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

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

Software and User's Guide: ftp://ftp.econ.duke.edu/pub/arg/gsm

Main Point: Things Change

- In the past, frequentist methods could solve problems that Bayesian methods could not.
- Now the converse is true.
- Especially when data are sparse.
- Because of modern languages, modern parallel equipment, and modern numerical methods.

Main Results: Habit Model

- Fits the data and agrees with frequentist and calibration results if conditional heterogeneity is suppressed by a sharp prior.

- Does not fit the data nor agree with frequentist or calibration results if conditional heterogeneity can manifest itself.

- It is the preference parameters, especially γ, that are most affected by the presence or absence of conditional heterogeneity.

Main Idea: Statistical, Scientific Models

- A statistical model determines the likelihood.

- A prior is placed on the statistical model by imposing
  1. a scientific model, which restricts its parameters to a lower dimensional manifold
  2. a prior on the parameters of the the scientific model
  3. a prior on functionals of the the scientific model

- This setup is especially advantageous when data are too sparse to estimate the statistical model without imposing a diffusion of this prior that attracts the parameters of the statistical model to the lower dimensional manifold.

Tinker Toy Example

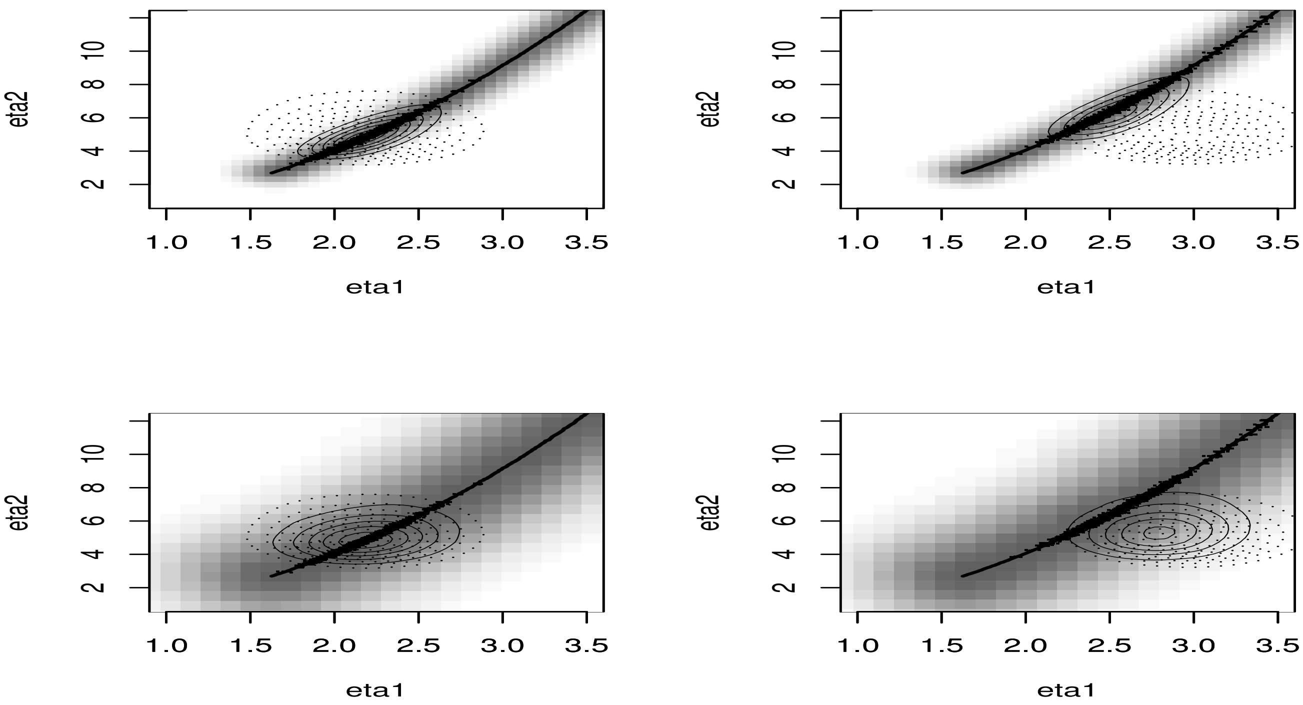
Scientific Model: $p(y \,|\, \theta) = n(y \,|\, \theta, \theta^2)$

Statistical Model: $f(y \,|\, \eta) = n(y \,|\, \eta_1, \eta_2)$

Implied map: $g : \theta \mapsto \eta = (\eta_1, \eta_2) = (\theta, \theta^2)$

Later the scientific model will be a habit persistence asset pricing model, the statistical model will be GARCH, and y will be consumption growth and stock returns.

Fig 1. The Scientific and Statistical Model



The 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.

Determination of Model Adequacy

- Relax the prior determined by the scientific model.

- The parameters of the statistical model are now free to move off the lower dimensional manifold into their natural higher dimensional parameter space.

- But the prior retains enough influence to overcome data sparseness.

- Model adequacy is determined by observing if parameters or functionals of the statistical model that have scientific relevance change.

Tinker Toy Example

Scientific Model: $p(y \,|\, \theta) = n(y \,|\, \theta, \theta^2)$

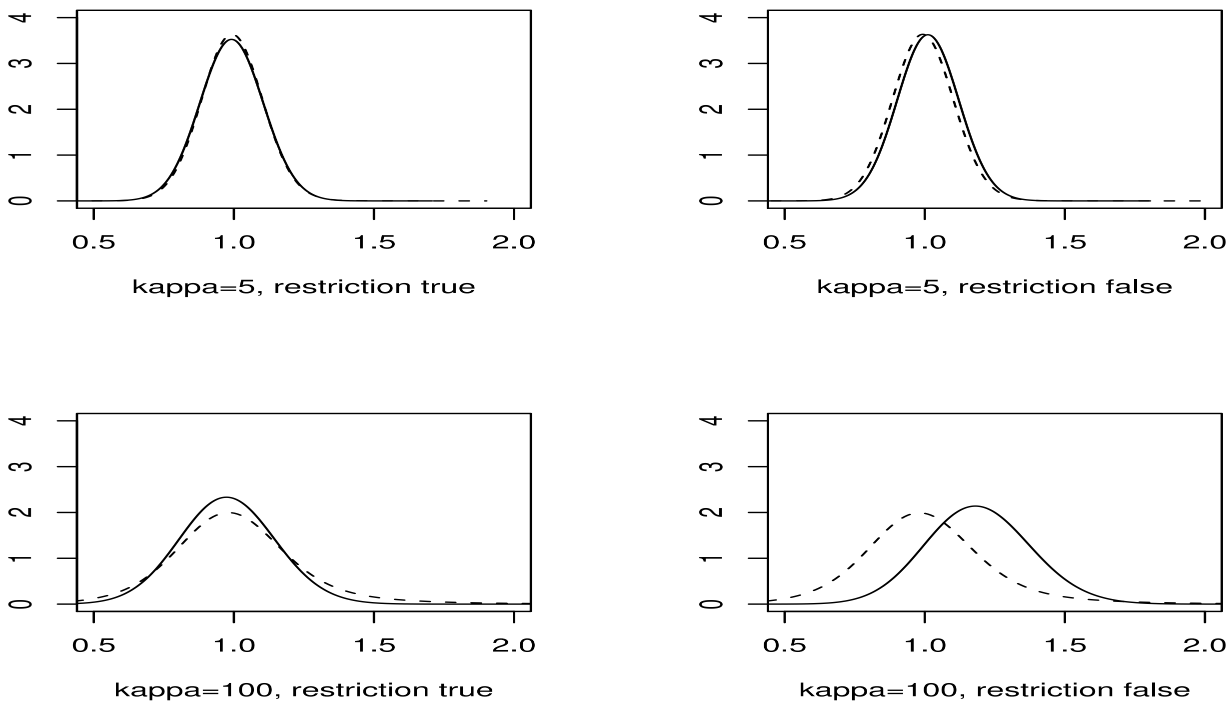
Statistical Model: $f(y \,|\, \eta) = n(y \,|\, \eta_1, \eta_2)$

Implied map: $g : \theta \mapsto \eta = (\eta_1, \eta_2) = (\theta, \theta^2)$

κ-Prior: $\pi_\kappa(\eta) \propto \exp\left(-\frac{1}{2\kappa} \min \{\|g(\theta) - \eta\|^2 : \theta \in \Theta\}\right)$

Later the scientific model will be a habit persistence asset pricing model, the statistical model will be GARCH, and y will be consumption growth and stock returns.

Fig 2. Determination of Model Adequacy



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

The Devil is in the Details

- The details of the implementation are the rest of the talk.

- We shall also discover precisely why the habit model fails.

- Implementation would be impossible without recent advances in object oriented programming languages and MCMC algorithms.

- The availability of a Linux cluster, while not essential, is helpful.

Common Characteristics of Scientific Models

- Likelihood not available
  - Partially observed states or discretely sampled.
  - Expressed as a system of differential equations with system noise.
  - Latent variables.
- Constraints on functionals of the model.
- Model can be simulated.

Examples of Such Models

- Epidemiology: The SEIR model determines those susceptible, exposed, infected, and recovered from a disease whereas usually data are from case reports that report only those infected (Olsen and Schaffer, 1990).

- Continuous and discrete time stochastic volatility models of speculative markets from finance (Ghysels, Harvey, and Renault, 1995),

- General equilibrium models from economics (Gennotte and Marsh, 1993),

- Compartment models from pharmacokinetics (Mallet, Mentré, Steimer, and Lokiec, 1988).

Example Used Here

- Habit persistence asset pricing model.

- Has these four characteristics:
  - Likelihood not available.
  - Prior information $\pi(\theta)$ on model parameters is available.
  - Prior information $\pi(\theta, \psi)$ on functionals is available.
  - Model can be simulated.

Habit Persistence Asset Pricing Model

Driving Processes

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

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

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

The time increment is one month. Lower case denotes logarithms of upper case quantities; i.e. $c_t = \log(C_t)$, $d_t = \log(D_t)$. From Campbell and Cochrane (1999).

Simulation

- The consumption and dividend processes are trivially easy to simulate.

- We need two returns processes: the risk free rate and the returns on the asset that pays the dividend stream.

- To compute these returns, we assume an economy with a representative agent endowed with a habit persistence utility function.

Habit Persistence Asset Pricing Model

Utility function

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

Habit persistence

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

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

$\mathcal{E}_t$ is conditional expectation with respect to S_t, S_{t-1}, ... Lower case denotes logarithms of upper case quantities: $s_t = \log(S_t)$. $\bar{S}$ and $s_{\max}$ can be computed from model parameters $\theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma)$ as $\bar{S} = \sigma\gamma/\phi$, $s_{\max} = \bar{s}(1 - \bar{S}^2)/2$. From Campbell and Cochrane (1999).

Simulation

- To generate returns, we must solve the agent's optimization problem, which is the next task.

Simulation

The computational problem is this: We must find the price-dividend function $V(\cdot)$ that solves

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

and then evaluate $V(\cdot)$ over our simulated values $\{C_t, D_t, S_t\}_{t=1}^N$ to get the corresponding returns process $\{r_{dt}\}_{t=1}^N$ using

$$\begin{aligned} P_{t-1} &= D_{t-1} \cdot V(S_{t-1}) \\ P_t &= D_t \cdot V(S_t) \\ r_{dt} &= \log \left(\frac{P_t + D_t}{P_{t-1}} \right) \end{aligned}$$

Solution Method

To simulate the model one is obliged to generate a long simulation of consumption, dividends, and surplus ratio in both logs and levels

$$\{c_t\}_{t=1}^{N}\,\{C_t\}_{t=1}^{N}\,N \sim 50,000$$

$$\{s_t\}_{t=1}^{N}\,\{S_t\}_{t=1}^{N}\,N \sim 50,000$$

$$\{d_t\}_{t=1}^{N}\,\{D_t\}_{t=1}^{N}\,N \sim 50,000$$

regardless. The idea is to take advantage of their existence to solve the asset pricing equations.

Solution Method (A Colocation Method)

Posit that the log price-dividend function

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

can be represented as a low order expansion in a set of basis functions:

$$v(s) = \alpha_0 + \alpha_1\phi_1(s) + \alpha_2\phi_2(s) + \cdots + \alpha_K\phi_K(s)$$

Define the instrumental variable:

$$Z_t = [1, \phi_1(S_t), \phi_2(S_t), ..., \phi_K(S_t)]'$$

Solution Method (A Colocation Method)

The conditional Euler condition

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

implies the unconditional instrumental variables condition

$$0 = \mathcal{E} \left(Z_t \left\{ e^{v(s_t)} - \delta \left(\frac{S_{t+1} C_{t+1}}{S_t C_t} \right)^{-\gamma} \left(\frac{D_{t+1}}{D_t} \right) \left[1 + e^{v(s_{t+1})} \right] \right\} \right)$$

which can be computed as

$$0 = \frac{1}{N} \sum_{t=1}^{N} \left(Z_t \left\{ e^{v(s_t)} - \delta \left(\frac{S_{t+1} C_{t+1}}{S_t C_t} \right)^{-\gamma} \left(\frac{D_{t+1}}{D_t} \right) \left[1 + e^{v(s_{t+1})} \right] \right\} \right)$$

for large N. This is a nonlinear equation in $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_K$ because

$$v(s) = \alpha_0 + \alpha_1 \phi_1(s) + \alpha_2 \phi_2(s) + \cdots + \alpha_K \phi_K(s).$$

Solution Method (A Colocation Method)

Solve the equations

$$0 = \frac{1}{N}\sum_{t=1}^{N}\left(Z_t\left\{e^{v(s_t)} - \delta\left(\frac{S_{t+1}C_{t+1}}{S_tC_t}\right)^{-\gamma}\left(\frac{D_{t+1}}{D_t}\right)\left[1 + e^{v(s_{t+1})}\right]\right\}\right)$$

where

$$v(s) = \alpha_0 + \alpha_1\phi_1(s) + \alpha_2\phi_2(s) + \cdots + \alpha_K\phi_K(s).$$

for α by Newton's method.

A concern to be addressed later: For each parameter vector $\theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma)$ for which returns are required, Newton's method needs a good starting value for α to assure convergence.

Solution Method Summary

Approximate the log price-dividend function by a series expansion

$$v(s_t) = \alpha_0 + \alpha_1\phi_1(s_t) + \alpha_2\phi_2(s_t) + \cdots + \alpha_K\phi_K(s_t)$$

where the Hermite polynomials are a good choice for ϕ_i.

Define the instrumental variables

$$Z_t = [1, \phi_1(S_t), \phi_2(S_t), ..., \phi_K(S_t)]'$$

Simulate $\{C_t, D_t, S_t\}_{t=1}^{N}$ and compute

$$0 = \frac{1}{N}\sum_{t=1}^{N}\left(Z_t\left\{e^{v(s_t)} - \delta\left(\frac{S_{t+1}C_{t+1}}{S_tC_t}\right)^{-\gamma}\left(\frac{D_{t+1}}{D_t}\right)\left[1 + e^{v(s_{t+1})}\right]\right\}\right)$$

which is a nonlinear equation in $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_K$.

Solve for $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_K$ using Newton's method.

Habit Persistence Asset Pricing Model

Risk Free Rate

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

r_{ft} is the logarithmic return on an asset that pays one real dollar one month hence with certainty. From Campbell and Cochrane (1999).

Solution method is similar to the foregoing.

Model Output

For given model parameters

$$\theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma)$$

the model produces simulated consumption and returns data at an annual frequency:

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

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

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

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

Available Data

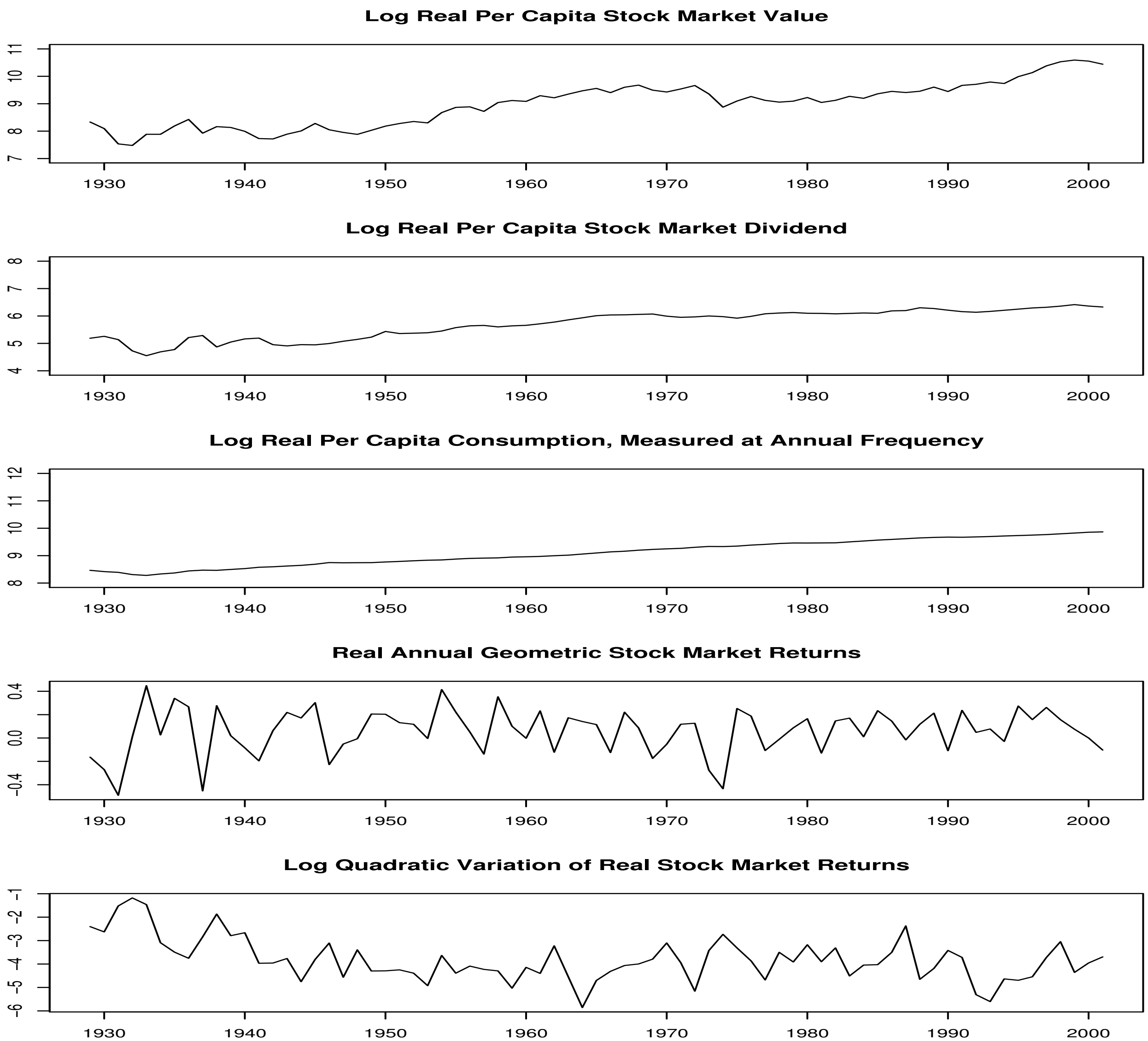
Annual observations 1929–2001, 72 years, on

P^a_{dt}	end-of-year per capita stock market value
D^a_t	annual aggregate per capita dividend
C^a_t	annual per capita consumption
r^a_{dt}	annual real geometric return
Q^a_t	annual quadratic variation

Data are real, i.e. inflation adjusted.

Source: Bansal, R., A. R. Gallant, and G. Tauchen (2003). "Rational Pessimism, Rational Exuberance, and Markets for Macro Risks," Working paper, Department of Economics, Duke University.

Fig 3. Data



Log Real Per Capita Stock Market Value
Log Real Per Capita Stock Market Dividend
Log Real Per Capita Consumption, Measured at Annual Frequency
Real Annual Geometric Stock Market Returns
Log Quadratic Variation of Real Stock Market Returns
1930
1940
1950
1960
1970
1980
1990
2000

Cointegrating Relationships

$$\begin{aligned} p^a_{dt} - d^a_t &= I(0) && \text{Well documented in the literature} \\ d^a_t - c^a_t &= I(0) && \text{Bansal, Gallant, Tauchen (2003)} \\ c^a_t - c^a_{t-1} &= I(0) && \text{Well documented in the literature} \end{aligned}$$

Jointly Stationary Data for Estimation

Used by methods proposed here: Contemporaneous and one lag of

$$\begin{pmatrix} c_t^a - c_{t-1}^a \\ r_{dt}^a \end{pmatrix}$$

Used by estimates compared with: Contemporaneous and one lag of

$$\begin{pmatrix} d_t^a - c_t^a \\ c_t^a - c_{t-1}^a \\ p_{dt}^a - d_t^a \\ r_{dt}^a \end{pmatrix}$$

Types of Prior Information

Analytical: Positivity restrictions on positive valued parameters and non-explosive restrictions on autoregressive parameters: $\pi_1(\theta)$

Numerical: Existence of solution to Euler condition: $\pi_1(\theta, \psi)$

Functional: Requires a successful simulation to compute: $\pi_2(\theta, \psi)$

$$P\left(\left|\mathcal{E}(r_f^a) - 0.89\%\right| < 1\%\right) = 0.95$$

where $P(\cdot)$ is standard normal in the application.

Prior Information Grouped by Cost

1. Can be determined cheaply knowing model parameters alone and evaluates to 0 or 1

$$\pi_1(\theta) \qquad\qquad \theta = (g, \sigma, \rho, \sigma_w, \phi, \delta, \gamma)$$

2. Requires a simulation to determine and evaluates to 0 or 1

$$\pi_1(\theta, \psi) \qquad\qquad \psi_1 \mapsto (success, failure)$$

3. Requires a simulation to determine and $0 \le \pi_2 < \infty$

$$\pi_2(\theta, \psi) \qquad\qquad \psi_2 = \mathcal{E}(r^a_{ft})$$

The overall prior is the product of the three:

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

Thus Far

- Done: Description of problem
- Done: Description of example
- Done: Description of data
- To Do: List options available
- To Do: Implement Bayesian solution
- To Do: Compare and discuss

Estimation Options Available

- Asymptotic Equivalent of MLE
  Gallant and Tauchen (2001)

- Bayesian with Synthesized Likelihood
  (proposed here)

- Simulated Method of Moments
  Duffie and Singleton (1993)

- Other Criterion Functions
  Cramer von Mises – Gallant and Hong (2005)

Cites are to the most closely related papers. They are not attributions.

Estimation Options Discussed Here

- Bayesian with Synthesized Likelihood
  Proposed here

- Asymptotic Equivalent of MLE
  Gallant and Tauchen (2001)
  EMM (Efficient Method of Moments)

Implementation of Bayesian Method: Notation

The transition density implied by the scientific model is

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

where $x = (y_{t-L},\ldots,y_{t-1})$.

Priors are $\pi_1(\theta)$, $\pi_1(\theta,\psi)$, $\pi_2(\theta,\psi)$, where computation of

$$\psi = \Psi[p(\cdot|\cdot,\theta)]$$

requires a simulation from $p(y_t|x_{t-1},\theta)$.

A simulation from the model is denoted as $\{\hat{y}_t\}_{t=1}^{N}$ where N is on the order of 50,000.

Data is denoted as $\{\tilde{y}_t\}_{t=1}^{n}$.

Computational Strategy

1. Find a statistical model

$$f(y_t|x_{t-1}, \eta)$$

that fits the observed data $\{\tilde{y}_t\}_{t=1}^n$ well.

2. Find the implied map $g : \theta \mapsto \eta$ that satisfies

$$p(y|x, \theta) = f[y|x, g(\theta)]$$

3. Use

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

as the likelihood.

Statistical Model

For the example we use an VAR(1)–GARCH(1,1) system with normal innovations as the statistical model

$$f(y_t|x_{t-1}, \eta).$$

The statistical model has sixteen parameters.

Implied Map

We use the map $g(\cdot)$ defined by

$$g : \theta \mapsto \underset{\eta}{\text{argmax}} \iint \log f(y|x,\eta)\, dp(y|x,\theta)\, dp(x|\theta)$$

where (recall) $p(y|x,\theta)$ is the transition density implied by the scientific model and $p(x|\theta)$ is its corresponding stationary density.

Computing the Implied Map

Use the map $g(\cdot)$ defined by

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

where $\{y_t\}_{t=1}^{N}$ is a simulation from the scientific model with parameters set to θ and $\hat{x}_{t-1} = (\hat{y}_{t-L}, \ldots, \hat{y}_{t-1})$.

Compute the maximizing value $\hat{\eta}$ using a Metropolis-Hastings chain ...

Metropolis-Hastings for η

Proposal density: $T(\eta_{here}, \eta_{there})$

Proposal: η_{prop} drawn from $T(\eta_{old}, \eta)$

Likelihood: $\mathcal{L}(\eta) = \prod_{t=1}^{N} f(\widehat{y}_t | \widehat{x}_{t-1}, \eta)$

Put η_{new} to η_{prop} with probability

$$\alpha = \min \left[1, \frac{\mathcal{L}(\eta_{prop}) T(\eta_{prop}, \eta_{old})}{\mathcal{L}(\eta_{old}) T(\eta_{old}, \eta_{prop})} \right]$$

else put η_{new} to η_{old}.

This can be interpreted as simulated annealing where the temperature parameter is the size N of the simulation from the scientific model $p(y|x, \theta)$.

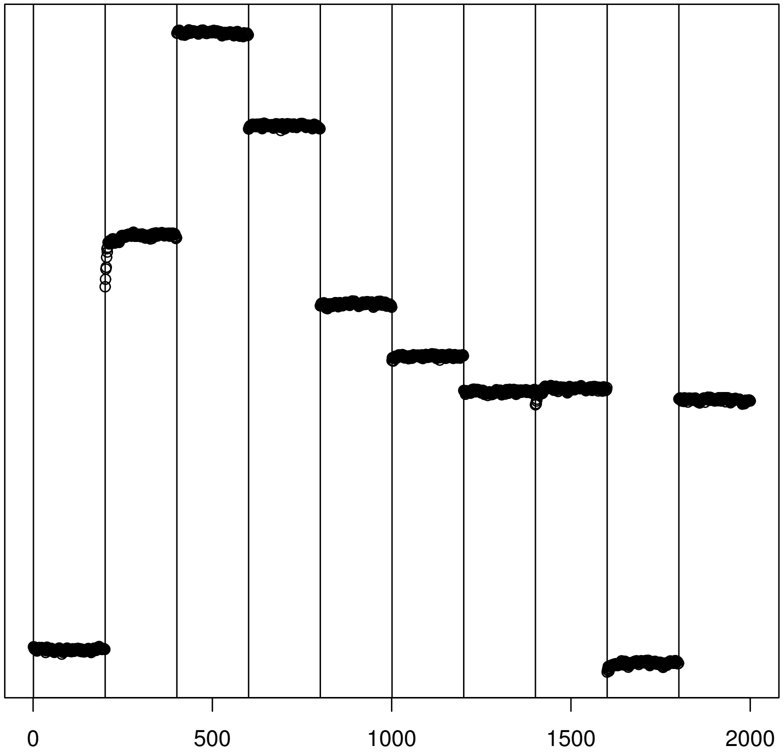
Results

The plot that follows exhibit the performance of the algorithm to compute the implied map

$$g : \theta \mapsto \eta$$

for a sequence of different θ.

Fig 4. $\mathcal{L}(\eta)$



For a sequence of MCMC proposals for θ, $\{\widehat{y}_t\}$ is generated and the η chain iterated. Plotted is $\mathcal{L}(\eta)$ evaluated at the iterates. Vertical bars mark where θ changes. The θ chain may or may not move, depending on $\pi(\theta, \psi)\mathcal{L}(\theta)$. Jumps are because $\{\widehat{y}_t\}$ changes at each vertical bar.

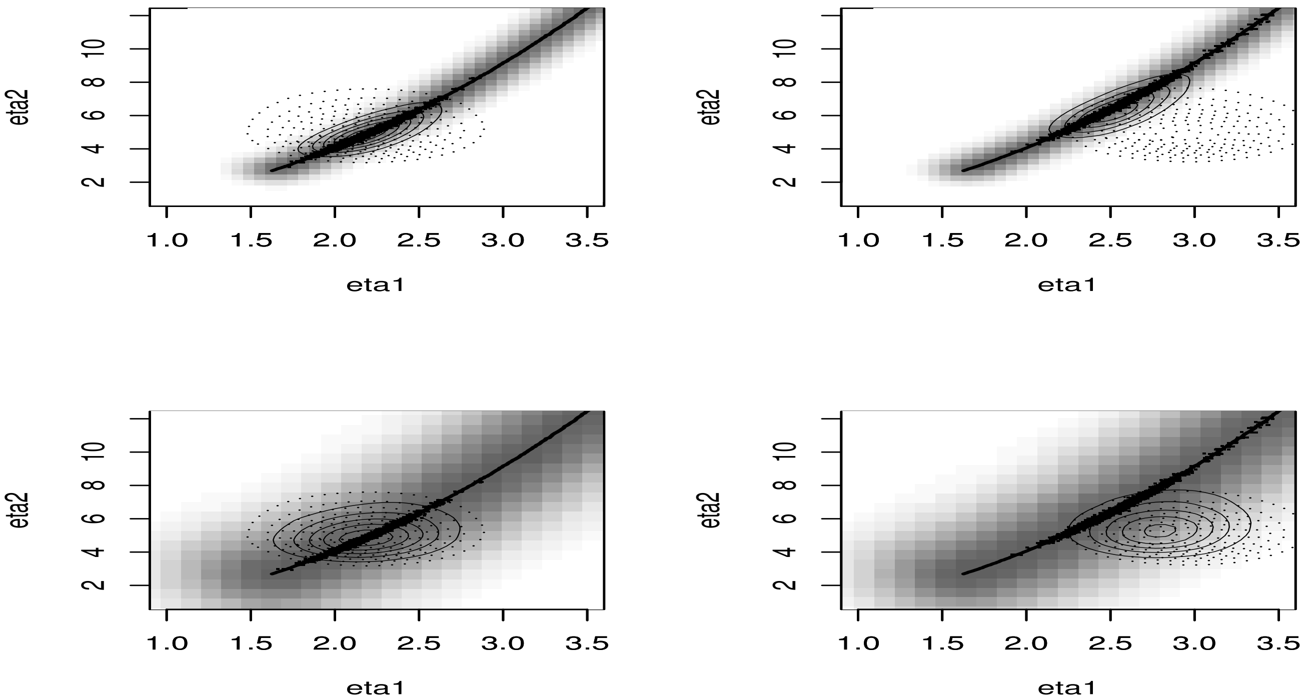
The Manifold

- We can now compute the implied map $\eta = g(\theta)$.

- Therefore we can now compute the manifold that was shown in Figure 1.

- Figure 1 is repeated next slide.

Fig 1. The Scientific and Statistical Model



The the solid ellipses are contours of the posterior of the statistical model $f(y|x,\eta)$; the dashed are for the likelihood. The line is the prior on $\pi(\eta)$ determined jointly by the implied map $\eta = g(\theta)$ from the parameters θ of scientific model to the parameters of the statistical model and by the prior on θ. The thickness of the line is proportional to the posterior of η. In the upper panels the scientific model is true; in the lower it is false. The adequacy of the scientific model can be assessed by relaxing $\pi(\eta)$ to see how functionals of $f(\cdot|\cdot,\eta)$ that relate to the scientific inquiry change. The shaded area indicates such a relaxation. The right panels are more relaxed than the left. In envisaged applications the statistical model is too richly parameterized to be estimable without imposition of the relaxed prior.

The Posterior

- Our next task is to compute the posterior shown as the thickening of the line in Figure 1.

- We first find the posterior in the seven dimensional space of the parameters of the scientific model $p(y|x,\theta)$ using the the implied map $\eta = g(\theta)$ and the statistical model to synthesize a likelihood, vis. $\mathcal{L}(g(\theta)) = \prod_{t=1}^{n} f(\tilde{y}_t|\tilde{x}_{t-1}, g(\theta))$.

- Once the posterior on θ has been determined, the posterior in the sixteen dimensional space of the parameters of the statistical model $f(y|x,\eta)$ can be found via the map $\eta = g(\theta)$.

Considerations

If θ is only moved slightly between computations, then the last computed value $\hat{\eta}$ will be a good start for the MCMC chain that computes the next $\eta = g(\theta)$.

If θ is only moved slightly between computations, then the last computed value $\hat{a} = (\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_K)$ of the asset pricing function

$$\mathcal{V}(\cdot) = \alpha_0 + \alpha_1 \phi_1(\cdot) + \alpha_2 \phi_2(\cdot) + \ldots + \alpha_K \phi_K(\cdot)$$

will make a good start for the nonlinear equation solver that computes the next.

These considerations argue for small moves in the MCMC chain used to get the posterior distribution of θ.

Implementing the Posterior Computation

Because we have the map $\eta = g(\theta)$, a synthesized likelihood

$$\mathcal{L}(g(\theta)) = \prod_{t=1}^{n} f(\tilde{y}_t | \tilde{x}_{t-1}, g(\theta))$$

for the scientific model is now available.

We can proceed to devise an MCMC chain that takes into account that:

- Small, local moves are required to preserve the stability of the algorithm for $\mathcal{L}(\eta)$.

- $\mathcal{L}(\eta)$ should only be computed if $\pi_1(\theta) = 1$ and is only available if $\pi_1(\theta, \psi) = 1$.

Additional Considerations

The parameters θ of the scientific model can be restricted to a grid and past values of $\mathcal{L}(g(\theta))$ saved so that once $\mathcal{L}(g(\theta))$ has been computed it never has to be computed again.

The grid increment should be scientifically meaningful. The number of grid points is irrelevant, only the increment matters.

An associative map can be used to store only those grid points that have been visited; this keeps memory usage manageable.

Implemented in this fashion, the MCMC chain for θ runs faster as it becomes longer.

Implemented in this fashion, errors in computing the likelihood are less harmful.

Proposal Density $T(\theta_{here}, \theta_{there})$

Allocate mass to grid points proportional to a normal density centered at θ_{here} with the mass assigned to θ_{here} put to zero.

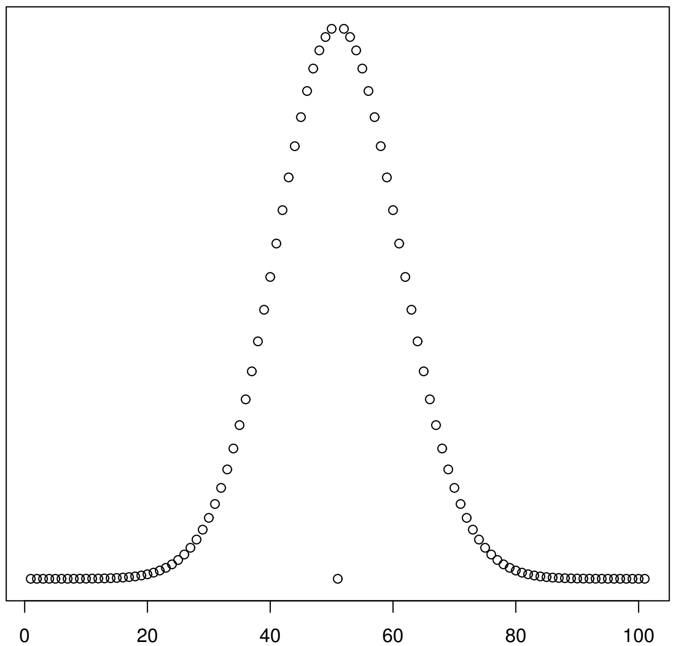
Normalize by summing the mass over the grid and dividing the mass at each grid point by the total.

If support conditions are built into the proposal, then when θ_{here} is near the boundary the density will be proportional to a right or left truncated normal and the normalization must be done twice at each move, because both

$$T(\theta_{old}, \theta_{prop}) \qquad \text{and} \qquad T(\theta_{prop}, \theta_{old})$$

need to be computed for Metropolis-Hastings.

Fig 6. Proposal Density



Mass proportional to a normal density. Becomes a truncated normal when the center nears the grid boundary. No mass at $x = 0$.

MCMC Implications of Prior Information

- Do not move if $\pi_1(\theta) = 0$.

- Do not move if $\pi_1(\theta, \psi) = 0$.

- Use $\pi_2(\theta, \psi)$ as the prior if a move is possible.

- Implied prior: $\pi(\theta, \psi) = \pi_1(\theta, \psi)\pi_1(\theta, \psi)\pi_2(\theta, \psi)$

Metropolis-Hastings

Proposal density: $T(\theta_{here}, \theta_{there})$

Proposal: θ_{prop} drawn from $T(\theta_{old}, \theta)$

Likelihood: $\mathcal{L}(g(\theta)) = \sum_{t=1}^{n} \log f(\tilde{y}_t | \tilde{x}_{t-1}, g(\theta))$

Put θ_{new} to θ_{prop} with probability

$$\alpha = \min\left[1, \frac{\pi(\theta_{prop}, \psi)\mathcal{L}(g(\theta_{prop}))T(\theta_{prop}, \theta_{old})}{\pi(\theta_{old}, \psi)\mathcal{L}(g(\theta_{old}))T(\theta_{old}, \theta_{prop})}\right]$$

else put θ_{new} to θ_{old}.

Results

The plots that follow exhibit the performance of the chain for the parameters θ of the habit model $p(y|x,\theta)$ using the synthesized likelihood

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

The prior information imposed is as follows:

Prior Information

$\pi_1(\theta)$: Reasonable bounds on all parameters to include positivity restrictions on positive valued parameters and non-explosive restrictions on autoregressive parameters.

$\pi_1(\theta, \psi)$: Existence of solution to Euler condition.

$\pi_2(\theta, \psi)$: Normal prior used by methods proposed here

$$P\left(\left|\mathcal{E}(r_f^a) - 0.89\%\right| < 1\%\right) = 0.95$$

$$P\left(|\rho - 0.2| < 0.1\right) = 0.95$$

$$P\left(|\phi - 0.9884| < 0.01\right) = 0.95$$

$\pi_2(\theta, \psi)$: Uniform prior by estimates compared with

$$P\left(\left|\mathcal{E}(r_f^a) - 0.89\%\right| < 0.5\%\right) = 1.00$$

Fig 7. Iterates 500 to 10,000 by 8.

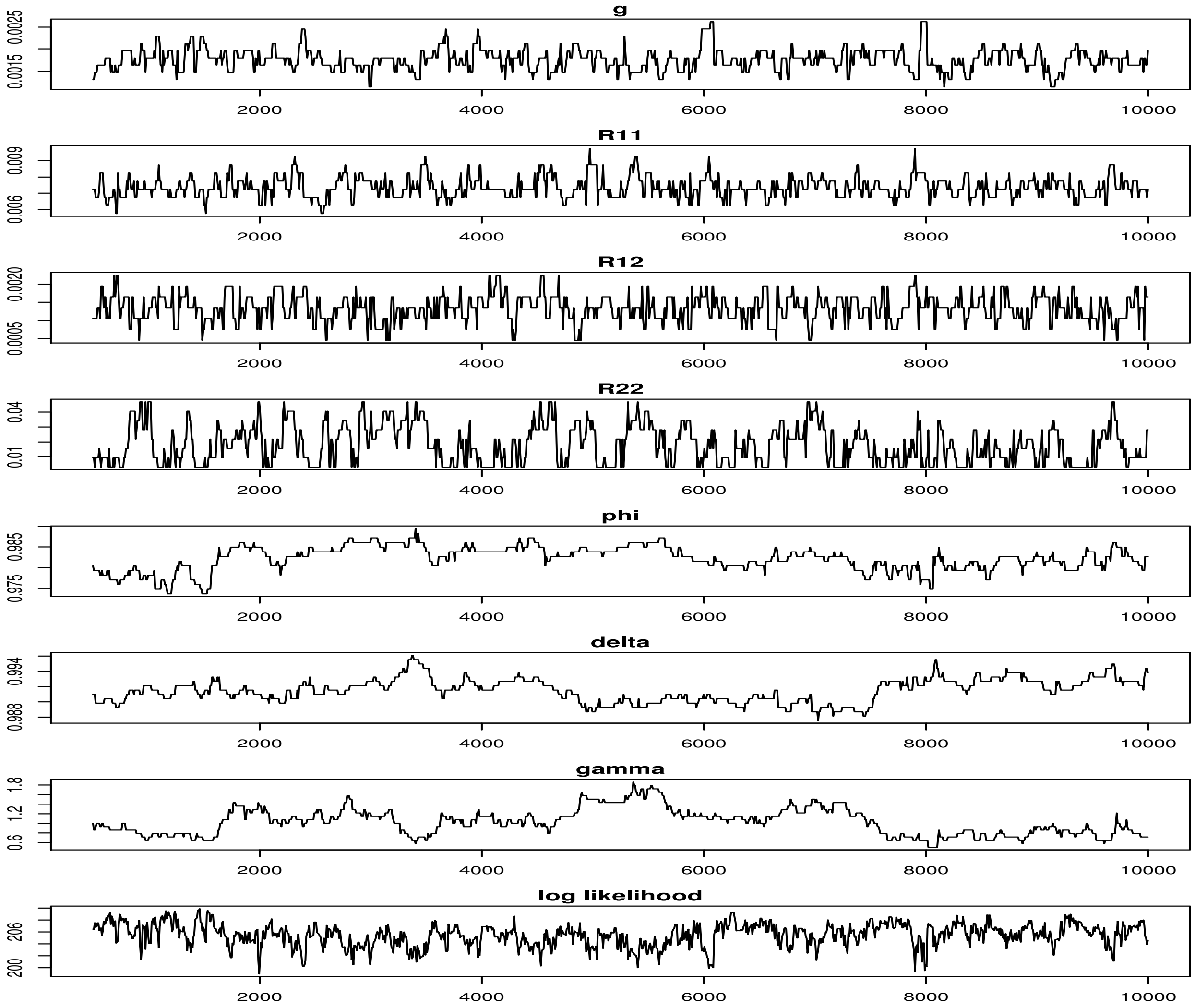


Fig 8. Correlations, Iterates 1 to 800,000 by 800

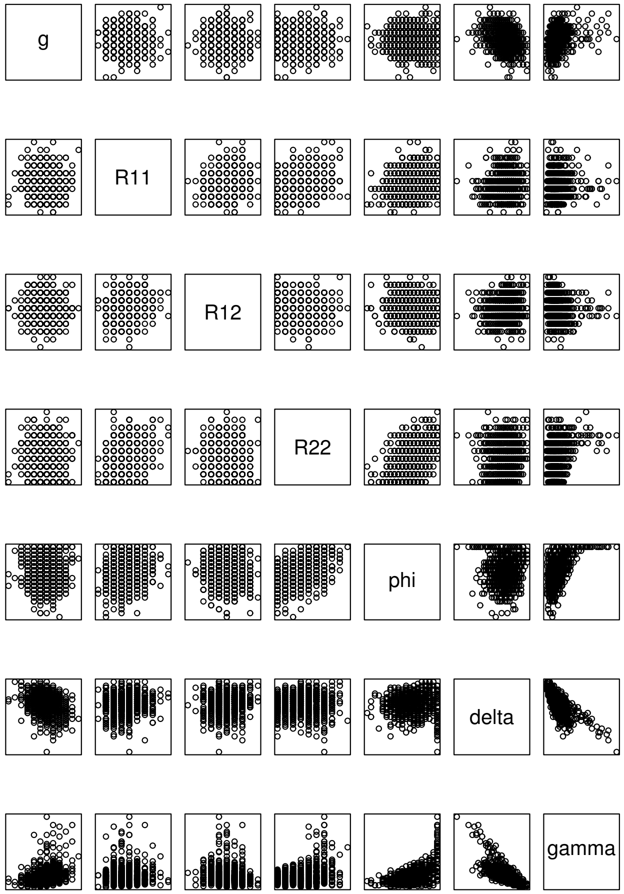


Fig 9. Autocorrelations, Iterates 1 to 800,000 by 8

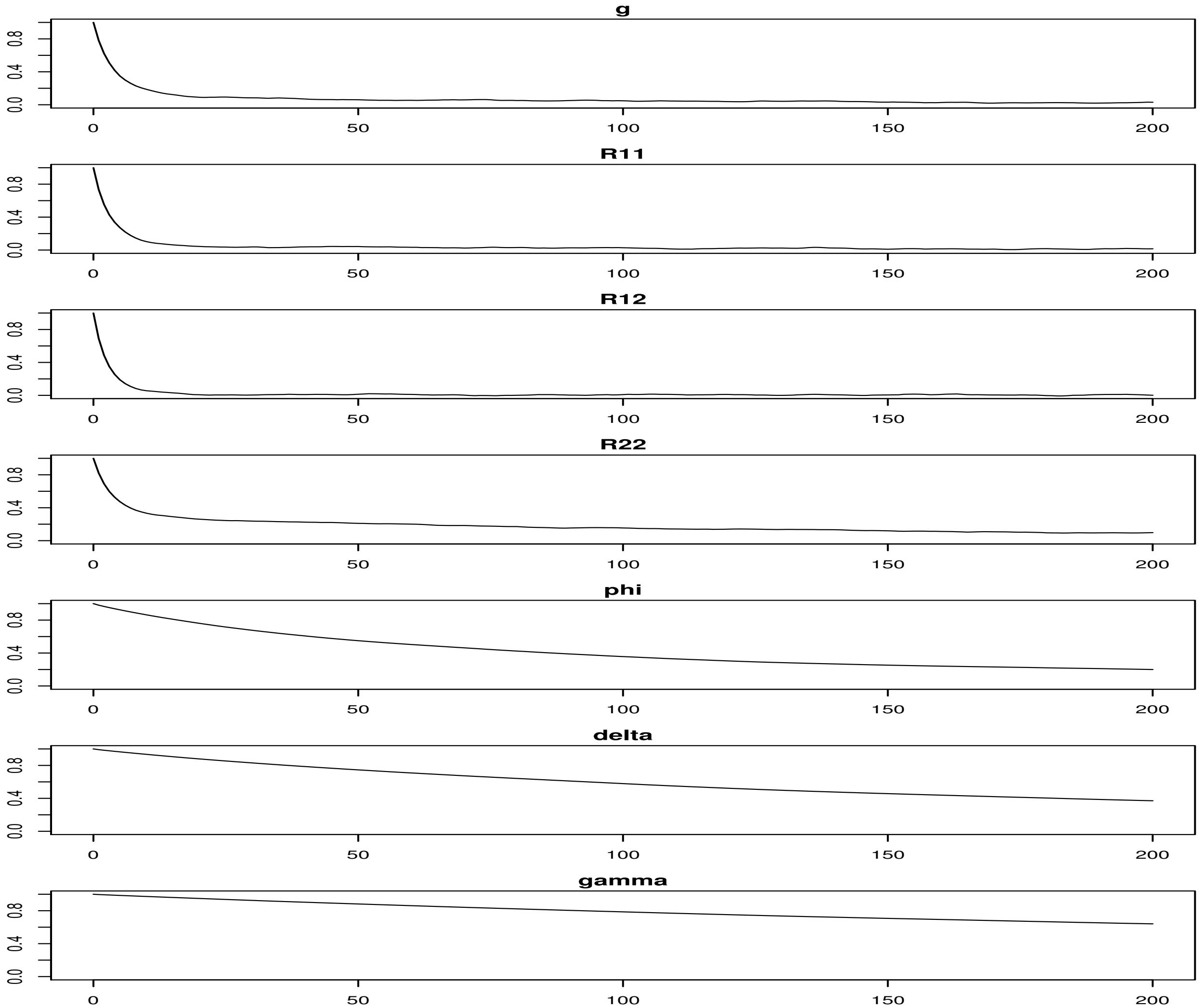


Fig 10. Posterior, Iterates 1 to 800,000 by 8

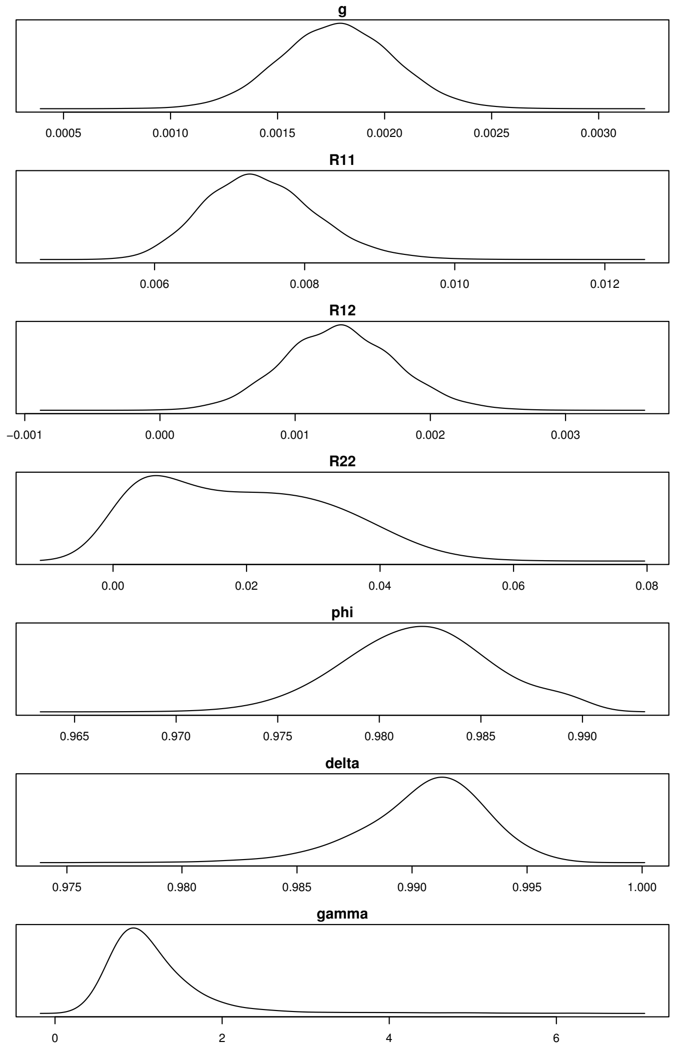
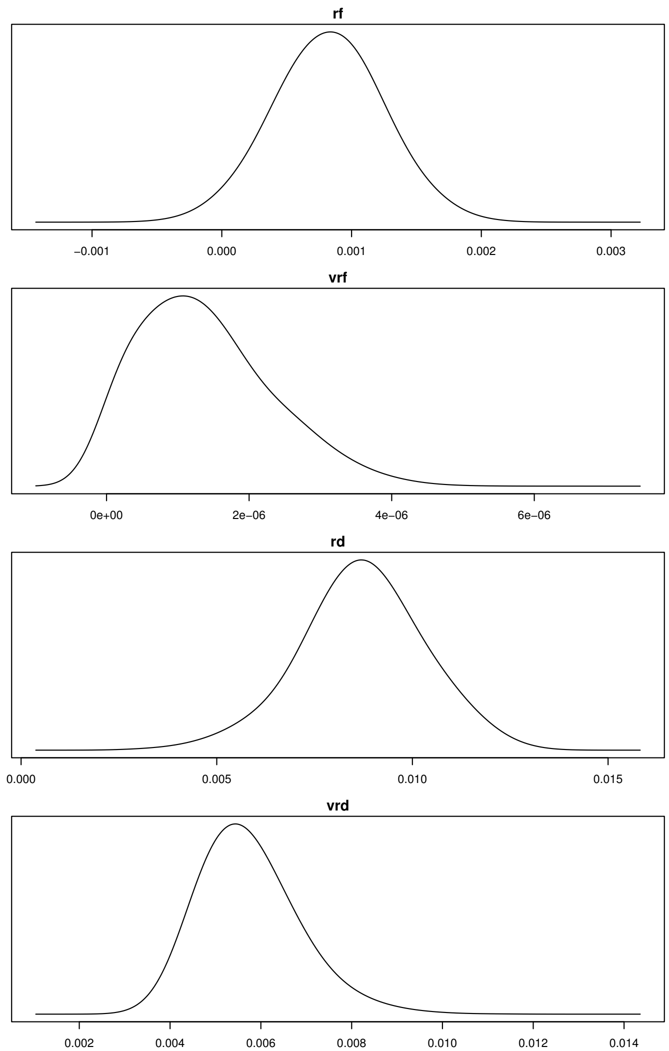


Fig 11. Posterior, Iterates 1 to 800,000 by 8



Monthly returns: 0.8% monthly = 11% annual

Results Will Be Compared to EMM Estimates

EMM Heuristics: For any QMLE estimator

$$\tilde{\eta}_n = \underset{\eta}{\text{argmax}} \frac{1}{n} \sum_{t=1}^{n} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \eta),$$

a sample average satisfies

$$0 = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \eta} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \tilde{\eta}_n)$$

because these are the first order conditions of the optimization problem.

Therefore a large simulation from a putative DGP $p(y_t | x_{t-1}, \theta)$ will satisfy

$$0 = m(\theta, \tilde{\eta}_n) = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial}{\partial \eta} \log f(\hat{y}_t | \hat{x}_{t-1}, \tilde{\eta}_n),$$

except for sampling variation in $\tilde{\eta}_n$. The equality holds exactly in the limit as n and N tend to infinity.

The EMM estimator attempts to find θ that solves these estimating equations as nearly as possible:

$$\hat{\theta}_n = \underset{\theta}{\text{argmin}}\, m^{'}(\theta, \tilde{\eta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\theta, \tilde{\eta}_n)$$

Table 1. Parameter Estimates (Monthly Frequency)

	EMM Estimates		Bayes-GARCH		
Parameter	Estimate	Std. Err.	Mode	Mean	Std. Dev.
g	0.002116	0.000250	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}			0.007254	0.007417	0.001903
r_{12}			0.001350	0.001336	0.001068
r_{22}			0.003125	0.018852	0.034435
ϕ	0.9853	0.0026	0.9804	0.9818	0.0095
δ	0.9939	0.0005	0.9898	0.9907	0.0070
γ	0.8386	0.2462	1.0744	1.1747	1.7638
	$\chi^2(5) = 14.476$ (0.0129)		$R = 800,000$		

Note: EMM uses data on the price dividend ratio and the consumption dividend ratio in addition to consumption growth and stock returns and imposes cointegration among consumption, dividends, and price. Variance parameters relate as

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

where μ_{dc}, ρ_s, and ψ_{22} are the location, autoregressive, and scale prameters of the cointegration relation between c_t and d_t.

Table 2. Parameter Estimates (Annual Frequency)

	Data	EMM Estimates		Bayes-GARCH		
Parameter	Estimate	Estimate	Std Err	Mode	Mean	Std Dev
g		2.539	0.0866	2.164	2.136	0.2369
σ		2.1308		2.5589	2.6106	0.2513
ρ		0.1650		0.1830	0.1773	0.0507
σ_w		12.9118		1.0825	6.5306	4.4984
ϕ		0.8373	0.0090	0.7890	0.8023	0.0328
δ		0.9292	0.0018	0.8845	0.8934	0.0244
γ		0.8386	0.2462	1.0744	1.1747	1.7638
r_{dt}^a	6.02	6.54		11.14	10.45	0.5487
$SD(r_{dt}^a)$	19.29	16.9		24.22	26.14	2.4735
r_{ft}^a		1.07		1.21	0.99	0.1451
$SD(r_{ft}^a)$		3.23		0.42	0.38	0.1489

Note: The mode of $\theta = (g, \sigma, \rho, \sigma_w, \delta, \gamma)$ is the mode of the multivariate posterior, not the mode of the marginal posteriors, and modal returns are those that correspond to that mode.

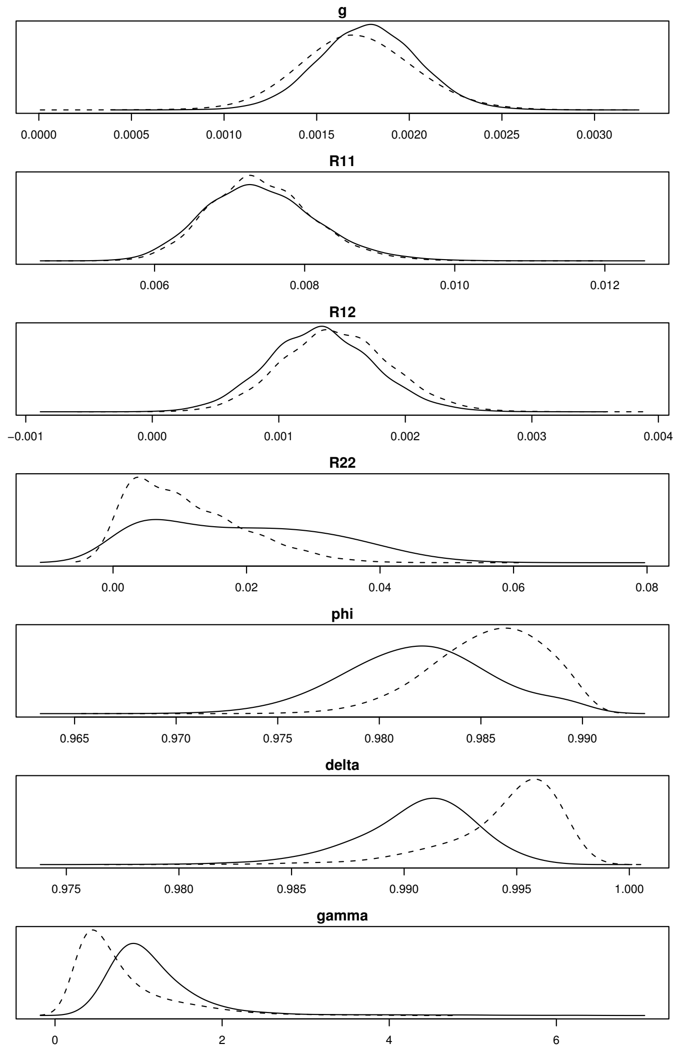
GARCH Implications

The Bayes estimates imply a huge risk premium whereas the EMM estimates imply a risk premium close to the data.

This is due to EMM's ignoring GARCH.

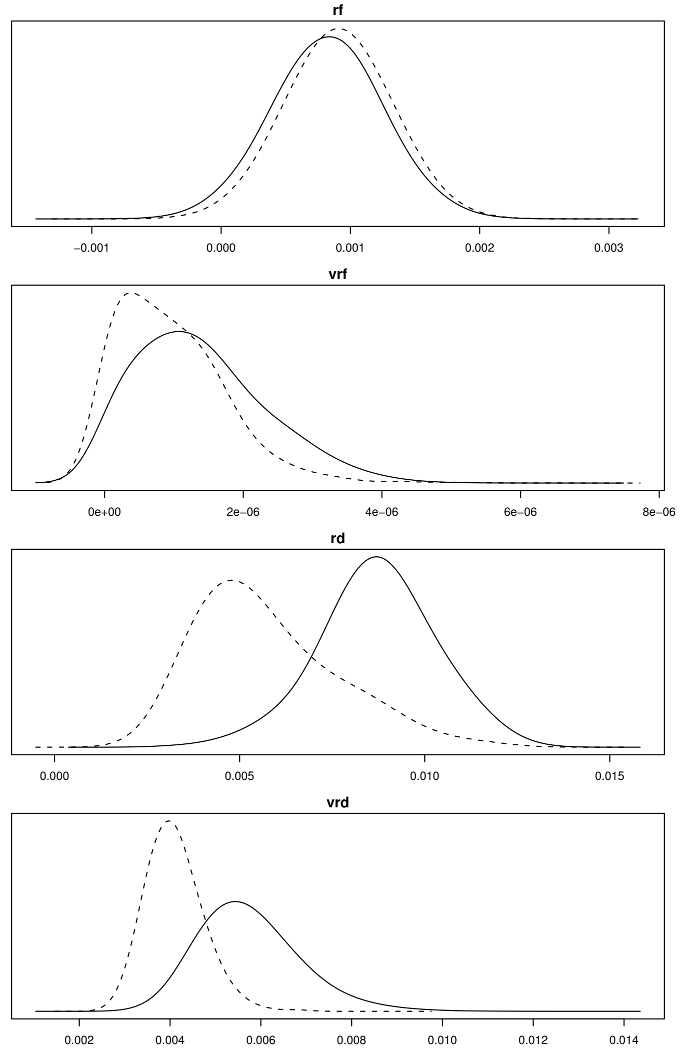
We'll show this by repeating the Bayesian analysis with a sharp prior that suppresses GARCH.

Fig 12. Posterior, Iterates 1 to 800,000 by 8



Solid line are GARCH; dashed lines are VAR.

Fig 13. Posterior, Iterates 1 to 800,000 by 8



Solid line are GARCH; dashed lines are VAR.

Table 3. Parameter Estimates (Monthly Frequency)

	EMM Estimates		Bayes-VAR		
Parameter	Estimate	Std. Err.	Mode	Mean	Std. Dev.
g	0.002116	0.000250	0.001639	0.001739	0.000258
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
r_{12}			0.001350	0.001451	0.000403
r_{22}			0.003125	0.008109	0.006025
ϕ	0.9853	0.0026	0.9861	0.9857	0.0024
δ	0.9939	0.0005	0.9955	0.9937	0.0030
γ	0.8386	0.2462	0.5726	0.9463	0.6179
	$\chi^2(5) = 14.476$ (0.0129)		$R = 800,000$		

Note: EMM uses data on the price dividend ratio and the consumption dividend ratio in addition to consumption growth and stock returns and imposes cointegration among consumption, dividends, and price. Variance parameters relate as

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

where μ_{dc}, ρ_s, and ψ_{22} are the location, autoregressive, and scale parameters of the cointegration relation between c_t and d_t.

Table 4. Parameter Estimates (Annual Frequency)

	Data	EMM-Estimates		Bayes-VAR		
Parameter	Estimate	Estimate	Std Err	Mode	Mean	Std Dev
g		2.539	0.0866	1.9672	2.087	0.0895
σ		2.1308		2.3857	2.5870	0.2202
ρ		0.1650		0.1960	0.1943	0.0508
σ_w		12.9118		1.0825	2.8090	2.1496
ϕ		0.8372	0.0090	0.8450	0.8412	0.0084
δ		0.9292	0.0018	0.9477	0.9269	0.0102
γ		0.8386	0.2462	0.5726	0.9463	0.6179
r^a_{dt}	6.02	6.54		6.00	7.58	0.6930
$SD(r^a_{dt})$	19.29	16.9		20.49	21.48	1.5343
r^a_{ft}		1.07		1.20	1.16	0.1389
$SD(r^a_{ft})$		3.23		0.33	0.27	0.1388

Note: The mode of $\theta = (g, \sigma, \rho, \sigma_w, \delta, \gamma)$ is the mode of the multivariate posterior, not the mode of the marginal posteriors, and modal returns are those that correspond to that mode.

Table 5. Parameter Estimates (Annual Frequency)

	Data	Bayes-VAR			Bayes-GARCH		
Param.	Est.	Mode	Mean	Std. Dev.	Mode	Mean	Std. Dev.
g		1.9672	2.087	0.0895	2.164	2.136	0.2369
σ		2.3857	2.5870	0.2202	2.5589	2.6106	0.2513
ρ		0.1960	0.1943	0.0508	0.1830	0.1773	0.0507
σ_w		1.0825	2.8090	2.1496	1.0825	6.5306	4.4984
ϕ		0.8450	0.8412	0.0084	0.7890	0.8023	0.0328
δ		0.9477	0.9269	0.0102	0.8845	0.8934	0.0244
γ		0.5726	0.9463	0.6179	1.0744	1.1747	1.7638
r^a_{dt}	6.02	6.00	7.58	0.6930	11.14	10.45	0.5487
$SD(r^a_{dt})$	19.29	20.49	21.48	1.5343	24.22	26.14	2.4735
r^a_{ft}		1.20	1.16	0.1389	1.21	0.99	0.1451
$SD(r^a_{ft})$		0.33	0.27	0.1388	0.42	0.38	0.1489

Note: The mode of $\theta = (g, \sigma, \rho, \sigma_w, \delta, \gamma)$ is the mode of the multivariate posterior, not the mode of the marginal posteriors, and modal returns are those that correspond to that mode.

Preference Parameters Control GARCH

The estimation results suggest that the preference parameters ϕ, δ and γ control GARCH. They are the ones that had to move to put GARCH effects in the fit.

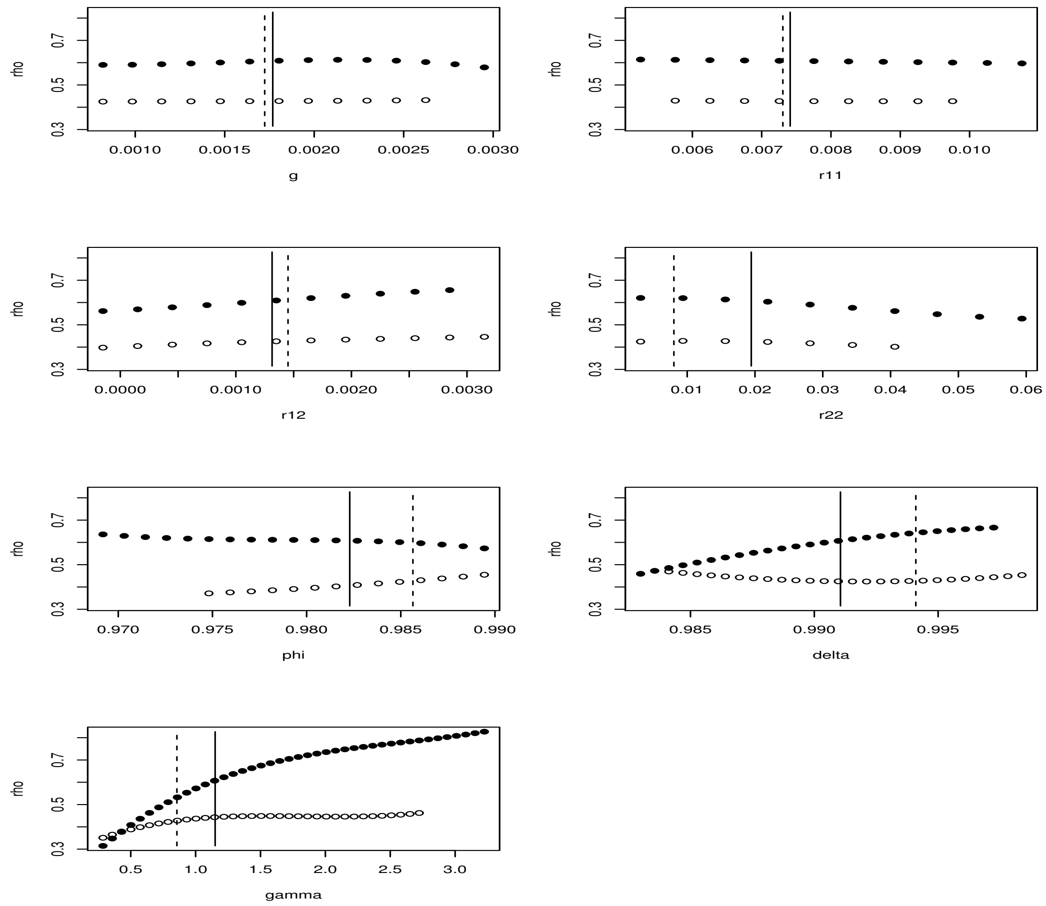
This can be seen visually by looking at plots of the map

$$g : \theta \to \underset{\eta}{\text{argmax}} \iint \log f(y|x,\eta)\, dp(y|x,\theta)\, dp(x|\theta)$$

where (recall) $p(y|x,\theta)$ is the transition density implied by the scientific model and $p(x|\theta)$ is its corresponding stationary density.

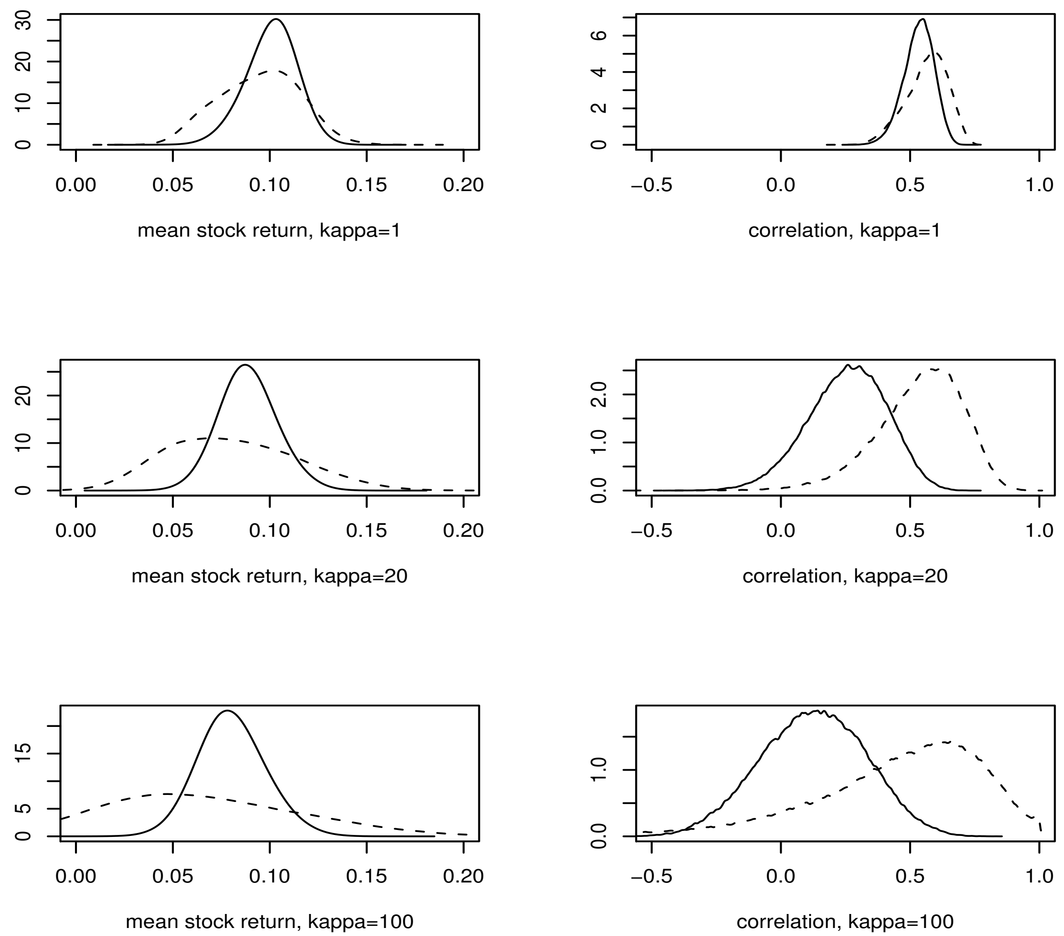
But the most dramatic visualization is a plot of the conditional correlation for the year 2002 as determined by the map, which follows.

Fig 14. Conditional Correlation



Conditional correlation between annual consumption growth and stock returns for 2002 plotted against scientific model parameters $\theta = (g, r_{11}, r_{21}, r_{22}, \phi, \delta, \gamma)$. Dots are for the GARCH statistical model and circles for the VAR. The solid vertical line is the posterior mean for θ for GARCH and the dashed line is the same for VAR.

Fig 15. Determination of Model Adequacy



The statistical model is a VAR-GARCH on annual consumption growth and stock returns for 1929–2001. The scientific model is a habit persistence consumption based asset pricing model. The habit model forces unrealistically high implied stock returns over the period 2002–2102 and an unrealistically high conditional correlation for the year 2002.

Table 6. 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

Things Change

- Previously frequentist statistical methods could solve problems that Bayesian statistical methods could not.

- Now the converse is true.

- Especially when data are sparse.

- Software: `ftp://ftp.econ.duke.edu/pub/arg/gsm`