibrium Models for VARS,” *International Economic Review* 45, 643–673.
- Duffee, Gregory R., and Richard H. Stanton (2005), “EMM and Indirect Inference Estimation of Near Unit-root Processes,” Manuscript, Haas School of Business, University of California, Berkeley, CA.
- Duffie, D., and K. J. Singleton (1993), “Simulated Moments Estimation of Markov Models of Asset Prices,” *Econometrica* 61, 929–952.
- Gallant, A. R., and J. Long (1997), “Estimating Stochastic Differential Equations Efficiently by Minimum Chi-squared,” *Biometrika* 84, 125–141.

- 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.
- Gourieroux, C., A. Monfort and E. Renault (1993) "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.
- Guveneren, Fatih (2003), "A Parsimonious Macroeconomic Model for Asset Pricing: Habit Formation or Cross-sectional Heterogeneity?" Manuscript, Department of Economics, University of Rochester.
- Ingram, B. F., and B. S. Lee (1991), "Simulation Estimation of Time Series Models," *Journal of Econometrics* 47, 197–250.
- Leon, S., Tsiatis, A.A., and Davidian, M. (2003) "Semiparametric Estimation of Treatment Effect in a Pretest-Posttest Study," *Biometrics* 59, 1046-1055.
- Lucas, R. E., Jr. (1978), "Asset Prices in and Exchange Economy, *Econometrica* 46, 1429–1445.
- Mallet, A., F. Mentré, J-L. Steimer, and F. Lokiec (1988), "Nonparametric Maximum Likelihood Estimation for Population Pharmacokinetics, with Application to Cyclosporine," *Journal of Pharmacokinetics and Biopharmaceutics* 16, 311–327.
- Mishkin, F.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.
- Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap." *Journal of the Royal Statistical Society (B)* 56, 3-48.
- Olsen, L. F., and W. M. Schaffer (1990), "Chaos Versus Noisy Periodicity: Alternative Hypotheses for Childhood Epidemics," *Science* 249, 499–504.

- Smith, A. A. (1993) “Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions,” *The Journal of Applied Econometrics* 8, S63–S84.
- Song, X., Davidian, M. , and Tsiatis, A.A. (2002), “A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data. *Biometrics* 58, 742-753.
- Theil, Henri, and Arthur S. Goldberger (1961), “On Pure and Mixed Estimation in Economics,” *International Economic Review* 2, 65–78.

Tables and Figures

Table 1. Parameter Estimates

Parameter	EMM Estimates		Bayes-VAR			Bayes-GARCH		
	Estimate	Std. Err.	Mode	Mean	Std. Dev.	Mode	Mean	Std. Dev.
Monthly								
g	0.002116	0.000250	0.001639	0.001739	0.000258	0.001803	0.001780	0.000684
ψ_{11}	0.006151	0.000896						
ψ_{22}	0.036503	0.007716						
ρ_s	0.971900	0.015449						
μ_{dc}	-3.3587	0.0380						
r_{11}			0.006753	0.007326	0.000627	0.007254	0.007417	0.001903
r_{12}			0.001350	0.001451	0.000403	0.001350	0.001336	0.001068
r_{22}			0.003125	0.008109	0.006205	0.003125	0.018852	0.034435
ϕ	0.9853	0.0026	0.9861	0.9857	0.0024	0.9804	0.9818	0.0095
δ	0.9939	0.0005	0.9955	0.9937	0.0030	0.9898	0.9907	0.0070
γ	0.8386	0.2462	0.5726	0.9463	0.6179	1.0744	1.1747	1.7638
Annualized								
g	2.539	0.0866	1.9672	2.087	0.0895	2.164	2.136	0.2369
σ	2.1308		2.3857	2.5870	0.2202	2.5589	2.6106	0.2513
ρ	0.1650		0.1960	0.1943	0.0508	0.1830	0.1773	0.0507
σ_w	12.9118		1.0825	2.8090	2.1496	1.0825	6.5306	4.4984
ϕ	0.8372	0.0090	0.8450	0.8412	0.0084	0.7890	0.8023	0.0328
δ	0.9292	0.0018	0.9477	0.9269	0.0102	0.8845	0.8934	0.0244
γ	0.8386	0.2462	0.5726	0.9463	0.6179	1.0744	1.1747	1.7638
$\mathcal{E}(r_{dt}^a)$	6.54		6.00	7.58	0.6930	11.14	10.45	0.5487
$SDev(r_{dt}^a)$	16.9		20.49	21.48	1.5343	24.22	26.14	2.4735
$\mathcal{E}(r_{ft}^a)$	1.07		1.20	1.16	0.1389	1.21	0.99	0.1451
$SDev(r_{ft}^a)$	3.23		0.33	0.27	0.1388	0.42	0.38	0.1489
	$\chi^2(5) = 7.11$ (0.21)		reps = 800,000 by 8			reps = 800,000 by 8		

Note: EMM estimates, from Bansal, Gallant, and Tauchen (2004), use data on the price dividend ratio and the consumption dividend ratio in addition to consumption growth and stock returns. EMM estimates impose $\mathcal{E}(r_f^a) = 0.89\%$ and cointegration among consumption, dividends, and price. Variance parameters relate as

$$\text{Var} \begin{pmatrix} c_t - c_{t-1} \\ d_t - d_{t-1} \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho \sigma \sigma_w \\ \text{sym} & \sigma_w^2 \end{pmatrix} = \begin{pmatrix} r_{11}^2 + r_{12}^2 & r_{12} r_{22} \\ \text{sym} & r_{22}^2 \end{pmatrix} = \begin{pmatrix} \psi_{11}^2 & \psi_{11}^2 \\ \text{sym} & \psi_{11}^2 + 2\psi_{22}^2(1 - \rho_s)^{-1} \end{pmatrix}$$

where EMM parameters μ_{dc} , ρ_s , and ψ_{22} are the location, autoregressive, and scale parameters of the cointegration relation between c_t and d_t . In the data, the mean of r_d^a is 6.02% and the standard deviation is 19.29%; for consumption growth the values are 1.95% and 2.24%. The mode of $\theta = (g, \sigma, \rho, \sigma_w, \delta, \gamma)$ is the mode of the multivariate posterior. All standard deviations are from marginal posteriors.

Table 2. Posterior Model Probabilities

κ	Model Probability
0	1.906874e-10
1	1.909023e-07
5	1.260361e-06
10	3.408552e-05
15	8.800083e-05
20	0.003628554
50	0.3379201
100	0.6583278

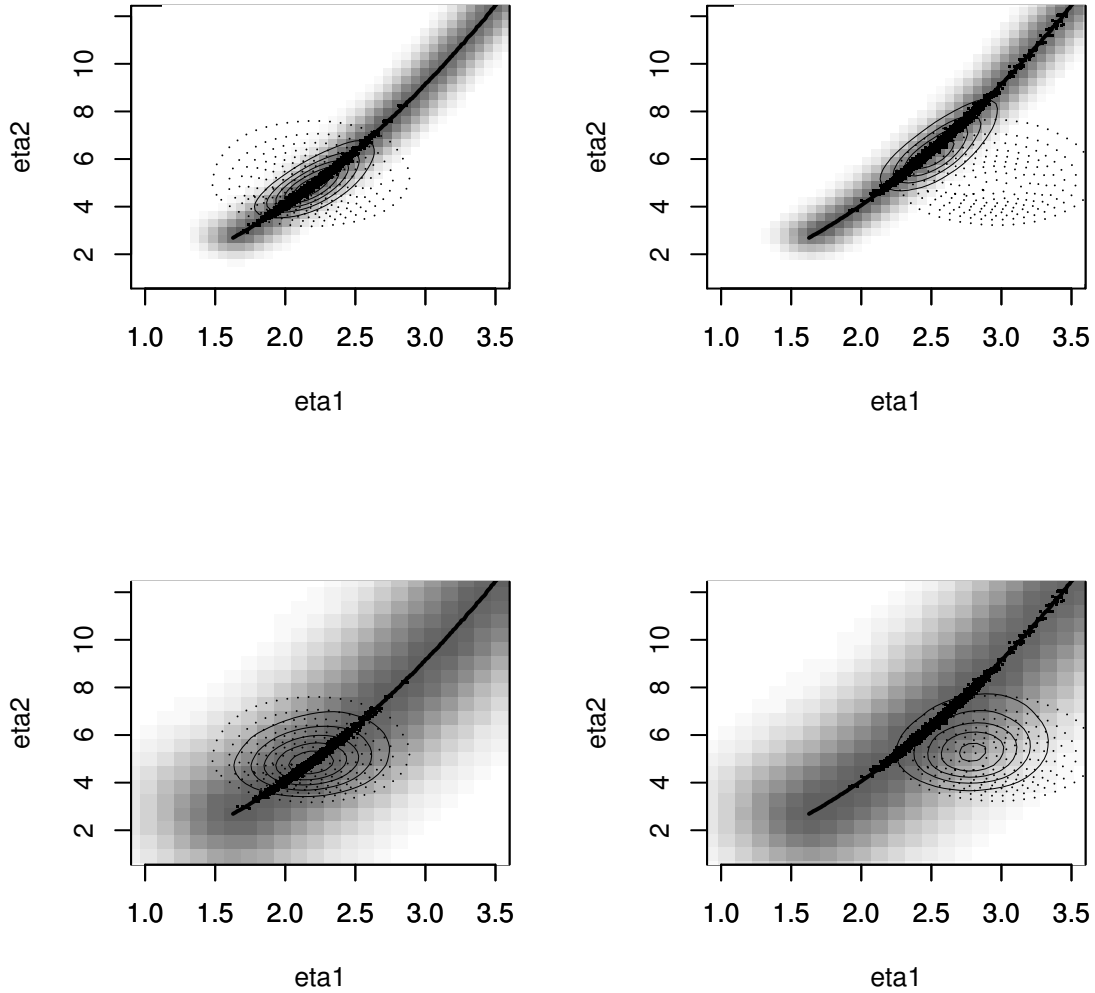
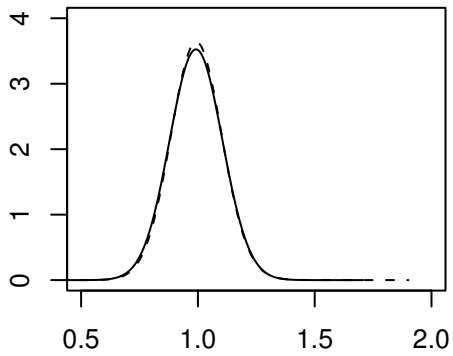
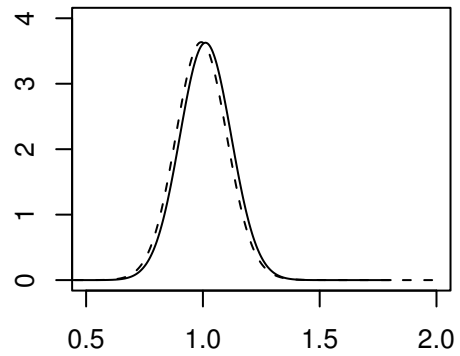


Figure 1. Priors and posteriors for the statistical model, tinker toy example.

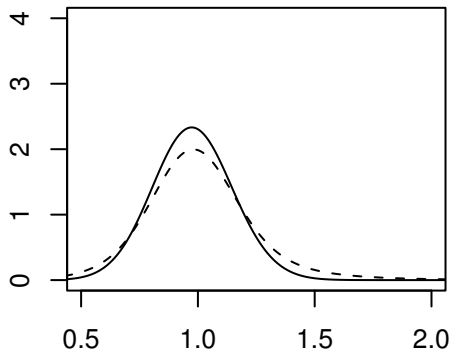
The dotted lines are contours of the likelihood of the statistical model $f(y|x, \eta)$ of the tinker toy example. The line is the prior on η determined by the implied map $\eta = g(\theta)$ from the parameters θ of scientific model $p(y|x, \theta)$ to the parameters η of the statistical model. In the left panels the scientific model is true, in the right it is false. The thickness of the line is proportional to the posterior of η . The prior $\pi(\eta)$ can be relaxed as indicated by the shading. The lower panels are more relaxed than the upper. The solid contours show the posterior under the relaxed prior. Relaxation causes the contours to enlarge in all cases. When the scientific model is false, the posterior shifts in search of the likelihood.



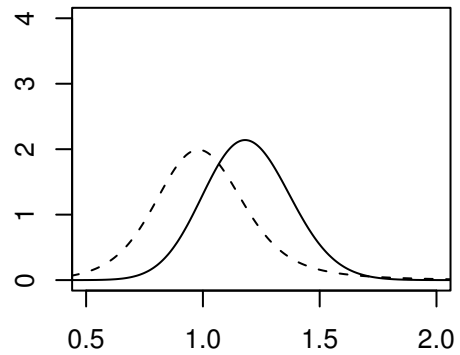
kappa=5, restriction true



kappa=5, restriction false



kappa=100, restriction true



kappa=100, restriction false

Figure 2. Priors and posteriors for a functional of the statistical model, tinker toy example. The posterior of the coefficient of variation for the tinker toy example is the solid line; the dashed line is prior. In the left panels the scientific model is true, in the right it is false. The prior is more relaxed in the lower panels than it is in the upper panels. The panels correspond to those of Figure 1

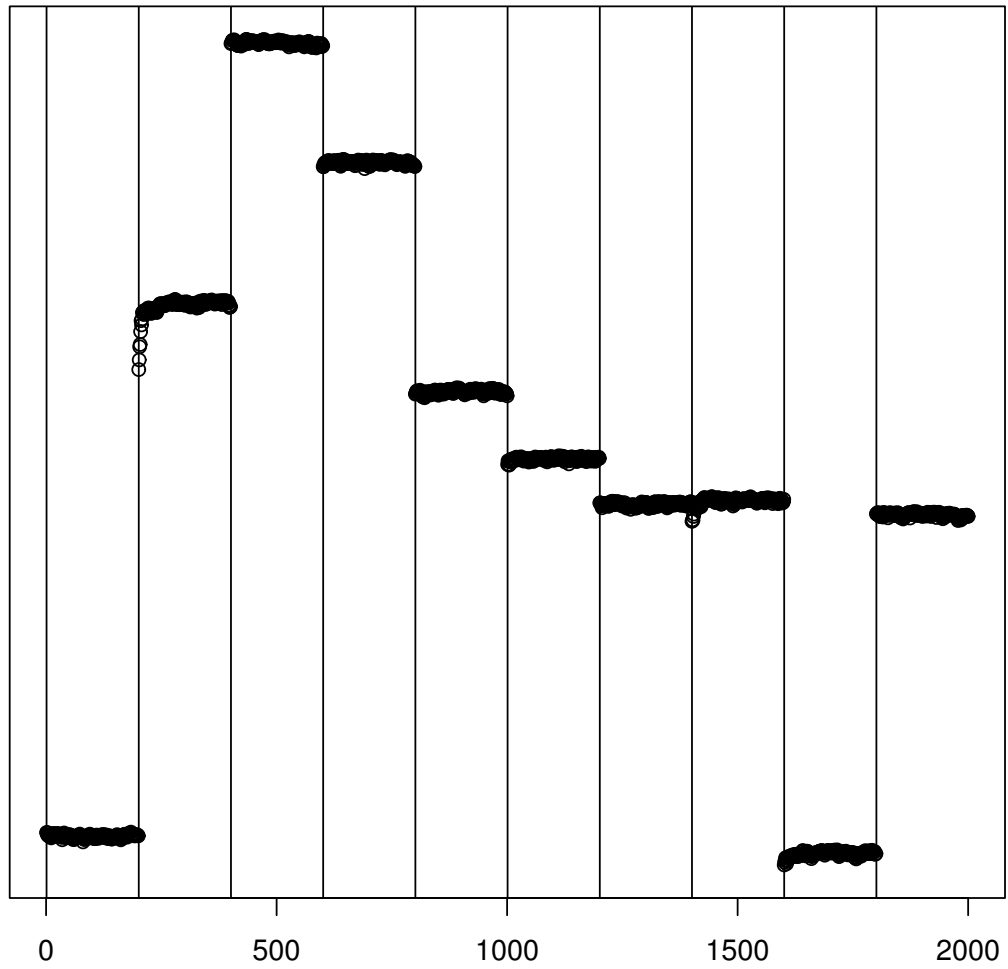


Figure 3. η Chain for the habit model. Ten successive runs of the η chain. Each run is 200 iterations. The log-likelihood of the simulated data set is plotted on the vertical axis. Vertical bars mark where θ changes. Jumps are because $\{\hat{y}_t\}_{t=1}^N$ changes at each vertical bar.

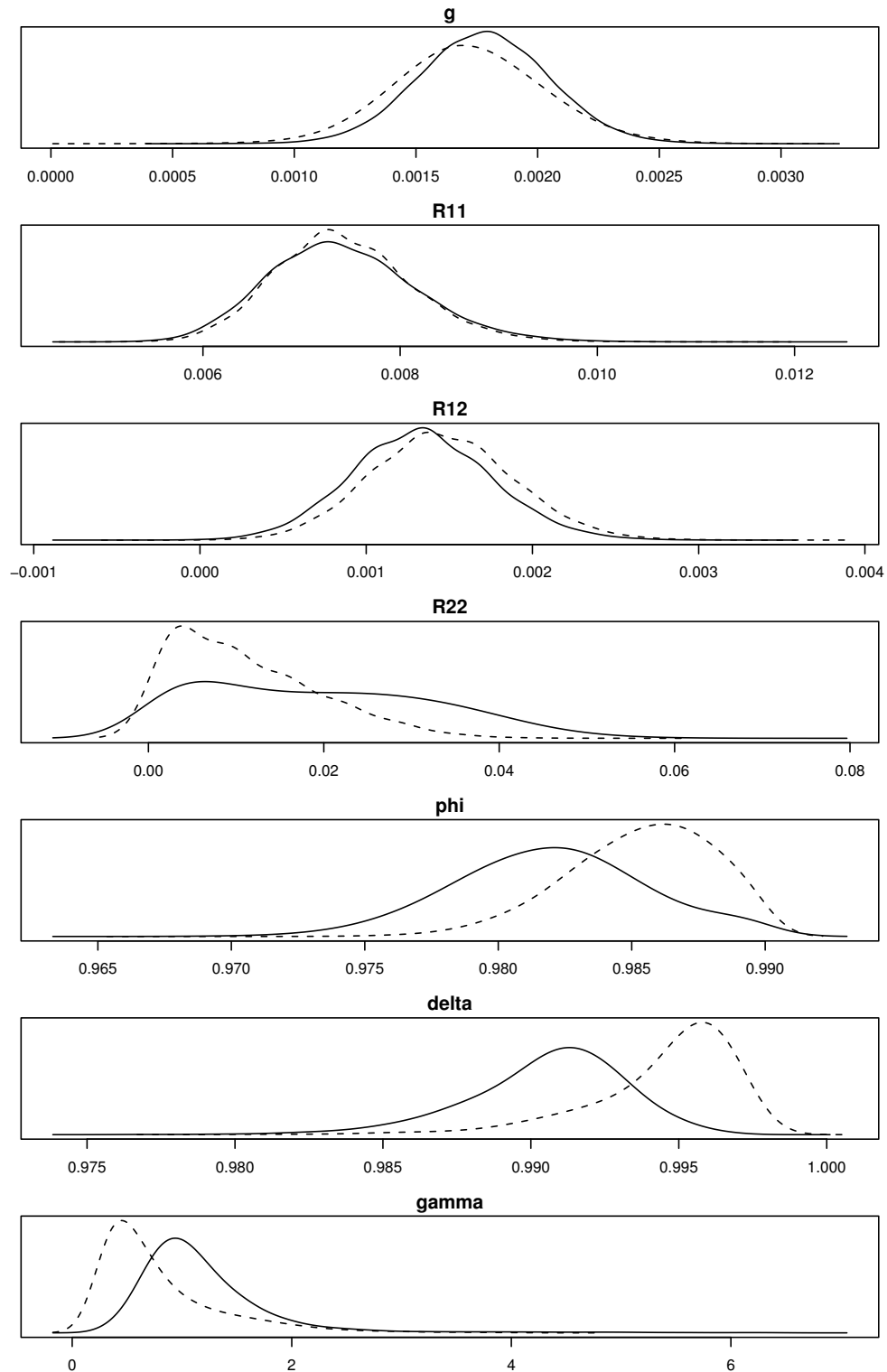


Figure 4. Density of the MCMC chain for θ . Shown is a kernel density estimate from iterates 1 to 800,000 by 8 of the MCMC chain for $\theta = (g, r_{11}, r_{12}, r_{22}, \psi, \delta, \gamma)$ at the monthly frequency.

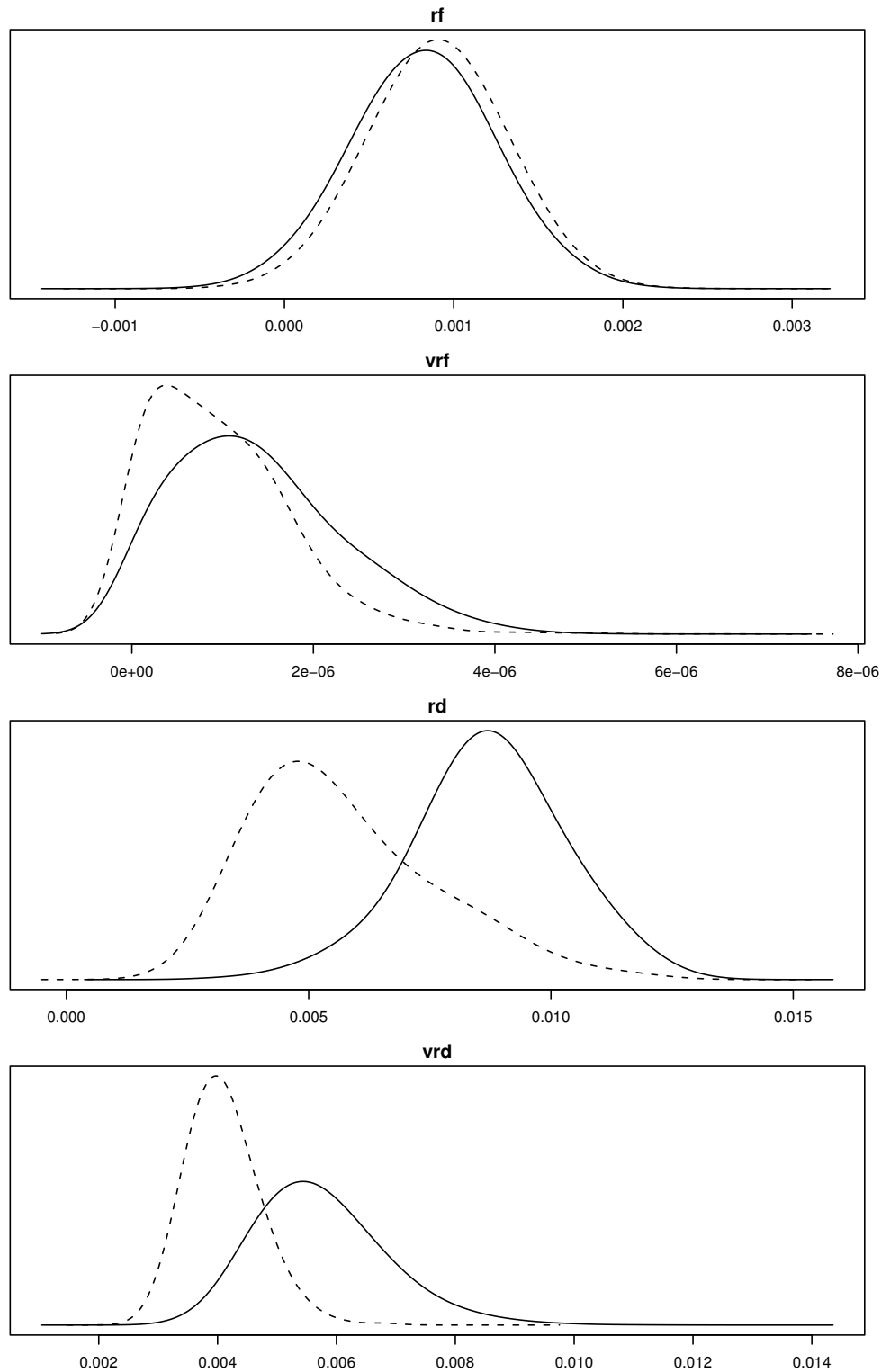


Figure 5. Density of the MCMC chain for Returns. Shown is a kernel density estimate of iterates 1 to 800,000 by 8 of the MCMC chain for $\mathcal{E}(r_f)$, $\sqrt{\text{Var}(r_f)}$, $\mathcal{E}(r_d)$, and $\sqrt{\text{Var}(r_d)}$ at the monthly frequency.

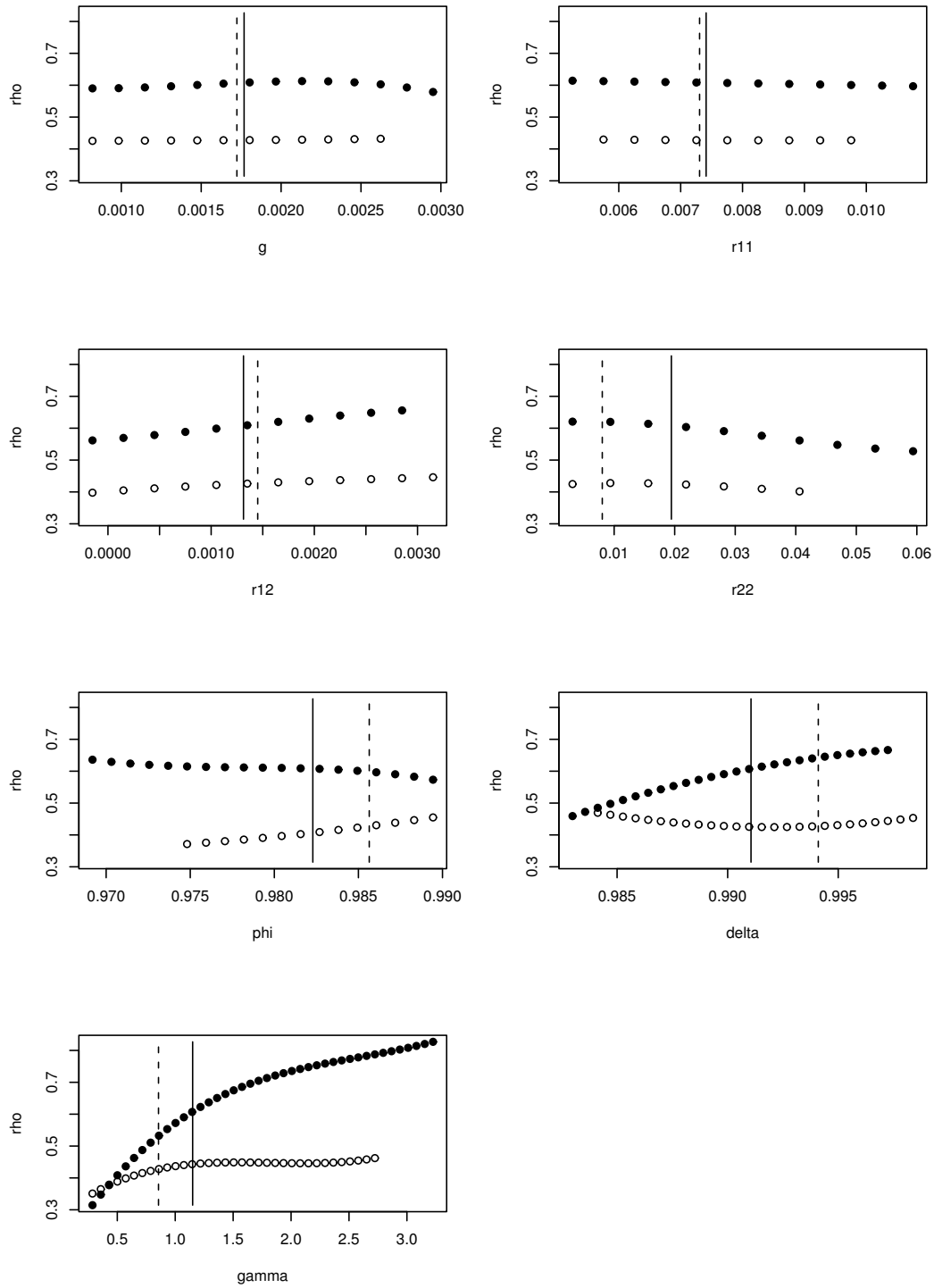


Figure 6. Conditional Correlation. The conditional correlation between annual consumption growth and stock returns for 2002 plotted against scientific model parameters $\theta = (g, r_{11}, r_{21}, r_{22}, \phi, \delta, \gamma)$ as determined from the map $g : \theta \mapsto \eta$. Dots are for the GARCH statistical model; circles for the VAR. The solid vertical line is the GARCH posterior mean and the dashed VAR.

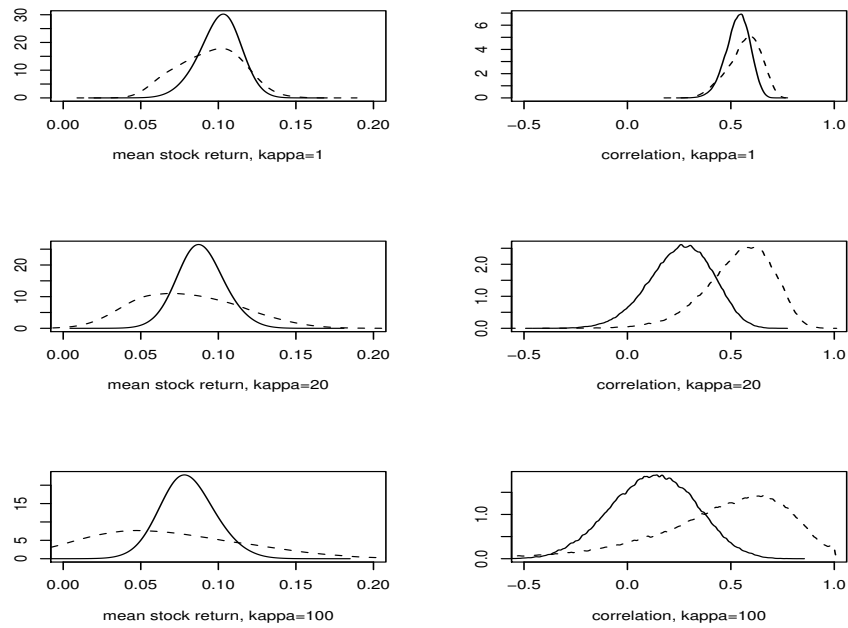


Figure 7. Priors and posterior for two functionals of the statistical model, habit model example. The statistical model is a bivariate GARCH model on annual consumption growth and stock returns. The scientific model is a habit persistence asset pricing model. The data covers 1929–2001. The left panels are the mean stock return for the period 2002–2102 implied by the GARCH model. The right panels are the conditional correlation between stock returns and consumption growth implied by the GARCH model for the year 2002. The prior is more relaxed in the lower panels than it is in the upper panels.