Bayesian Estimation of State Space Models Using Moment Conditions

A. Ronald Gallant
Penn State University

Raffaella Giacomini
University College London

Giuseppe Ragusa
Luiss University

Paper: http://www.aronaldg.org/papers/bliml.pdf
Slides: http://www.aronaldg.org/papers/blimlclr.pdf

The State Space Model

- Observed variables: $y = (y_1, y_2, ..., y_T)$
- Latent variables: $x = (x_1, x_2, ..., x_T)$
- Parameters θ
- Known transition density $p^o(x_{t+1} \,|\, x_t, \theta)$
- Prior $p^o(\theta)$
- Do not know the measurement density $p^o(y_{t+1} \,|\, x_{t+1}, \theta)$
- But do have moment conditions $\mathcal{E}\Big[g(y_{t+1}, x_{t+1}, \theta) \,|\, \mathcal{I}_t\Big] = 0$

Assumptions, 1 of 2

- Sample moment conditions

$$g_T(y, x, \theta) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(y_t, x_t, \theta)$$

- Weighting matrix (May have to use a HAC weighting matrix instead.)

$$\begin{aligned} \Sigma(y, x, \theta) &= \frac{1}{T} \sum_{t=1}^{T} \tilde{g}(y_t, x_t, \theta)' \tilde{g}(y_t, x_t, \theta) \\ \tilde{g}(y_t, x_t, \theta) &= g(y_t, x_t, \theta) - \frac{1}{\sqrt{T}} g_T(y, x, \theta) \end{aligned}$$

- Assert

$$Z = [\Sigma(y, x, \theta^o)]^{-1/2} g_T(y, x, \theta^o) \sim \Psi(z)$$

  – $\Psi(z)$ usually $N(z \,|\, 0, I)$ a la Chernozukov and Hong (2003)

Assumptions, 2 of 2

- Semi-pivotal condition: Let $\mathcal{Y} \times \mathcal{X}$ denote the support of (y, x), Θ the support of θ, $\mathcal{Z}$ the support of z, and

$$C^{(\theta,z)} = \{(y, x) \in \mathcal{Y} \times \mathcal{X} \,:\, Z(y, x, \theta) = z\}.$$

  We assume that $C^{(\theta,z)}$ is not empty for any $(\theta, z) \in \Theta \times \mathcal{Z}$.

- Sufficient is that each element of g_T is continuous with respect to at least one continuous element of (y, x), is neither bounded from above nor below as that continuous variable varies over its support, and that the residuals used to compute the weighting matrix are centered as above.

The Enabling Result (Gallant, 2015a&b, JFEC)

The probability space

$$(\mathcal{Y} \times \mathcal{X} \times \Theta, \mathcal{C}^o, P^o)$$

with density

$$p^o(y, x, \theta) = p^o(y \,|\, x, \theta) p^o(x \,|\, \theta) p^o(\theta),$$

where $\mathcal{C}^o$ the Borel subsets of $\mathcal{Y} \times \mathcal{X} \times \Theta$, can be replaced by the probability space

$$(\mathcal{Y} \times \mathcal{X} \times \Theta, \mathcal{C}^*, P^*)$$

with density

$$\begin{aligned} p^*(y \,|\, x, \theta) &= \text{adj}(y, x, \theta) \psi[Z(y, x, \theta)] \\ p^*(y, x, \theta) &= p^*(y \,|\, x, \theta) p^o(x \,|\, \theta) p^o(\theta) \end{aligned}$$

for the purposes of Bayesian inference, where $\mathcal{C}^* \subset \mathcal{C}^o$, which implies a loss of information.

◇ Usually set $\text{adj}(y, x, \theta) = 1$ (Gallant, 2015b).

Estimation Strategy

- Sample $\left\{\theta^{(i)}, x^{(i)}\right\}$ from the density

$$p(\theta, x \mid y) \propto (2\pi)^{-M/2} \exp\left\{-\frac{1}{2} g_T(y, x, \theta)' \left[\Sigma(y, x, \theta)\right]^{-1} g_T(y, x, \theta)\right\} p^o(x \mid \theta)\, p^o(\theta)$$

- Particle Gibbs algorithm
  - Sample $\theta^{(i)}$ given $x^{(i-1)}$ and $\theta^{(i-1)}$ using Metropolis
    - last draw of MCMC chain of length K.
  - Sample $x^{(i)}$ given $\theta^{(i)}$ and $x^{(i-1)}$ using a conditional PF
    - last particle of a modified particle filter of size N.
  - Iterate back and forth. (Can view it as an approximate EM algorithm.)

- Estimate and scale are mean and standard deviation of $\left\{\theta^{(i)}\right\}$.

Next:
Two Examples

- A Dynamic Stochastic General Equilibrium Model
  - Description
  - Estimates

- A Stochastic Volatility Model
  - Description
  - Estimates

A DSGE Model – 1 of 4

From Del Negro and Schorfheide (2008) simplified to permit an analytic solution by removing rigidities, investment, etc.

Three shocks:

$z_t = \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}$	Factor productivity
$\phi_t = \rho_\phi \phi_{t-1} + \sigma_\phi \epsilon_{\phi,t}$	Consumption/leisure preference
$\lambda_t = \rho_\lambda \lambda_{t-1} + \sigma_\lambda \epsilon_{\lambda,t}$	Price elasticity of intermediate goods

Three outputs:

w_t	Wages
y_t	Output
π_t	Inflation

A DSGE Model – 2 of 4

First order conditions

$$
\begin{aligned}
0 &= y_t + \frac{1}{\beta}\pi_t - \mathcal{E}_t(y_{t+1} + \pi_{t+1} + z_{t+1}) \\
0 &= w_t + \lambda_t \\
0 &= w_t - (1+\nu)y_t - \phi_t
\end{aligned}
$$

where ν is a labor supply elasticity and β is the discount rate.

The true values of the parameters are

$$
\begin{aligned}
\theta &= (\rho_z, \rho_\phi, \rho_\lambda, \sigma_z, \sigma_\phi, \sigma_\lambda, \nu, \beta) \\
&= (0.15, 0.68, 0.56, 0.71, 2.93, 0.11, 0.96, 0.996)
\end{aligned}
$$

We take w_t, y_t, and π_t as measured and z_t and ϕ_t as latent.

A DSGE Model – 3 of 4

A set of conditions that identify the model are

$$
\begin{aligned}
g_1 &= (w_t - \rho_\lambda w_{t-1})^2 - \sigma_\lambda^2 \\
g_2 &= w_{t-1}(w_t - \rho_\lambda w_{t-1}) \\
g_3 &= [w_{t-1} - (1+\nu)y_{t-1}][w_t - (1+\nu)y_t - \rho_\phi(w_{t-1} - (1+\nu)y_{t-1})] \\
g_4 &= [w_{t-1} - (1+\nu)y_{t-1}](\phi_t - \rho_\phi \phi_{t-1}) \\
g_5 &= [w_t - (1+\nu)y_t]^2 - \sigma_\phi^2 \\
g_6 &= w_{t-1}(y_{t-1} + \frac{1}{\beta}\pi_{t-1} - y_t - \pi_t - \rho_z z_{t-1}) \\
g_7 &= y_{t-1}(y_{t-1} + \frac{1}{\beta}\pi_{t-1} - y_t - \pi_t - \rho_z z_{t-1}) \\
g_8 &= \pi_{t-1}(y_{t-1} + \frac{1}{\beta}\pi_{t-1} - y_t - \pi_t - \rho_z z_{t-1}) \\
g_9 &= (y_{t-1} + \frac{1}{\beta}\pi_{t-1} - y_t - \pi_t)^2 - \frac{\rho_z^2 \sigma_z^2}{1 - \rho_z^2}
\end{aligned}
$$

A DSGE Model – 4 of 4

- An analytic expression for the likelihood $\mathcal{L}(\theta) = p(y|\theta)$ is available for this model

- Analysis of the likelihood shows that only one of the four parameters $\sigma_z, \sigma_\phi, \nu, \beta$ can be identified

- Three will have to be calibrated in order to apply frequentist methods and either calibrated or specified by a tight prior to apply Bayesian methods

- We calibrate σ_z, σ_ϕ, ν and leave β as the free parameter.

Table 1. Parameter Estimates, DSGE Model

Parameter	True Value	Mean	Mode	Standard Dev.
		Particle Gibbs		
ρ_z	0.15	0.21887	0.23069	0.09179
ρ_ϕ	0.68	0.59967	0.60750	0.04988
ρ_λ	0.56	0.50884	0.31473	0.28981
σ_λ	0.11	0.10797	0.11613	0.06896
β	0.996	0.98201	0.99634	0.01834
		Maximum Likelihood		
ρ_z	0.15	0.15165	0.15087	0.00583
ρ_ϕ	0.68	0.59185	0.59419	0.05044
ρ_λ	0.56	0.56207	0.56549	0.05229
σ_λ	0.11	0.11225	0.11189	0.00508
β	0.996	0.99640	0.99643	0.00186

Data with $T = 250$ simulated at true values. $N = 1000$ particles; $K = 50$ Metropolis draws. GMM mean, mode, and standard deviation are from MCMC chains of length $R = 9637$ with stride of 1; for MLE chain $R = 500000$, stride is 5.

Figure 1. PF Estimate of x, DSGE Model

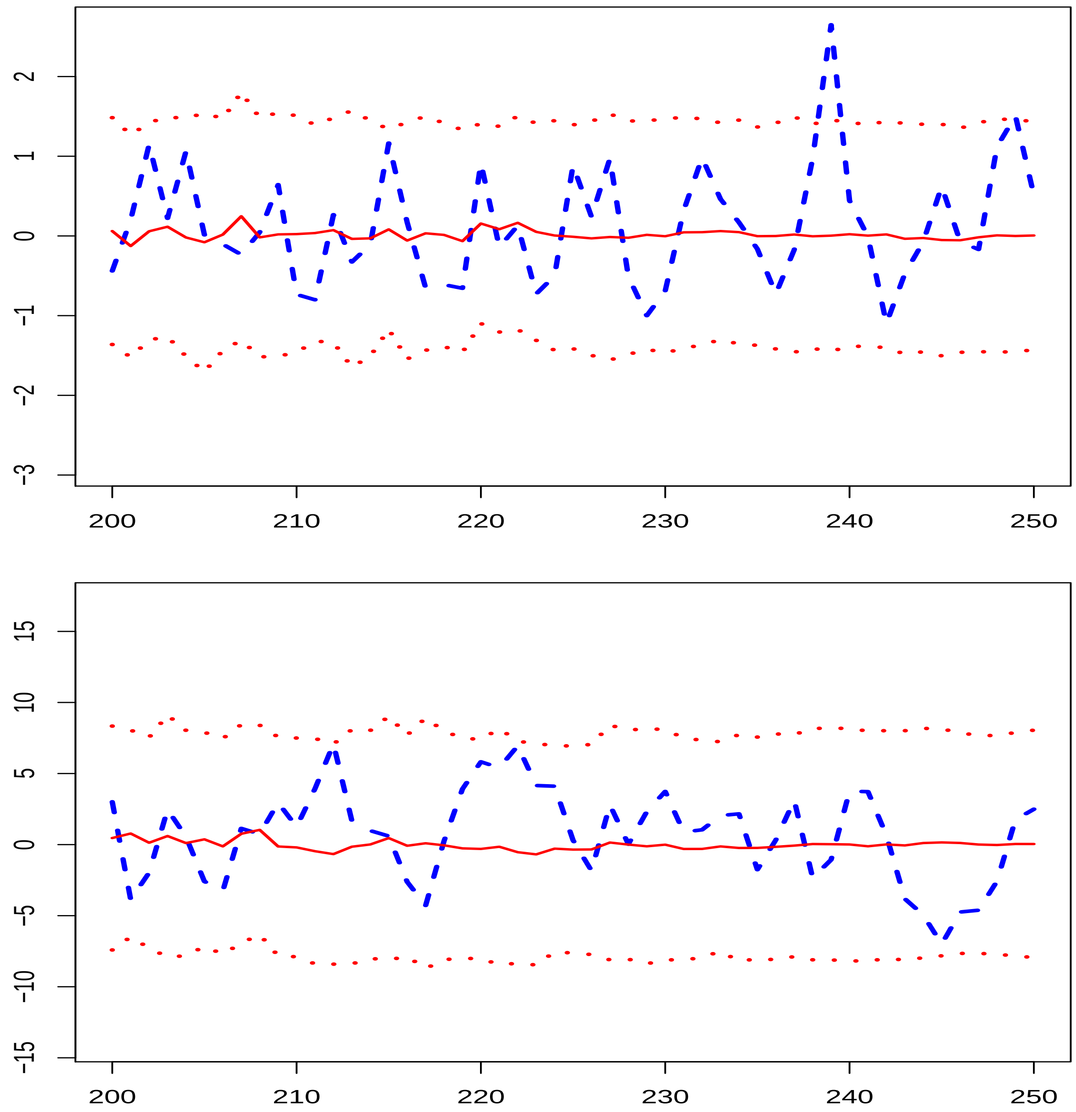
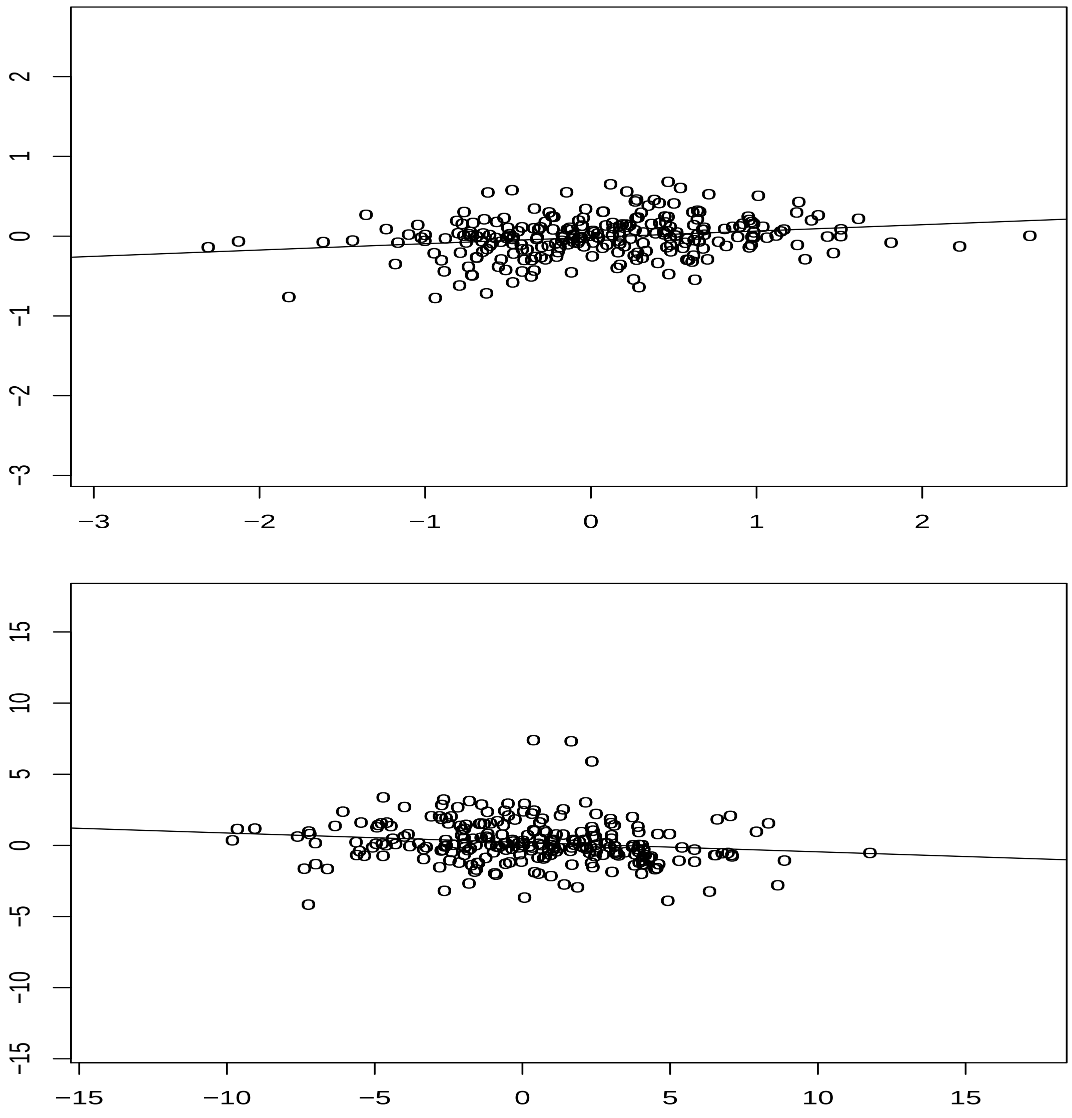


Figure 2. PF Estimate of x, DSGE Model



The Choice of Moments Does Matter 1 of 2

- It is possible to perform counter-factual (e.g. impulse-response) analysis using moment conditions alone.

- However, for it to work, one must do a much better job of estimating the history of the latent variables.

- To estimate latent variables, it is not necessary to identify model parameters.

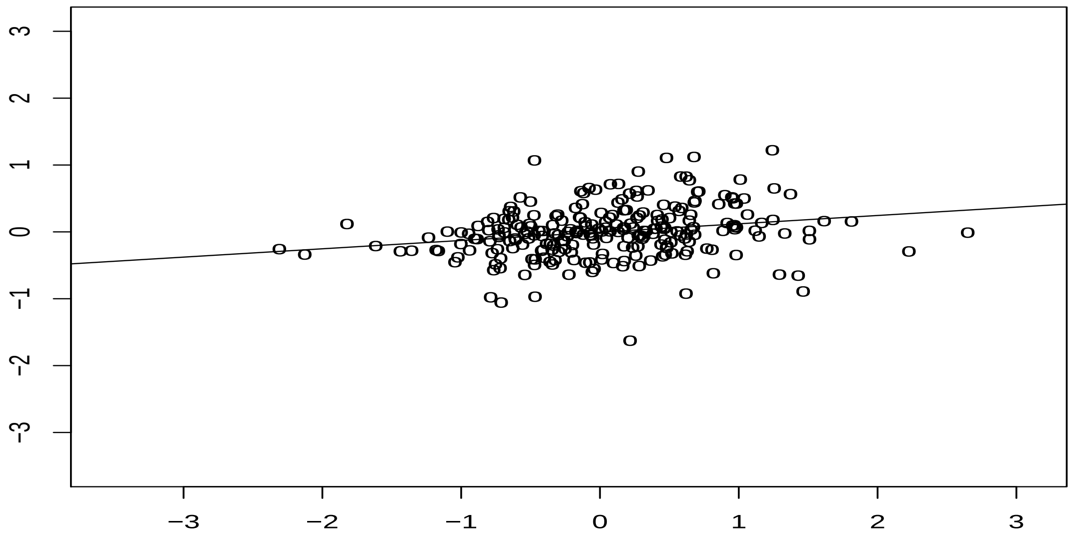
- Only the latent variables need to be identified.

The Choice of Moments Does Matter 2 of 2

Moment conditions for counter-factual analysis

$$
\begin{aligned}
h_1 &= y_{t-1} + \frac{1}{\beta}\pi_{t-1} - y_t - \pi_t - \rho_z z_{t-1} \\
h_2 &= w_{t-1}\, h_1 \\
h_3 &= y_{t-1}\, h_1 \\
h_4 &= \pi_{t-1}\, h_1 \\
h_5 &= w_t - (1+\nu) y_t - \phi_t \\
h_6 &= w_{t-1}\, h_5 \\
h_7 &= y_{t-1}\, h_5 \\
h_8 &= \pi_{t-1}\, h_5
\end{aligned}
$$

Figure 3. PF Estimate of x, DSGE Model



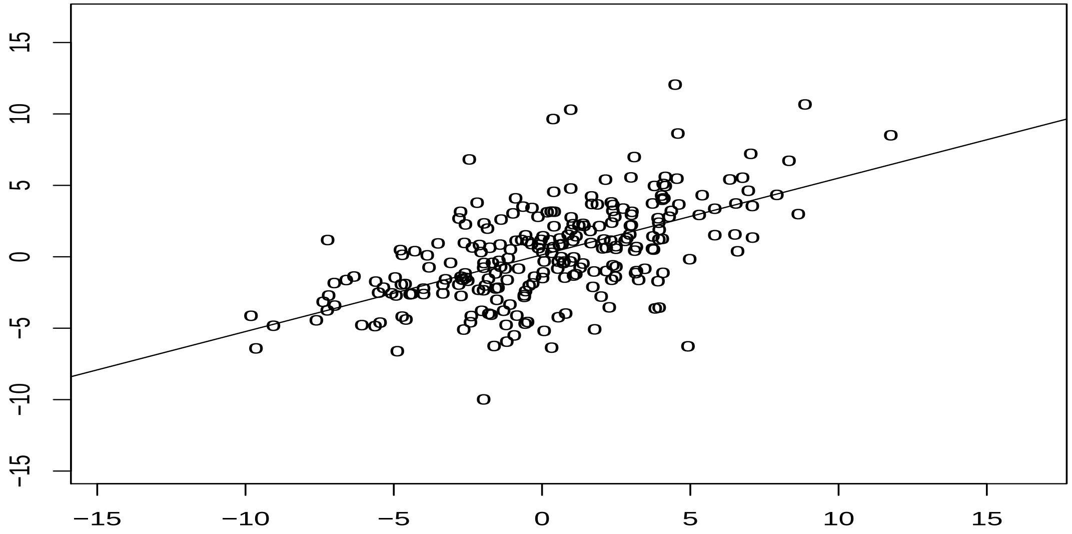
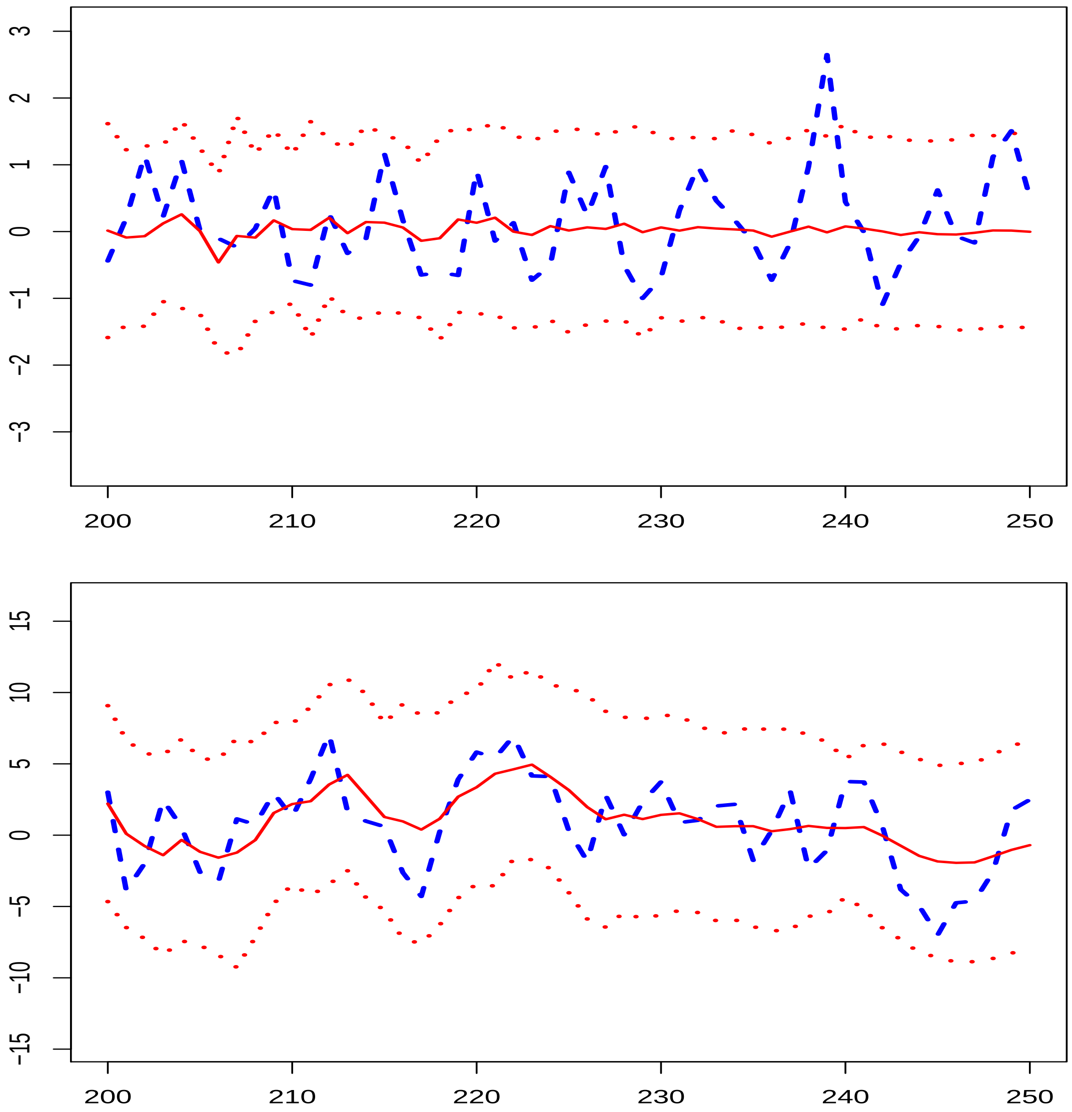


Figure 4. PF Estimate of x, DSGE Model



PF and Metropolis Moments Can Differ

If we use the moments that identify the model used for Table 1 for the Metropolis step and the moments designed for a counter-factual analysis used for Figures 3 through 4 for the PF step, we get slightly better results in the following Table 2.

Table 2. Alternative Parameter Estimates, DSGE Model

Parameter	True Value	Mean	Mode	Standard Dev.
		Particle Gibbs		
ρ_z	0.15	0.23508	0.15007	0.08975
ρ_ϕ	0.68	0.69870	0.58945	0.06127
ρ_λ	0.56	0.49904	0.46443	0.28418
σ_λ	0.11	0.11292	0.08924	0.06559
β	0.996	0.97465	0.99604	0.02479
		Maximum Likelihood		
ρ_z	0.15	0.15165	0.15087	0.00583
ρ_ϕ	0.68	0.59185	0.59419	0.05044
ρ_λ	0.56	0.56207	0.56549	0.05229
σ_λ	0.11	0.11225	0.11189	0.00508
β	0.996	0.99640	0.99643	0.00186

Data with $T = 250$ simulated at true values. $N = 1000$ particles; $K = 50$ Metropolis draws. GMM mean, mode, and standard deviation are from MCMC chains of length $R = 9637$ with stride of 1; for MLE chain $R = 500000$, stride is 5.

A Stochastic Volatility Model – 1 of 2

$$\begin{aligned} y_t &= \rho y_{t-1} + \exp(x_t)\, u_t & (1) \\ x_t &= \phi x_{t-1} + \sigma e_t & (2) \\ e_t &\sim N(0,1) & (3) \\ u_t &\sim N(0,1) & (4) \end{aligned}$$

The true values of the parameters are

$$\theta_0 = (\rho_0, \phi_0, \sigma_0) = (0.9, 0.9, 0.5) \quad \text{(plots)}$$

$$\theta_0 = (\rho_0, \phi_0, \sigma_0) = (0.25, 0.8, 0.1) \quad \text{(estimation)}$$

A Stochastic Volatility Model – 2 of 2

Moment Conditions

$$
\begin{aligned}
h_1 &= (y_t - \rho y_{t-1})^2 - [\exp(x_t)]^2 \\
h_2 &= |y_t - \rho y_{t-1}||y_{t-1} - \rho y_{t-2}| - \left(\frac{2}{\pi}\right)^2 \exp(x_t)\exp(x_{t-1}) \\
&\vdots \\
h_{L+1} &= |y_t - \rho y_{t-1}||y_{t-L} - \rho y_{t-L-1}| - \left(\frac{2}{\pi}\right)^2 \exp(x_t)\exp(x_{t-L}) \\
h_{L+2} &= y_{t-1}(y_t - \rho y_{t-1}) \\
h_{L+3} &= x_{t-1}(x_t - \phi x_{t-1}) \\
h_{L+4} &= (x_t - \phi x_{t-1})^2 - \sigma^2
\end{aligned}
$$

Table 3. Parameter Estimates, SV Model

Parameter	True Value	Mean	Mode	Standard Dev.
		Particle Gibbs		
ρ	0.25	0.30271	0.30939	0.076758
ϕ	0.8	0.15348	0.85765	0.643400
σ	0.1	0.11400	0.08435	0.070081
		Flury and Shephard Estimator		
ρ	0.25	0.30278	0.28555	0.059320
ϕ	0.8	0.17599	0.89189	0.509780
σ	0.1	0.09737	0.07839	0.064661

Data of length $T = 200$ was generated from the SV model at true values. In both panels the number of particles is $N = 1000$. For particle Gibbs, there are $K = 50$ Metropolis draws. For FS an unbiased estimate of the likelihood is computed from and unconditional PF and used for one Metropolis draw. The columns labeled mean, mode, and standard deviation are the mean, mode, and standard deviations of an MCMC chain of length 200000.

Figure 5. PF Estimate of x, SV Model

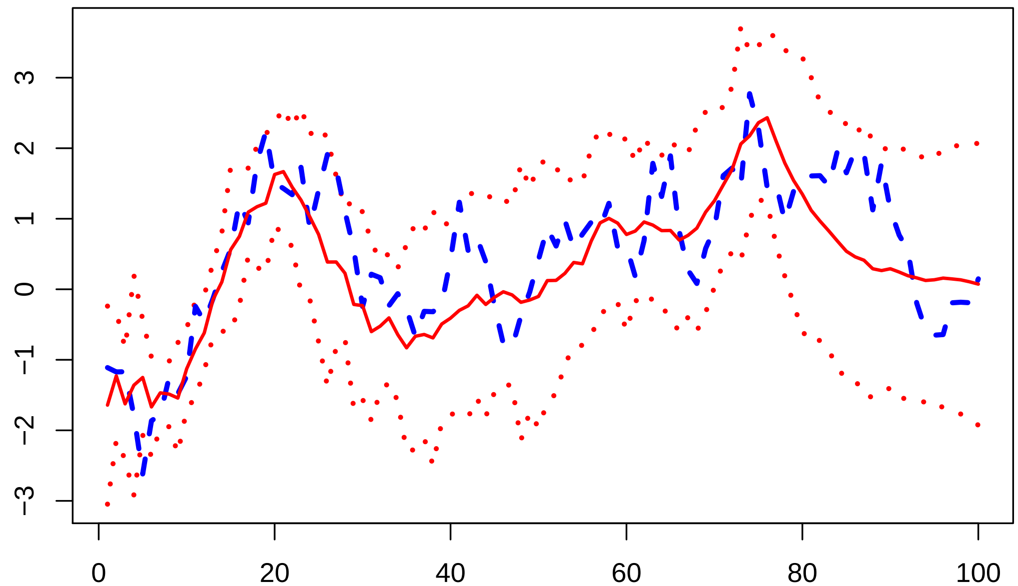


Figure 6. Flury-Shephard Estimate of x, SV Model

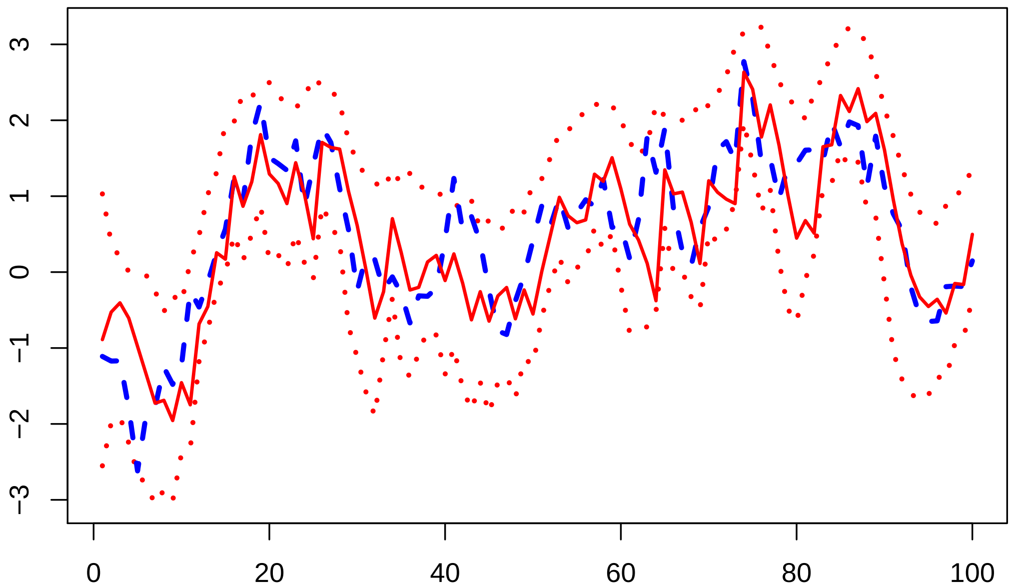


Figure 7. PF Estimate of x, SV Model

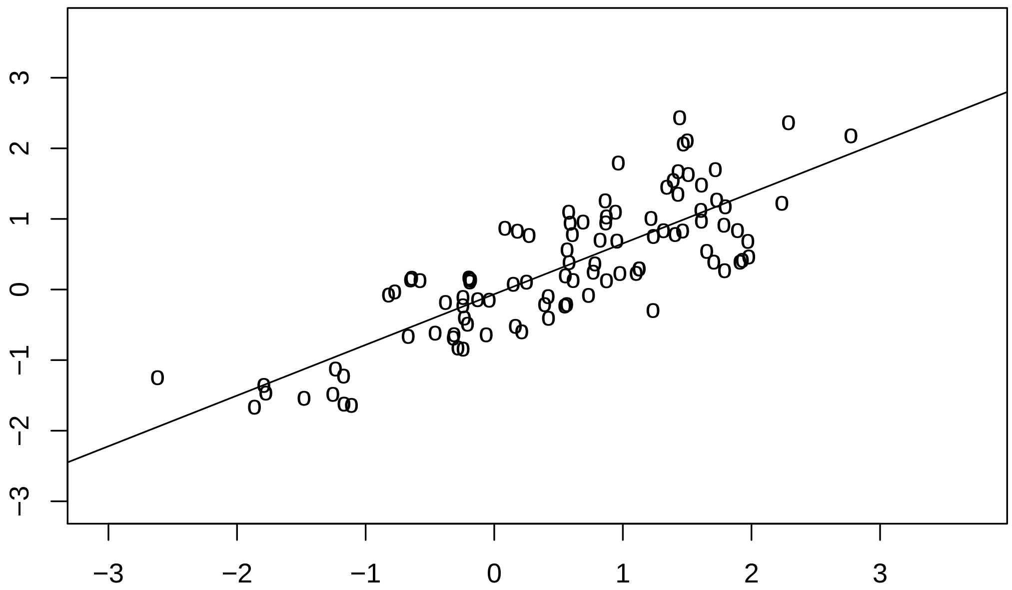
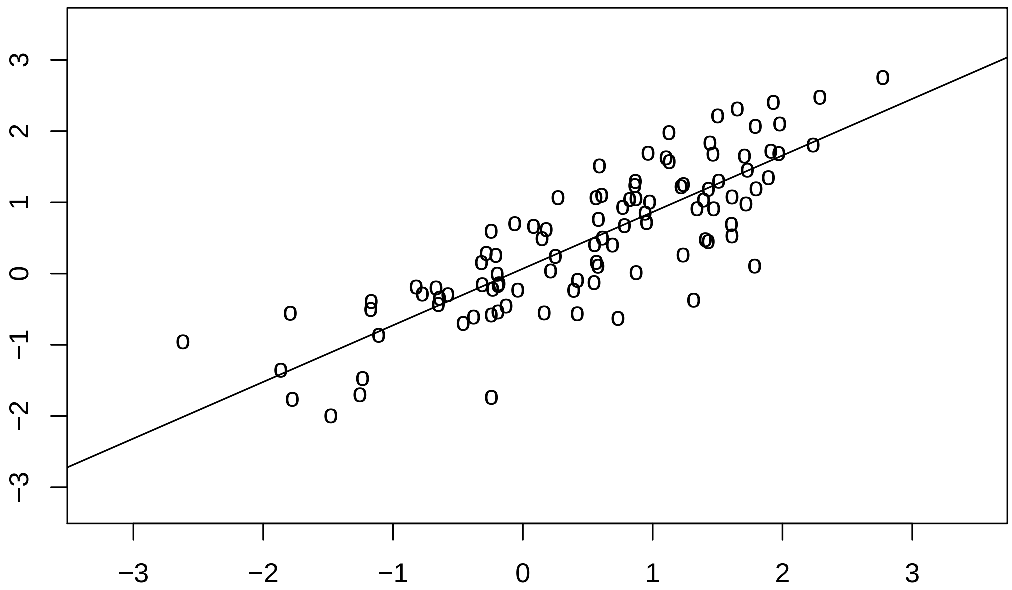


Figure 8. Flury-Shephard Estimate of x, SV Model



Next:
Application

- Extract the latent endowment (= consumption) process from a panel of assets in a Lucas (1978) economy

- Assuming that the latent endowment process is ARCH

- Using only the agents first order conditions

- And the scores of an ARCH process

- Thus avoiding payoff distribution assumptions and model solution

Lucas (1978) Economy

- Agent's first order conditions $1 = \mathcal{E}(M_{t+1} R_{t+1} \,|\, \mathcal{F}_t)$
- Marginal rate of substitution $M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{\gamma}$
- t is time in annual increments
- $\beta = 0.98$ is the discount factor
- $\gamma = 2$ is the risk aversion parameter,
- C_t is the endowment process,
- $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$ is the gross return
- $\mathcal{F}_t$ is the agent's information set at time t.

ARCH Process

- $x_t = \log(C_t/C_{t-1})$
- $x_t = \mu + \rho x_{t-1} + \sqrt{v_{t-1}} z_t$
- $v_{t-1} = \sigma^2 + [\tau(x_{t-1} - \mu - \rho x_{t-2})]^2$
- z_t is a standard normal

Implementation

- Data, from Gallant and Hong (2007), 1930–2004, annual
- Moments for conditional particle filter (dimension 700)
  - FOCs for returns on Fama-French portfolios, 30 day T-bills
  - Lagged returns on Fama-French portfolios, 30 day T-bills, instruments
  - Lagged consumption growth, labor income growth, instruments
  - Factor structure, known eigen vectors $\to$ diagonal variance matrix
- Moments for Metropolis step (dimension 4)
  - Scores of an ARCH process
- Endowment process x_t enters both, latent (dimension 75)

Differences Between GH and GGR

- GH posterior:

$$p_{gh}(y \,|\, \theta_{(1)})\, p_{gh}(\theta_{(1)} \,|\, \theta_{(2)})\, p_{gh}(\theta_{(2)})$$

- GGR posterior:

$$p_{gh}\{y \,|\, \beta[\text{exp}(x)]^{\gamma}\}\, p_{ggr}(x \,|\, \theta_{(3)})\, p_{ggr}(\theta_{(3)}),$$

- $\theta_{(1)}, \theta_{(2)}, \theta_{(3)}$ are different sets of parameters
- $p_{gh}(\cdot \,|\, \cdot)$ is the same in both
- GH have no latent variables: Inference by MCMC
- GGR model depends on the latent variables x

Prior

- Prior for $(\mu, \rho, \sigma, \tau)$ is a product of truncated normal densities
  - Location: $(\mu, \rho, \sigma, \tau) = (0.015, 0.35, 0.015, 0.01)$
  - Scale: Marginal probability of within $S \times 100\,\%$ of location is 95%
  - Truncation: $-100 < \mu < 100$, $-.999 < \rho < .999$, $0 < \sigma < 100$, and $0 < \tau < 100$,
- Prior for (β, γ) is $(\beta, \gamma) = (0.98, 2)$ with probability one.

Table 4. Estimates for MRS Model

Parameter	Mean	Standard Deviation
Prior Scale Factor = 50		
μ	0.18203	0.13172
ρ	0.56149	0.28256
σ	0.48836	0.0.2004
τ	0.23083	0.28940
β	0.98	0.0
γ	2.0	0.0
Prior Scale Factor = 500		
μ	0.22091	0.19625
ρ	0.51358	0.40154
σ	0.36577	0.34499
τ	1.68040	1.37980
β	0.98	0.0
γ	2.0	0.0
Prior Scale Factor = 1000		
μ	0.18904	0.25181
ρ	0.33079	0.56017
σ	0.29808	0.36470
τ	2.17500	1.75620
β	0.98	0.0
γ	2.0	0.0

Particles $N = 500$, Metropolis $K = 50$, chain $R = 9610$

Compare: consumption $(\mu, \rho, \sigma, \tau) = (0.015, 0.35, 0.015, 0)$

Figure 9. PF MRS Model

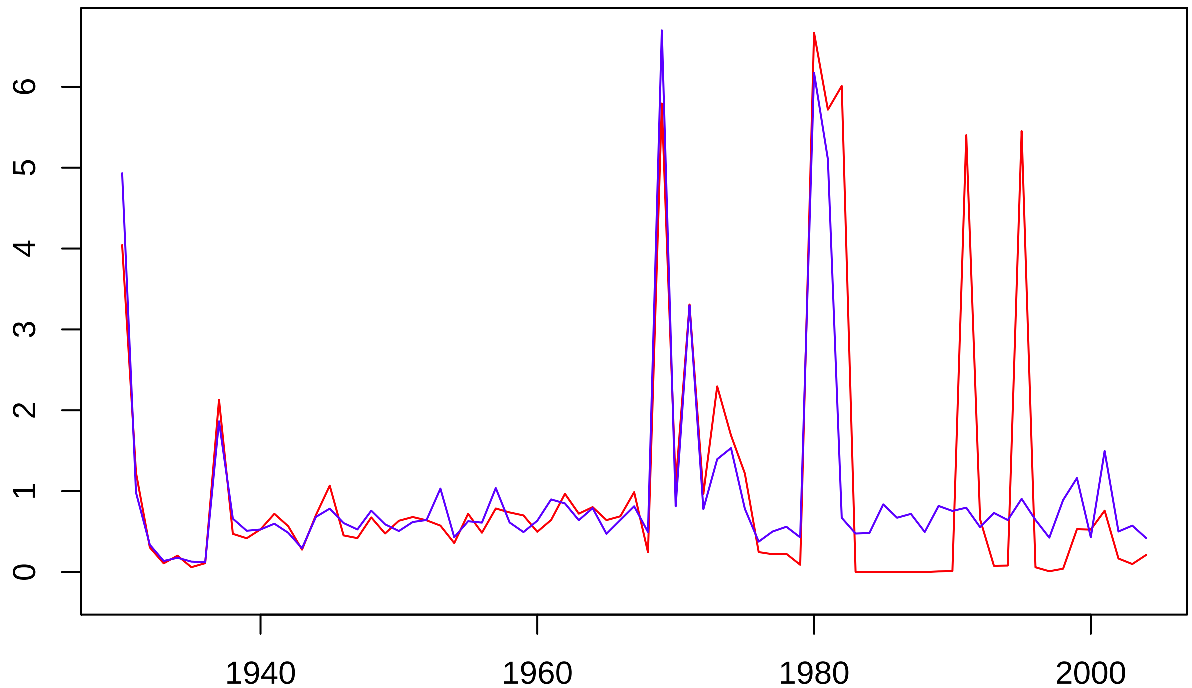


Figure 10. PF Estimate of MRS

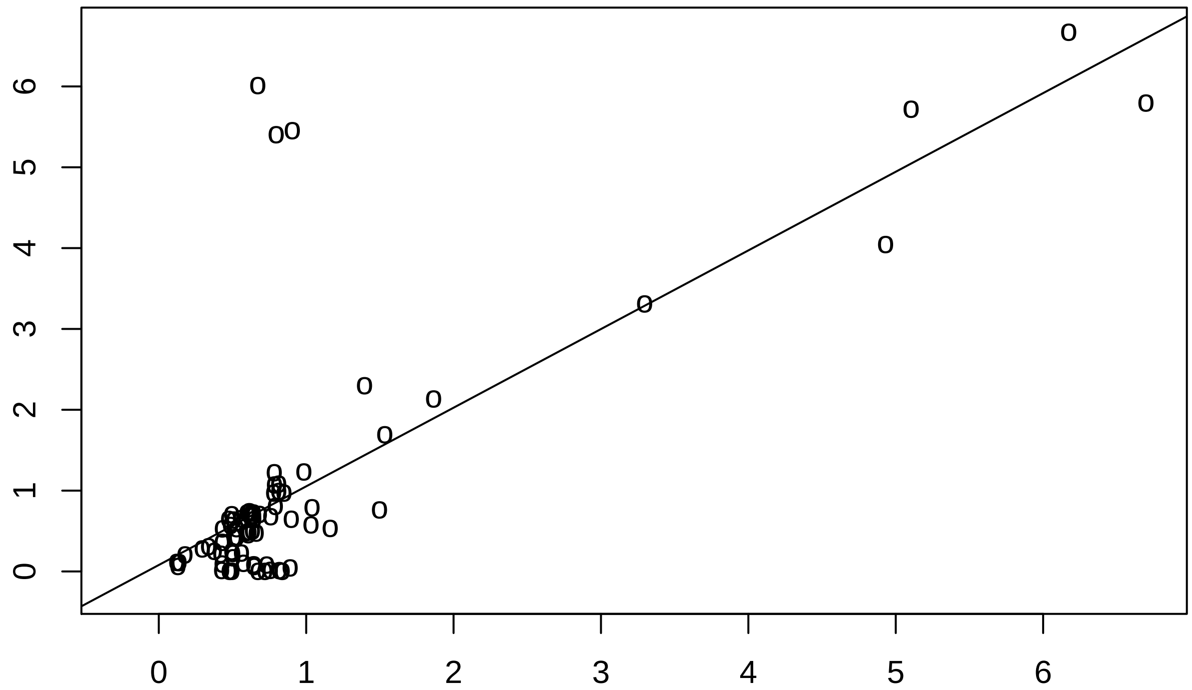
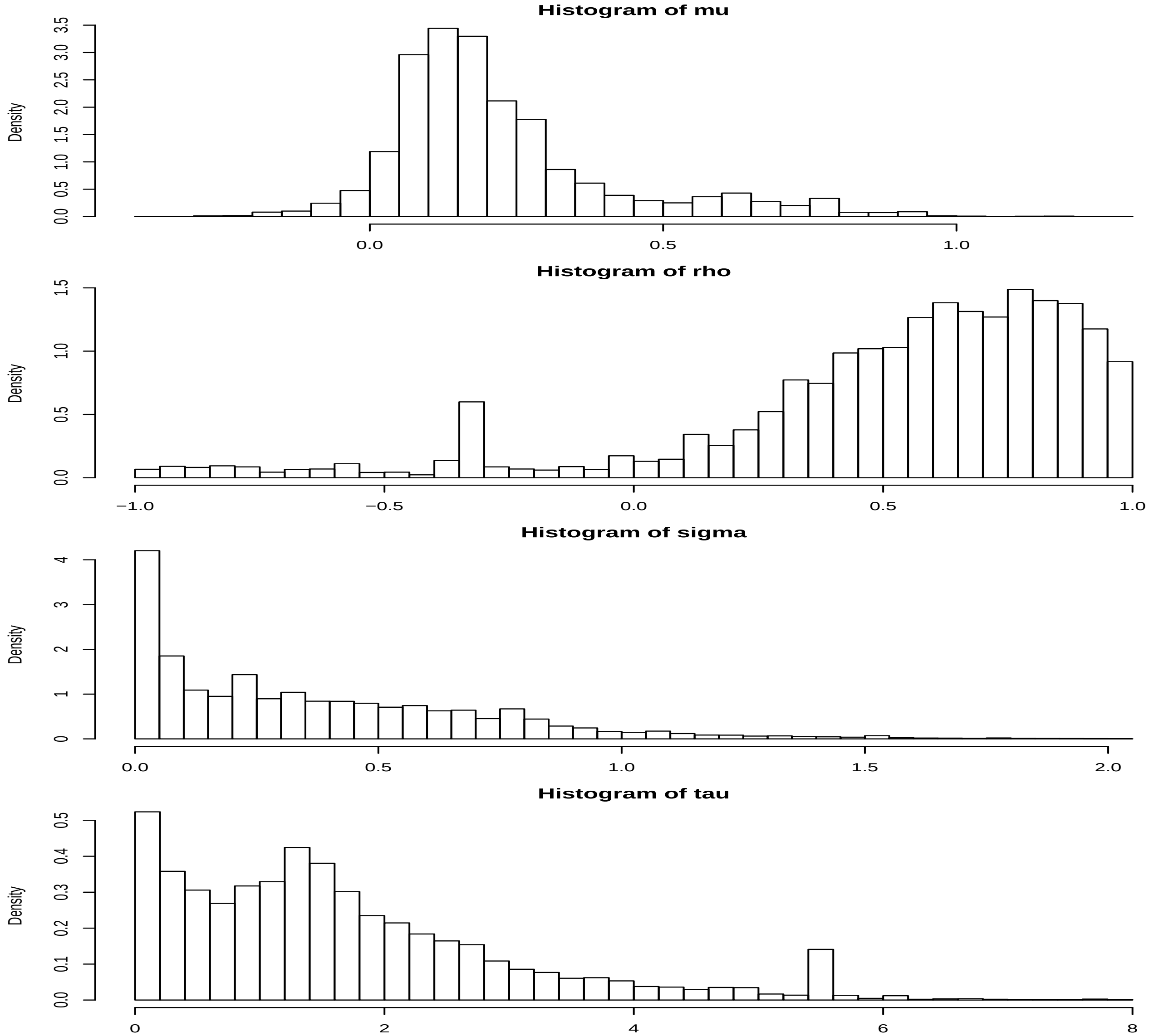


Figure 11. PF Estimate of MRS



Next:
The Three Algorithms

- A particle filter algorithm
  - Input: θ
  - Output: Draws $\left\{x^{(i)}\right\}_{i=1}^{R}$ from $p(x \,|\, y, \theta)$
- Conditional particle filter algorithm
  - Input: Draws $\theta^{(i-1)}$ and $x^{(i-1)}$
  - Output: A draw $x^{(i)}$ from $p(x \,|\, y, \theta)$
- Metropolis algorithm
  - Input: Draws $\theta^{(i-1)}$ and $x^{(i)}$
  - Output: A draw $\theta^{(i)}$ from $p(\theta \,|\, y, x)$

Notation

- $y_{1:t} = (y_1, ..., y_t)$

- $x_{1:t} = (x_1, ..., x_t)$

- $p(y_{1:t}, x_{1:t}, \theta)$
$$= (2\pi)^{-M/2} \exp\left\{ -\frac{1}{2} g_t(y_{1:t}, x_{1:t}, \theta)' \left[\Sigma(y_{1:t}, x_{1:t}, \theta)\right]^{-1} g_t(y_{1:t}, x_{1:t}, \theta) \right\}$$

Particle Filter Algorithm, 1 of 3

1. Initialization.

- Input θ (and y)

- Set T_0 to the minimum sample size required to compute $g_t(y_{1:t}, x_{1:t}, \theta)$.

- For $i = 1, \ldots, N$ sample $(x_1^{(i)}, x_2^{(i)}, \ldots, x_{T_0}^{(i)})$ from $p(x_t | x_{t-1}, \theta)$.

- Set t to $T_0 + 1$.

- Set $x_{1:t-1}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_{T_0}^{(i)})$

Particle Filter Algorithm, 2 of 3

2. Importance sampling step.

- For $i = 1, \ldots, N$ sample $\tilde{x}_t^{(i)}$ from $p(x_t | x_{t-1}^{(i)})$ and set

$$\tilde{x}_{1:t}^{(i)} = (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)}).$$

- For $i = 1, \ldots, N$ compute weights $\tilde{w}_t^{(i)} = p(y_{1:t}, \tilde{x}_{1:t}^{(i)}, \theta)$.

- Scale the weights to sum to one.

Particle Filter Algorithm, 3 of 3

3. Selection step.

- For $i = 1, \ldots, N$ sample with replacement particles $x_{1:t}^{(i)}$ from the set $\{\tilde{x}_{1:t}^{(i)}\}$ according to the weights.

4. Repeat

- If $t < T$, increment t and go to Importance Sampling Step;

- else output $\left\{x_{1:T}^{(i)}\right\}_{i=1}^{N}$.

Conditional Particle Filter Algorithm, 1 of 3

1. Initialization.

- Input $x_{1:T}^{(1)}$, θ (and y)

- Set T_0 to the minimum sample size required to compute $g_t(y_{1:t}, x_{1:t}, \theta)$.

- For $i = \mathbf{2}, \ldots, N$ sample $(x_1^{(i)}, x_2^{(i)}, \ldots, x_{T_0}^{(i)})$ from $p(x_t | x_{t-1}, \theta)$.

- Set t to $T_0 + 1$.

- Set $x_{1:t-1}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_{T_0}^{(i)})$

Conditional Particle Filter Algorithm, 2 of 3

2. Importance sampling step.

- For $i = \mathbf{2}, \ldots, N$ sample $\tilde{x}_t^{(i)}$ from $p(x_t | x_{t-1}^{(i)})$ and set

$$\tilde{x}_{1:t}^{(i)} = (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)}).$$

- For $i = \mathbf{1}, \ldots, N$ compute weights $\tilde{w}_t^{(i)} = p(y_{1:t}, \tilde{x}_{1:t}^{(i)}, \theta)$.

- Scale the weights to sum to one.

Conditional Particle Filter Algorithm, 3 of 3

3. Selection step.

- For $i = \mathbf{2}, \ldots, N$ sample with replacement particles $x_{1:t}^{(i)}$ from the set $\{\tilde{x}_{1:t}^{(i)}\}_{i=1}^{N}$ according to the weights.

4. Repeat

- If $t < T$, increment t and go to Importance Sampling Step;

- else output the particle $x_{1:T}^{(N)}$.

Metropolis Algorithm

Proposal density: $T(\theta_{here}, \theta_{there})$ (e.g., move one-at-time random walk)

- Input: x, θ_{old} (and y)

- Propose: Draw θ_{prop} from $T(\theta_{old}, \theta)$

- Accept-Reject: Put $\theta^{(i)}$ to θ_{prop} with probability

$$\alpha = \min\left[1, \frac{p(y, x, \theta_{prop})T(\theta_{prop}, \theta_{old})}{p(y, x, \theta_{old})T(\theta_{old}, \theta_{prop})}\right]$$

  else put $\theta^{(i)}$ to θ_{old}.

- Repeat: If $i < K$ put $\theta_{old} = \theta^{(i)}$ and go to Propose;

- else output $\theta^{(K)}$.