GMM with Latent Variables

A. Ronald Gallant
Penn State University

Raffaella Giacomini
University College London

Giuseppe Ragusa
Luiss University

Contribution

- The contribution of GMM (Hansen and Singleton, 1982) was to allow frequentist inference regarding the parameters of a nonlinear structural model without having to solve the model.
  - Provided there were no latent variables.

- The contribution of this paper is the same.
  - With latent variables.

The Requirements, 1 of 3

- A structural model with parameters θ and true value θ^o
- Observed variables: $X = (X_1, X_2, ..., X_T)$
- Latent variables: $\Lambda = (\Lambda_1, \Lambda_2, ..., \Lambda_T)$
- Known transition density: $\Lambda_{t+1} \sim P(\Lambda_{t+1} \,|\, \Lambda_t, \theta)$
- Conditional moment conditions: $\mathcal{E}\Big[g(X_{t+1}, \Lambda_{t+1}, \theta) \,|\, \mathcal{I}_t\Big] = 0$
  - That would identify θ if both X and Λ were observed.
  - $\mathcal{I}_t = \{X_{-\infty}, ..., X_t, \, \Lambda_{-\infty}, ..., \Lambda_t\}$

The Requirements, 2 of 3

- Sample moment conditions

$$g_T(X,\Lambda,\theta) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(X_t,\Lambda_t,\theta)$$

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

$$\begin{aligned} \Sigma(X,\Lambda,\theta) &= \frac{1}{T} \sum_{t=1}^{T} \tilde{g}(X_t,\Lambda_t,\theta)'\tilde{g}(X_t,\Lambda_t,\theta) \\ \tilde{g}(X_t,\Lambda_t,\theta) &= g(X_t,\Lambda_t,\theta) - \frac{1}{\sqrt{T}} g_T(X,\Lambda,\theta) \end{aligned}$$

- Such that

$$Z = [\Sigma(X,\Lambda,\theta^o)]^{-1/2} g_T(X,\Lambda,\theta^o) \xrightarrow{d} N(0,I)$$

  - Hansen and Singleton (1982)
  - Gallant and White (1987)

The Requirements, 3 of 3

- A sample $\left\{\theta^{(i)}\right\}_{i=1}^{R}$ from the density

$$p(\theta) = (2\pi)^{-M/2} \exp\left\{-\frac{1}{2} g_T(X, \Lambda, \theta)' \left[\Sigma(X, \Lambda, \theta)\right]^{-1} g_T(X, \Lambda, \theta)\right\}$$

  is a sample from the asymptotic distribution of the GMM estimator for large T.

  – Chernozhukov and Hong (2003)

Estimation Strategy

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

$$p(\theta,\Lambda) = (2\pi)^{-M/2} \exp\left\{-\frac{1}{2} g_T(X,\Lambda,\theta)' \left[\Sigma(X,\Lambda,\theta)\right]^{-1} g_T(X,\Lambda,\theta)\right\}$$

  - Might multiply $p(\theta,\Lambda)$ by a Jacobian term $[\det \Sigma(X,\Lambda,\theta)]^{-M/2}$

- Metropolis within Gibbs algorithm
  - Sample $\theta^{(i)}$ given $\Lambda^{(i-1)}$ and $\theta^{(i-1)}$ using Metropolis
    - ∗ last draw of MCMC chain of length K.
  - Sample $\Lambda^{(i)}$ given $\theta^{(i)}$ and $\Lambda^{(i-1)}$ using Gibbs.
    - ∗ 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 so

$$X_t = (w_t, y_t, \pi_t)$$

$$\Lambda_t = (z_t, \phi_t).$$

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 $\mathcal{L}(\theta) = p(X|\theta)$ 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

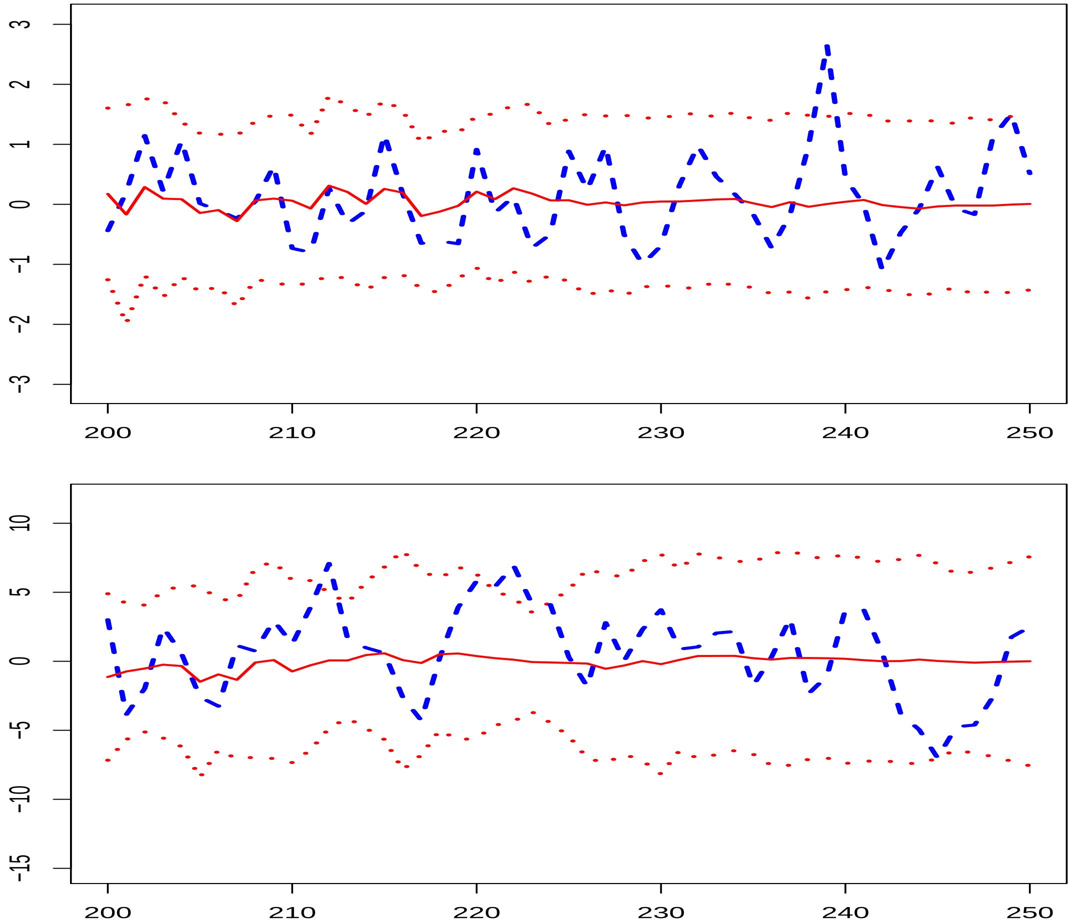
- We calibrate σ_z, σ_ϕ, ν and leave β as the free parameter.

Table 1. Parameter Estimates, DSGE Model

Parameter	True Value	Mean	Mode	Standard Error
		With Jacobian		
ρ_z	0.15	0.21596	0.15006	0.08632
ρ_ϕ	0.68	0.60098	0.58945	0.04988
ρ_λ	0.56	0.50134	0.46443	0.28818
σ_λ	0.11	0.10827	0.08923	0.06494
β	0.996	0.98429	0.99603	0.01476
		Without Jacobian		
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. Gibbs particles are $N = 1000$; Metropolis draws are $K = 50$. 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 Λ with Jacobian, DSGE Model



Remark: The Gibbs draw should evaluate the moments in the Metropolis step accurately; not necessarily approximate the history accurately.

Figure 2. PF Estimate of Λ without Jacobian, DSGE Model

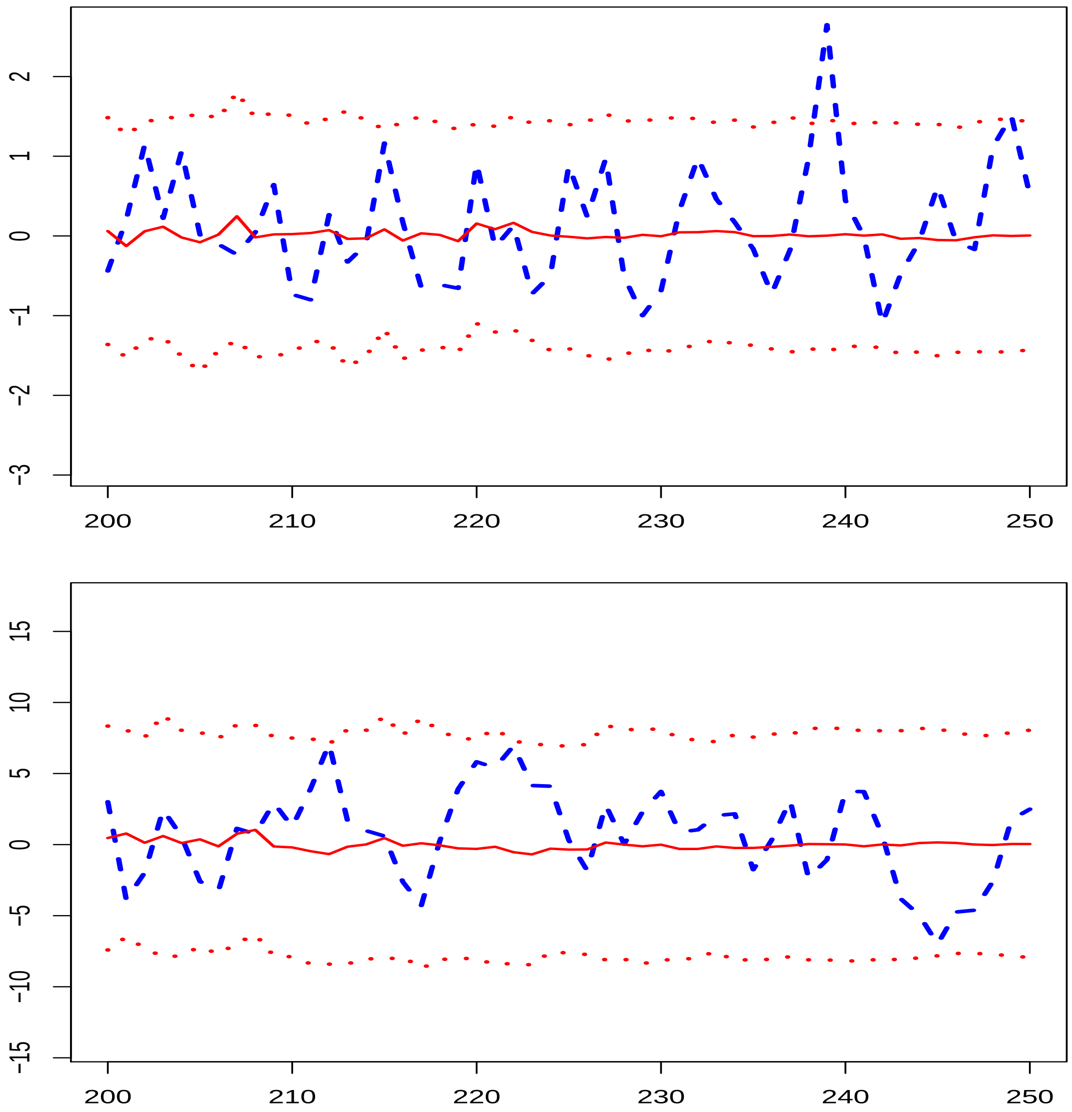
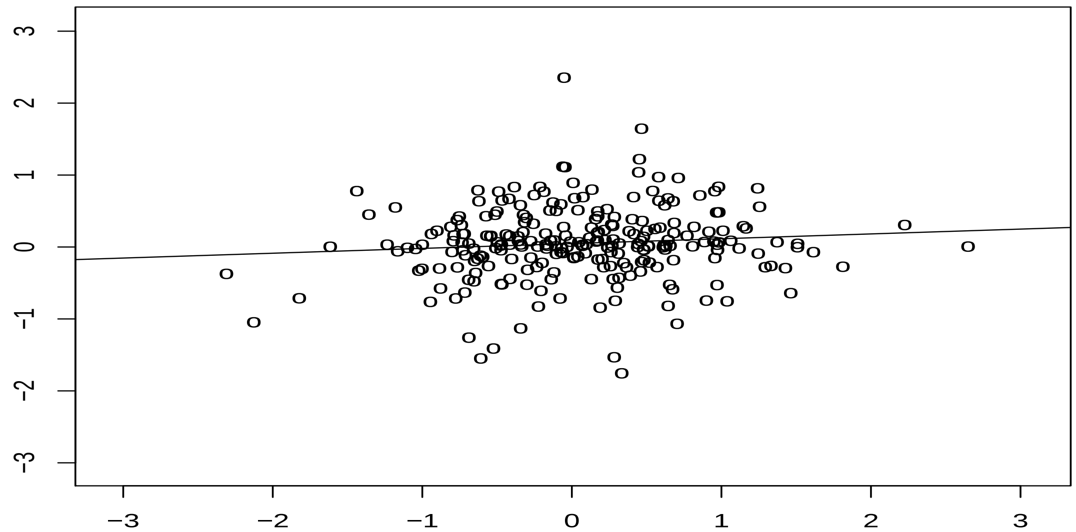


Figure 3. PF Estimate of Λ with Jacobian, DSGE Model



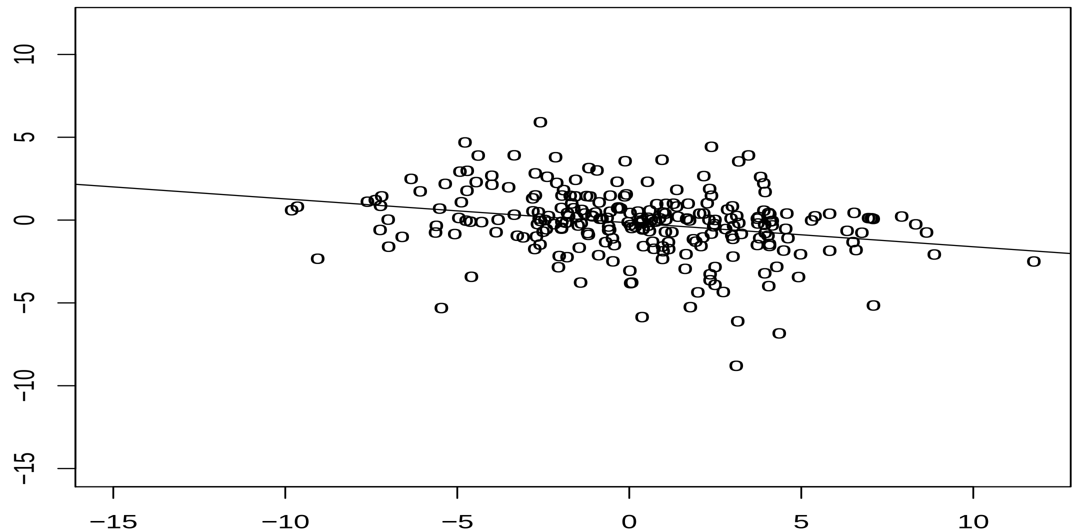
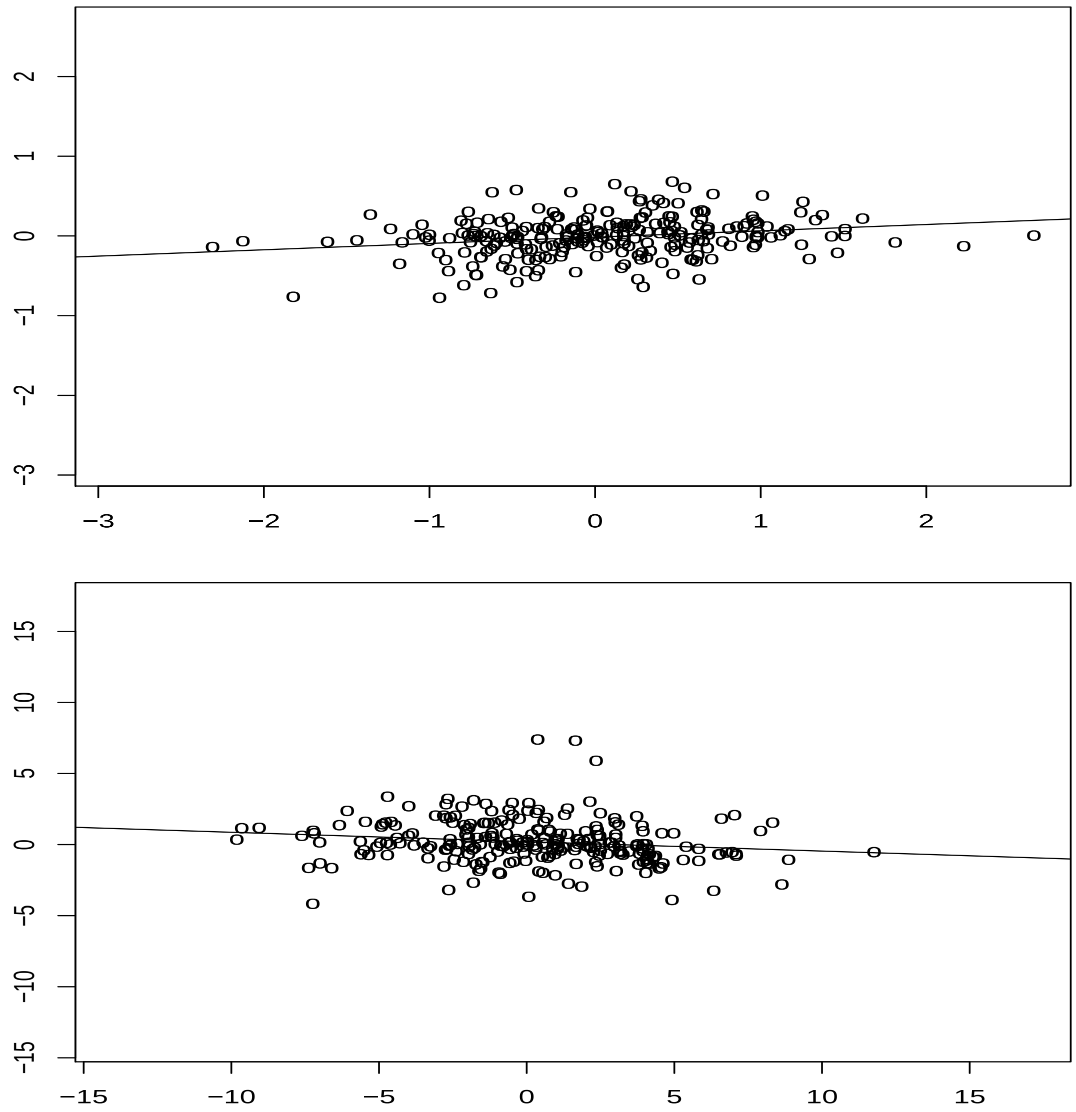


Figure 4. PF Estimate of Λ without Jacobian, DSGE Model



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 5. PF Estimate of Λ with Jacobian, DSGE Model

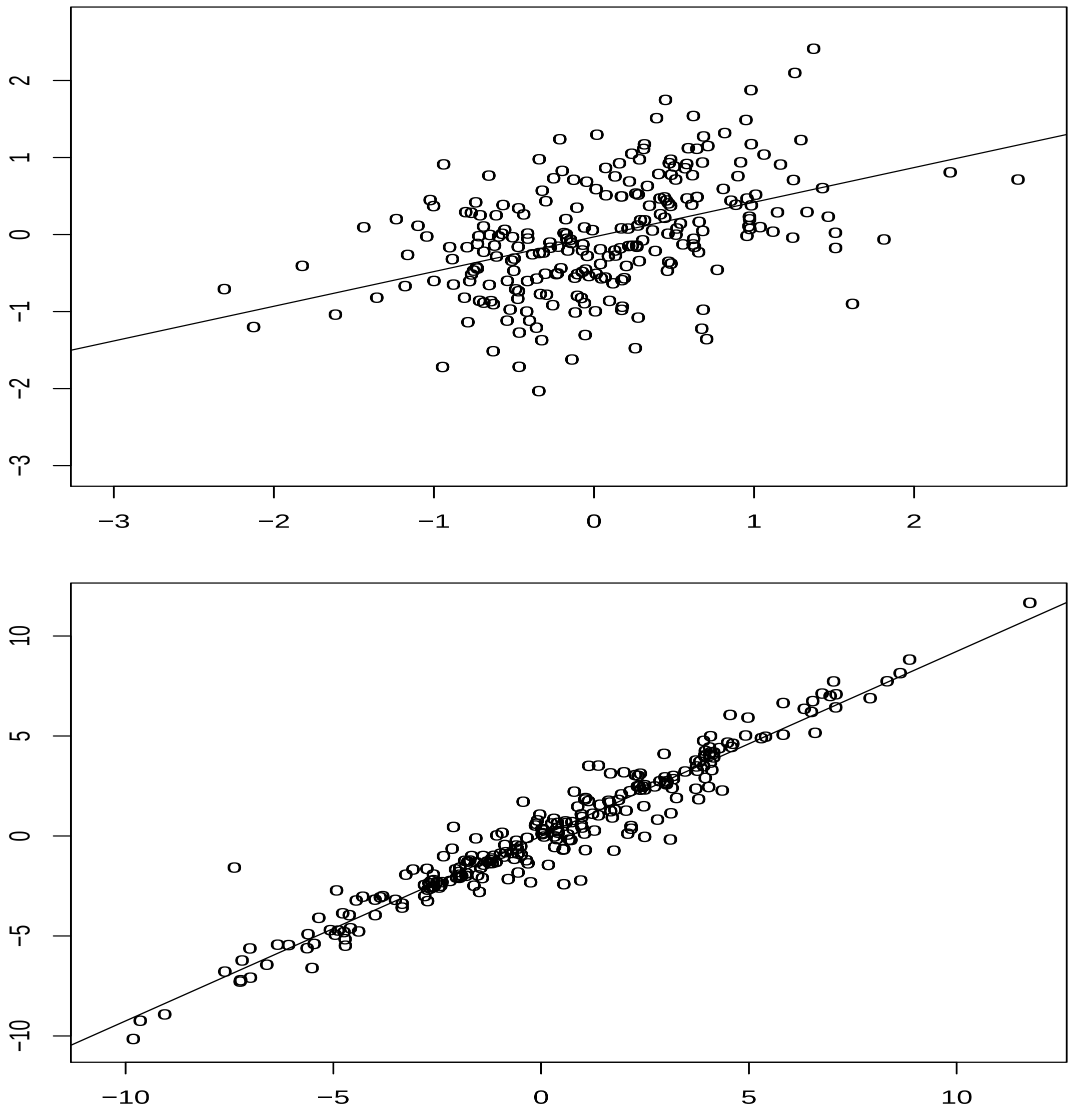


Figure 6. PF Estimate of Λ without Jacobian, DSGE Model

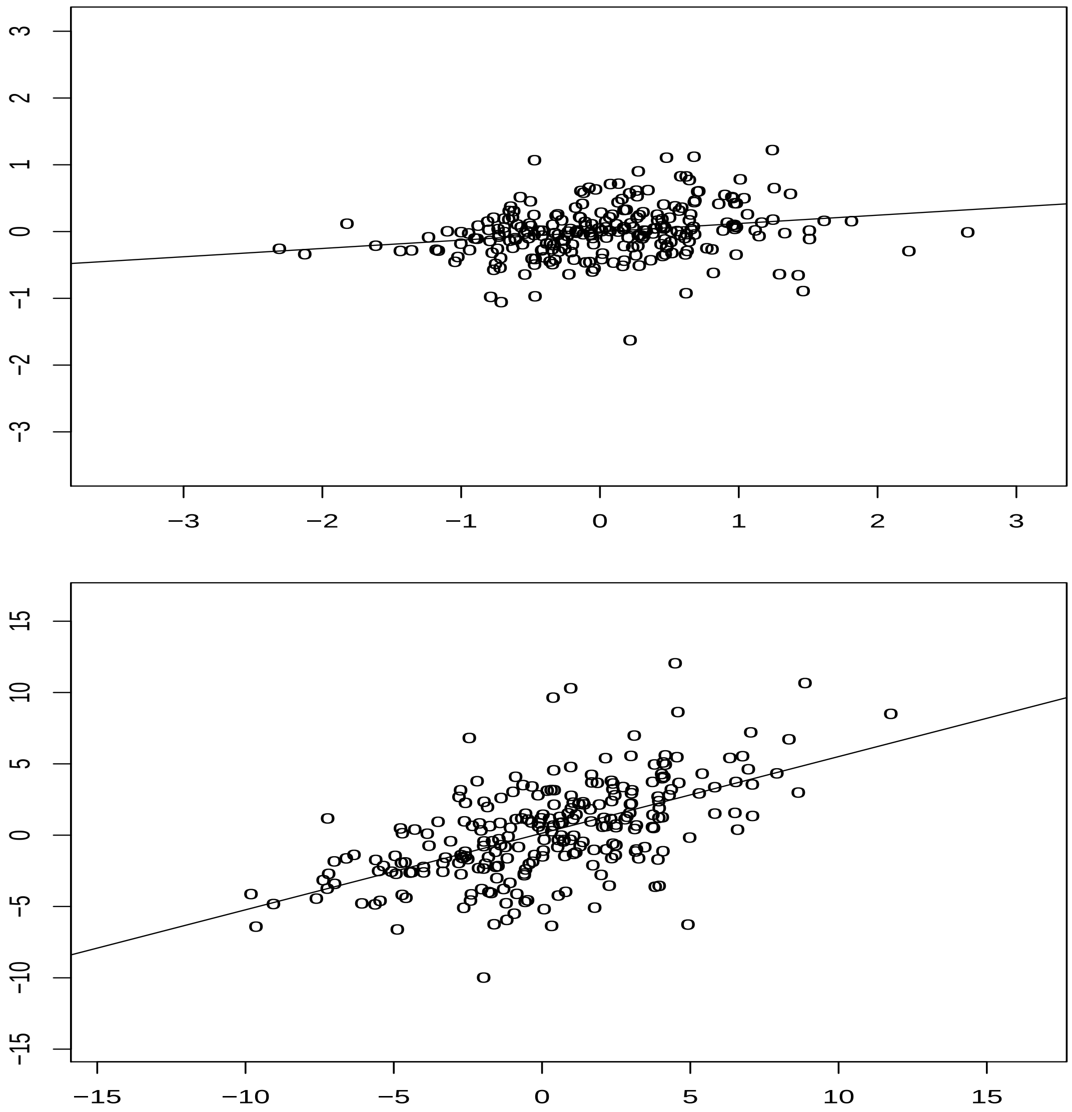


Figure 7. PF Estimate of Λ with Jacobian, DSGE Model

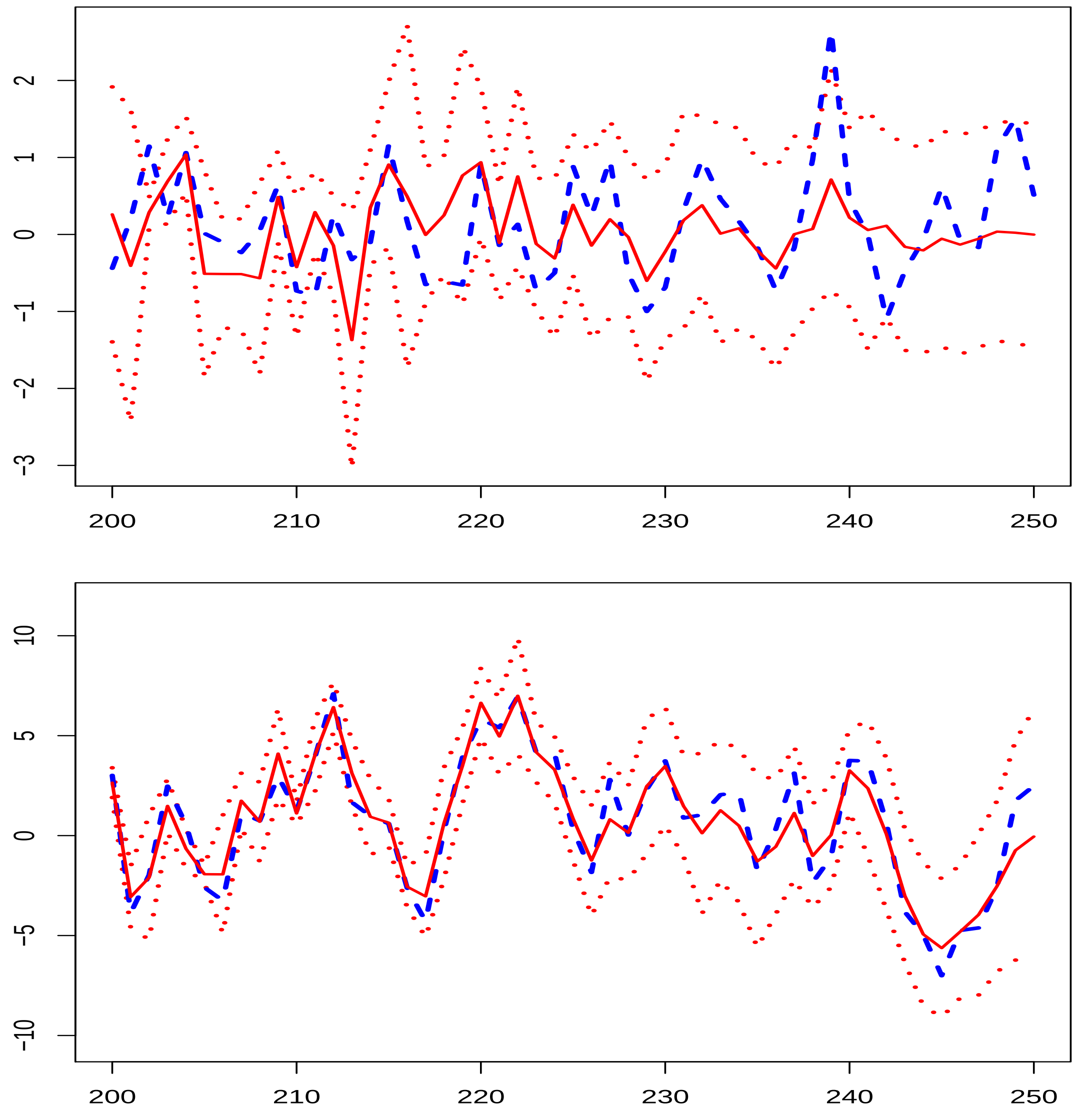
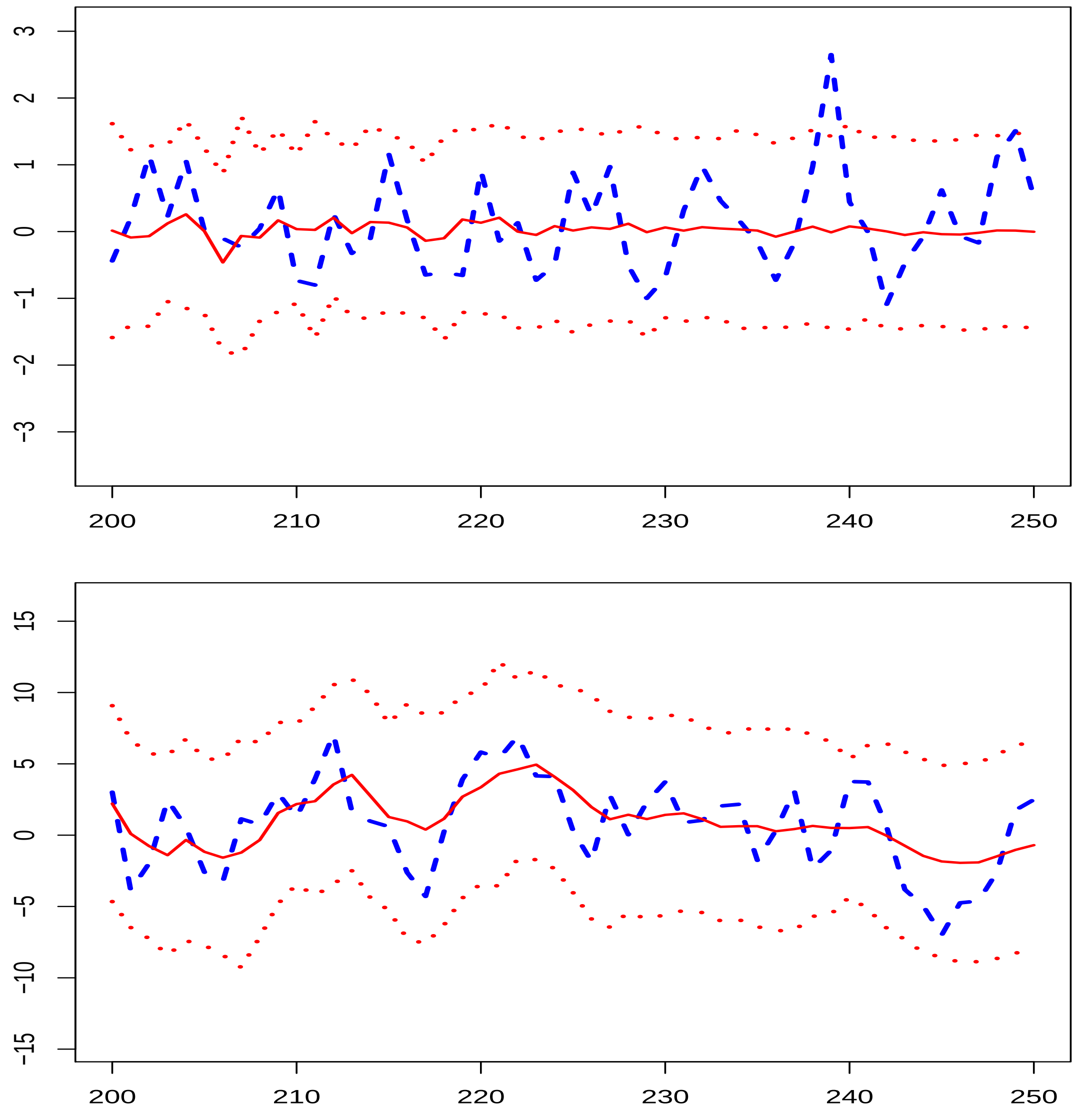


Figure 8. PF Estimate of Λ without Jacobian, DSGE Model



Gibbs 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 5 through 8 for the Gibbs step, we get slightly better results in the following Table 2.

Table 2. Alternative Parameter Estimates, DSGE Model

Parameter	True Value	Mean	Mode	Standard Error
		With Jacobian		
ρ_z	0.15	0.21702	0.15006	0.08367
ρ_ϕ	0.68	0.61408	0.58945	0.05102
ρ_λ	0.56	0.50082	0.46443	0.28344
σ_λ	0.11	0.11086	0.08924	0.06493
β	0.996	0.98740	0.99603	0.01056
		Without Jacobian		
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. Gibbs particles are $N = 1000$; Metropolis draws are $K = 50$. 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

$$X_t = \rho X_{t-1} + \exp(\Lambda_t)\, u_t \quad (1)$$

$$\Lambda_t = \phi \Lambda_{t-1} + \sigma e_t \quad (2)$$

$$e_t \sim N(0,1) \quad (3)$$

$$u_t \sim N(0,1) \quad (4)$$

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

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

Table 3. Parameter Estimates, SV Model

Parameter	True Value	Mean	Mode	Standard Error
		With Jacobian Term		
ρ	0.25	0.30488	0.30961	0.074778
ϕ	0.8	0.09153	0.94851	0.660790
σ	0.1	0.09023	0.06702	0.050229
		Without Jacobian		
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 all panels the number of particles is $N = 1000$. The columns labeled mean, mode, and standard deviation are the mean, mode, and standard deviations of an MCMC chain of length 200000.

Figure 9. PF Estimate of Λ with Jacobian Term, SV Model

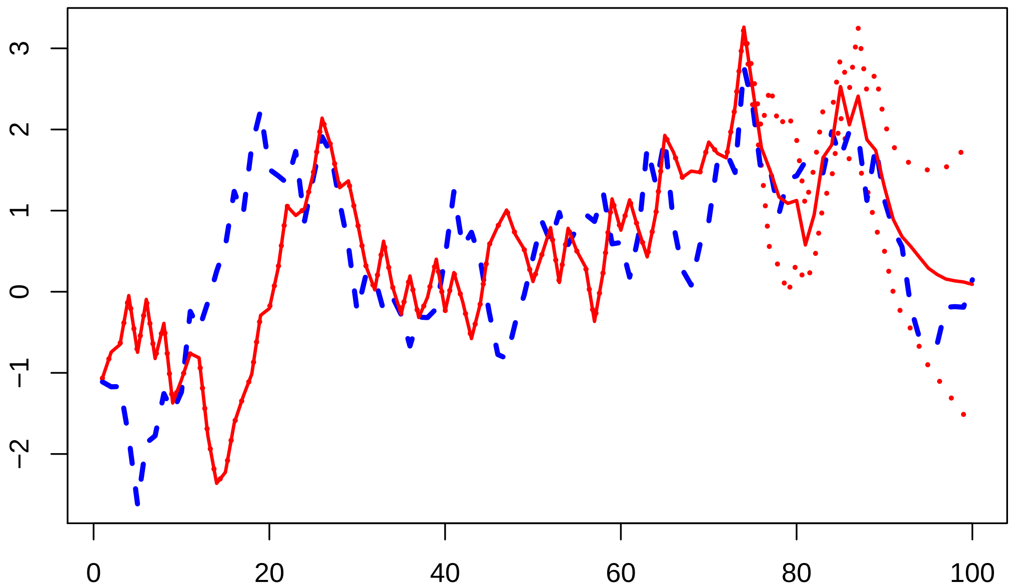


Figure 10. PF Estimate of Λ without Jacobian, SV Model

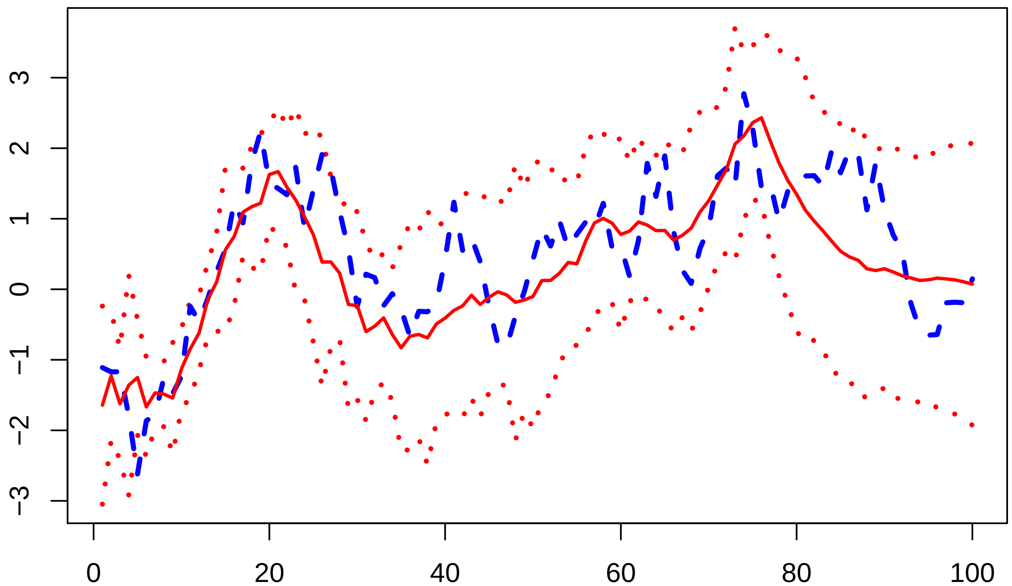


Figure 11. Flury-Shephard Estimate of Λ, SV Model

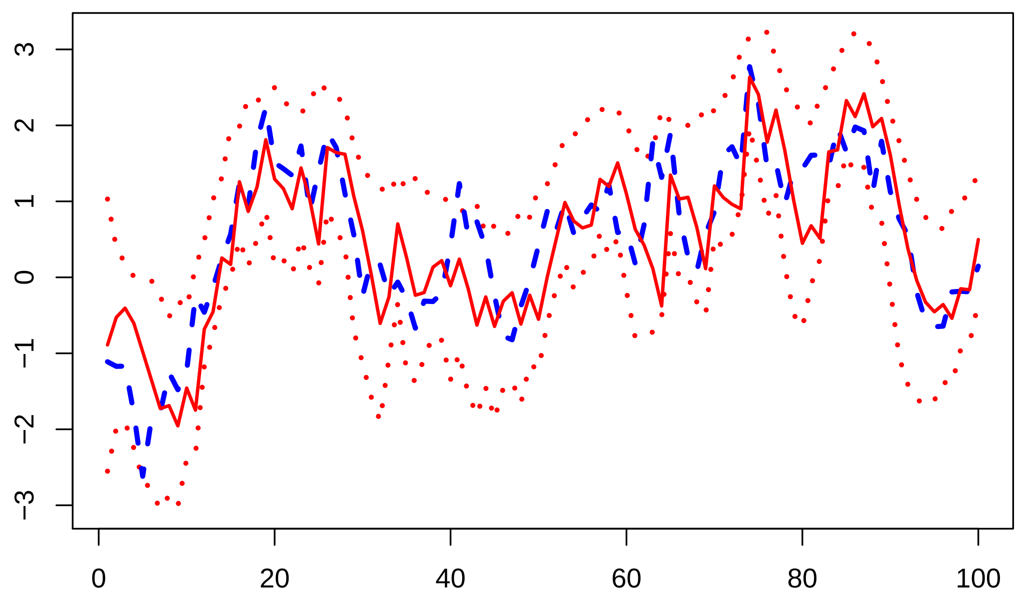


Figure 12. PF Estimate of Λ with Jacobian Term, SV Model

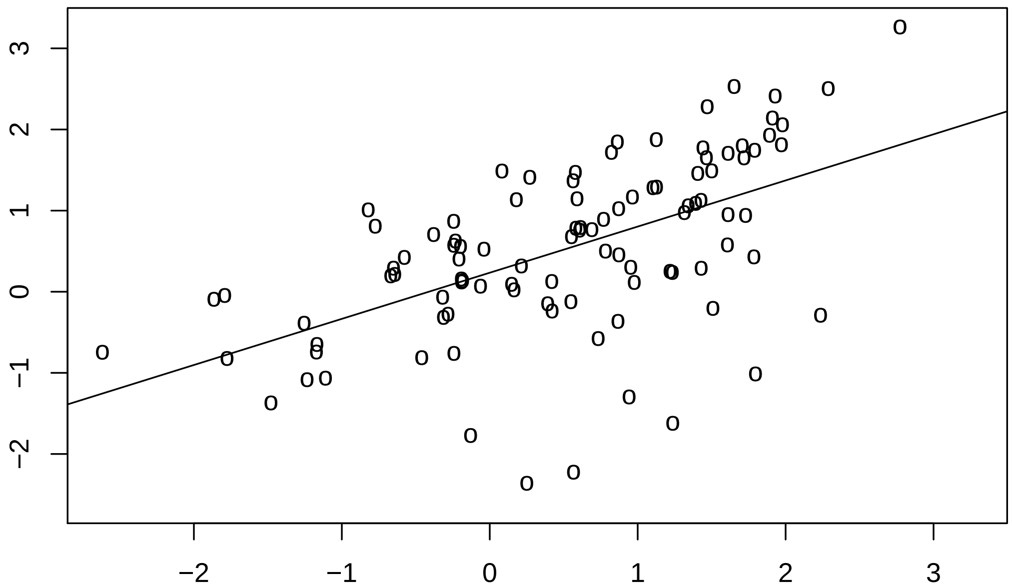


Figure 13. PF Estimate of Λ without Jacobian, SV Model

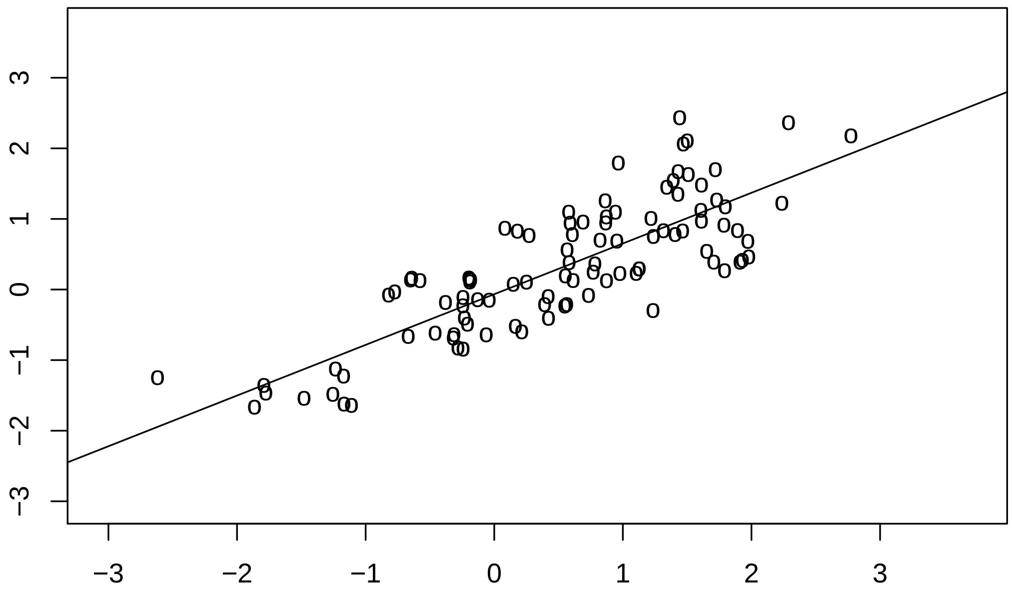
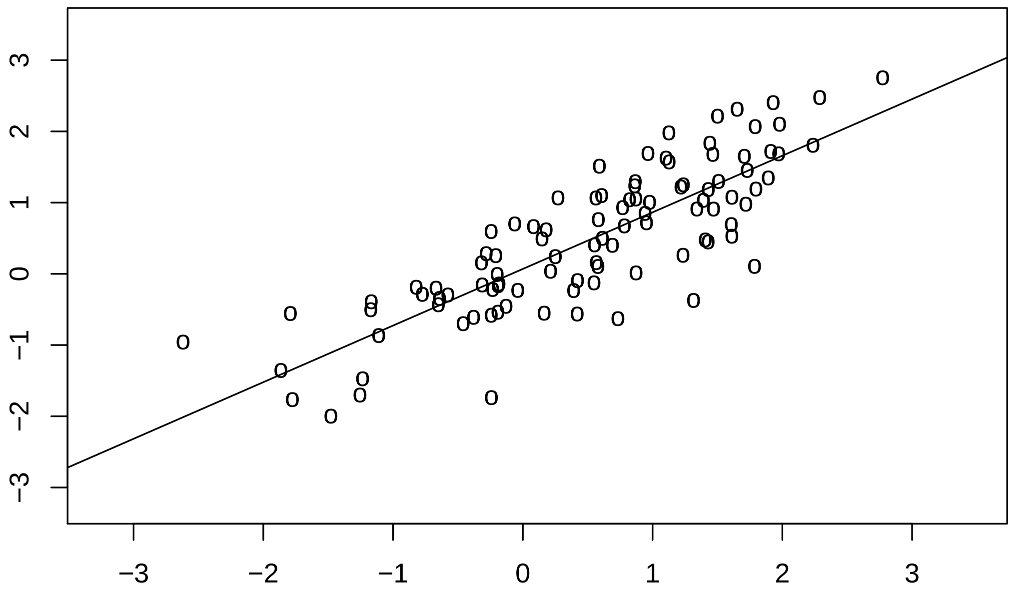


Figure 14. Flury-Shephard Estimate of Λ, SV Model



Next:
The Three Algorithms

- A particle filter algorithm
  - Input: θ
  - Output: Draws $\left\{\Lambda^{(i)}\right\}_{i=1}^{R}$ from $P(\Lambda \,|\, X, \theta)$
- Gibbs algorithm
  - Input: Draws $\theta^{(i-1)}$ and $\Lambda^{(i-1)}$
  - Output: A draw $\Lambda^{(i)}$ from $P(\Lambda \,|\, X, \theta)$
- Metropolis algorithm
  - Input: Draws $\theta^{(i-1)}$ and $\Lambda^{(i)}$
  - Output: A draw $\theta^{(i)}$ from $P(\theta \,|\, X, \Lambda)$

Notation

- $X_{1:t} = (X_1, ..., X_t)$

- $\Lambda_{1:t} = (\Lambda_1, ..., \Lambda_t)$

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

Particle Filter Algorithm, 1 of 3

1. Initialization.

- Input θ (and X)

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

- For $i = 1, \ldots, N$ sample $(\Lambda_1^{(i)}, \Lambda_2^{(i)}, \ldots, \Lambda_{T_0}^{(i)})$ from $p(\Lambda_t | \Lambda_{t-1}, \theta)$.

- Set t to $T_0 + 1$.

- Set $\Lambda_{1:t-1}^{(i)} = (\Lambda_1^{(i)}, \Lambda_2^{(i)}, \ldots, \Lambda_{T_0}^{(i)})$

Particle Filter Algorithm, 2 of 3

2. Importance sampling step.

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

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

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

- 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 $\Lambda_{1:t}^{(i)}$ from the set $\{\tilde{\Lambda}_{1:t}^{(i)}\}$ according to the weights.

4. Repeat

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

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

Gibbs Algorithm, 1 of 3

1. Initialization.

- Input $\Lambda_{1:T}^{(1)}$, θ (and X)

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

- For $i = \mathbf{2}, \ldots, N$ sample $(\Lambda_1^{(i)}, \Lambda_2^{(i)}, \ldots, \Lambda_{T_0}^{(i)})$ from $p(\Lambda_t | \Lambda_{t-1}, \theta)$.

- Set t to $T_0 + 1$.

- Set $\Lambda_{1:t-1}^{(i)} = (\Lambda_1^{(i)}, \Lambda_2^{(i)}, \ldots, \Lambda_{T_0}^{(i)})$

Gibbs Algorithm, 2 of 3

2. Importance sampling step.

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

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

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

- Scale the weights to sum to one.

Gibbs Algorithm, 3 of 3

3. Selection step.

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

4. Repeat

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

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

Metropolis Algorithm

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

- Input: Λ, θ_{old} (and X)

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

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

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

  else put $\theta^{(i)}$ to θ_{old}.

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

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

Next:
Why Does this Work?

- Prove that the particle filter works using the notion of Gallant, A. Ronald, and Han Hong (2007), "A Statistical Inquiry into the Plausibility of Recursive Utility," *Journal of Financial Econometrics*, that GMM induces a probability space;
  - next several slides

- That the Metropolis algorithm works follows from Chernozhukov, Victor, and Han Hong (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics*.

- That the Gibbs algorithm works follows from Andrieu, C., A. Douced, and R. Holenstein (2010), "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Series B*.

Joint Density Induced by GMM, Dice Example

Table 4. Tossing two correlated dice (X, Λ) when the probability of the difference $D = X - \Lambda$ is the primitive.

Preimage	d	$P(D = d)$	$P(D = d \mid \Lambda = 1)$	$P(D = d \mid \Lambda = 2)$
$C_{-5} = \{(1,6)\}$	-5	0	0	0
$C_{-4} = \{(1,5),(2,6)\}$	-4	0	0	0
$C_{-3} = \{(1,4),(2,5),(3,6)\}$	-3	0	0	0
$C_{-2} = \{(1,3),(2,4),(3,5),(4,6)\}$	-2	0	0	0
$C_{-1} = \{(1,2),(2,3),(3,4),(4,5),(5,6)\}$	-1	4/18	0	4/18
$C_{0} = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$	0	10/18	10/14	10/18
$C_{1} = \{(2,1),(3,2),(4,3),(5,4),(6,5)\}$	1	4/18	4/14	4/18
$C_{2} = \{(3,1),(4,2),(5,3),(6,4))\}$	2	0	0	0
$C_{3} = \{(4,1),(5,2),(6,3)\}$	3	0	0	0
$C_{4} = \{(5,1),(6,2)\}$	4	0	0	0
$C_{5} = \{(6,1)\}$	5	0	0	0

Conditional probability is $P(D = d \mid \Lambda = \lambda) = P(C_d \cap O_\lambda)/P(O_\lambda)$, where O_λ is the union of the events that can occur. $Q(\Lambda = \lambda) = P(O_\lambda)$ is the marginal in the sense that $P(D = d) = \sum_{\lambda=1}^{6} P(D = d \mid \Lambda = \lambda) Q(\Lambda = \lambda)$

Conditional Density, Dice Example, 1 of 2

- Let $\mathcal{C}$ be the smallest σ-algebra that contains the preimages in Table 4.

- Any $\mathcal{C}$-measurable f must be constant on the preimages.

- For such f the formula

$$\mathcal{E}(f \mid \Lambda = 2) = \sum_{x=1}^{6} f(x, 2) \sum_{d=-5}^{5} I_{C_d}(x, 2) P(D = d \mid \Lambda = 2) \tag{5}$$

  can be used to compute conditional expectation because f can be regarded as a function of d and the right hand side of (5) equals

$$\sum_{d=-5}^{5} f(d) P(D = d \mid \Lambda = 2).$$

Conditional Density, Dice Example, 2 of 2

- Equation (5) implies that we can view $P(D = d)$ as defining a conditional density function

$$P(X = x \,|\, \Lambda = \lambda) = \sum_{d=-5}^{5} I_{C_d}(x, \lambda) P(D = d \,|\, \Lambda = \lambda) \tag{6}$$

  that is a function of x as long as we only use it in connection with $\mathcal{C}$-measurable f.

- To get an expression that agrees with the expressions in Gallant and Hong (2007) note that we can write equation (6) as

$$P(X = x \,|\, \Lambda = \lambda) = \frac{P(D = x - \lambda)}{\sum_{x=1}^{6} P(D = x - \lambda)} \tag{7}$$

- Similarly,

$$P(\Lambda = \lambda \,|\, X = x) = \frac{P(D = x - \lambda)}{\sum_{\lambda=1}^{6} P(D = x - \lambda)} \tag{8}$$

Abstraction

- A GMM criterion $Z(X, \Lambda, \theta)$ defines a probability space

$$(\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}, P_\theta)$$

  - $\mathcal{C}$ is the smallest σ-algebra containing the preimages of Z

- On which there are notions of joint

$$p(X, \Lambda, \theta) = (2\pi)^{-M/2} \exp\left\{ -\frac{1}{2} g_T(X, \Lambda, \theta)' \left[\Sigma(X, \Lambda, \theta)\right]^{-1} g_T(X, \Lambda, \theta) \right\},$$

conditional $p(X \mid \Lambda, \theta)$, and marginal densities.

- If P_θ^o denotes the data generating process, then

$$(\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}, P_\theta) = (\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}, P_\theta^o)$$

What if One Knows a Marginal?, Dice Example

- Then one knows the probabilities $P(R_\lambda)$ of the rectangles

$$
\begin{aligned}
R_\lambda &= \mathbb{D} \times \{\lambda\} \\
\mathbb{D} &= \{1, 2, 3, 4, 5, 6\}
\end{aligned}
$$

- Let $\mathcal{C}^*$ be the smallest σ-algebra containing $\{C_d\}_{d=-5}^{5}$ and $\{R_\lambda\}_{\lambda=1}^{6}$

- The singleton sets $\{(x, \lambda)\}$ are in $\mathcal{C}^*$ so joint probability P^* on $\mathcal{C}^*$ and conditional densities have their conventional definition

  - $P^*(X = x \,|\, \Lambda = \lambda) = \frac{P^*(\{(x,\lambda)\})}{P^*(R_\lambda)}$

  - $P^*(\Lambda = \lambda \,|\, X = x) = \frac{P^*(\{(x,\lambda)\})}{P^*(R_x)}$

Indeterminacy, Dice Example, 1 of 2

For $P^*(\{(x,\lambda)\})$ we have nine equations in sixteen unknowns:

$$\begin{aligned}
\frac{4}{18} &= \sum_{i=1}^{5} P^*(\{(i,i+1)\}) \\
\frac{10}{18} &= \sum_{i=1}^{6} P^*(\{(i,i)\}) \\
\frac{4}{18} &= \sum_{i=1}^{5} P^*(\{(i+1,i)\}) \\
\frac{1}{6} &= P^*(\{(1,1)\}) + P^*(\{(2,1)\}) \\
\frac{1}{6} &= P^*(\{(1,2)\}) + P^*(\{(2,2)\}) + P^*(\{(3,2)\}) \\
\frac{1}{6} &= P^*(\{(2,3)\}) + P^*(\{(3,3)\}) + P^*(\{(4,3)\}) \\
\frac{1}{6} &= P^*(\{(3,4)\}) + P^*(\{(4,4)\}) + P^*(\{(5,4)\}) \\
\frac{1}{6} &= P^*(\{(4,5)\}) + P^*(\{(5,5)\}) + P^*(\{(6,5)\}) \\
\frac{1}{6} &= P^*(\{(5,6)\}) + P^*(\{(6,6)\})
\end{aligned}$$

There is one linear dependency leaving eight equations in sixteen unknowns.

Indeterminacy, Dice Example, 2 of 2

- The fact that for $P^*(\{(x, \lambda)\})$ we have only eight equations in sixteen unknowns is fatal.

- We have no logical basis for choosing a particular solution.

- The particle filter depends on the choice of solution.

A Second Example, Mimics Fisher (1930), 1 of 2

$$
\begin{aligned}
P[Z(X,\Lambda)=z] &= \frac{1-p}{1+p}p^{|z|} \\
Z(X,\Lambda) &= X-\Lambda \\
X &\in \mathbb{N} \\
\Lambda &\in \mathbb{N} \\
\mathbb{N} &= \{0, \pm 1, \pm 2, \ldots\}
\end{aligned}
$$

- The preimages of $Z(x,\lambda)$ are

$$
C_z = \{(x,\lambda) : x = z+\lambda,\ \lambda \in \mathbb{N}\} \qquad z \in \mathbb{N}
$$

  which lie on 45 degree lines in the (x,λ) plane.

- Given λ, for every $z \in \mathbb{N}$ there is an $x \in \mathbb{N}$ with $(x,\lambda) \in C_z$ so every C_z can occur. Therefore $O_\lambda = \cup_{z\in\mathbb{N}} C_z$ and $P(O_\lambda) = 1$ for every $\lambda \in \mathbb{N}$.

A Second Example, Mimics Fisher (1930), 2 of 2

- If $P(O_\lambda) = 1$ for every $\lambda \in \mathbb{N}$.

- Then
$$P(Z = z \,|\, \Lambda = \lambda) = \frac{P(C_z \cap O_\lambda)}{P(O_\lambda)} = P(C_z) = \frac{1-p}{1+p}\, p^{|z|},$$
which does not depend on λ.

- Consequently,
$$P(X = x \,|\, \Lambda = \lambda) = P(Z = x - \lambda)$$

- Provides a rationale for choosing a solution: The conditional robability of X given Λ should be the same under P^*_θ and P_θ.
$$\begin{aligned} P^*(X = x, \Lambda = \lambda) &= P(Z = x - \lambda)\, P^*(R_\lambda) \\ P^*(X = x \,|\, \Lambda = \lambda) &= P(Z = x - \lambda). \end{aligned}$$

Abstraction, 1 of 3

- A GMM criterion $Z(X, \Lambda, \theta)$
- And knowledge of $P^o(R_B)$
  - $R_B = \mathbb{R}^{\text{dim}(X)} \times B$
  - $B \in \mathbb{R}^{\text{dim}(\Lambda)}$ is Borel
- Defines a probability space

$$(\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}^*, P^*_\theta)$$

  - $\mathcal{C}^*$ is the smallest σ-algebra containing the preimages of Z and rectangles R_B,
- On which there are notions of joint $p^*(X, \Lambda, \theta)$, conditional $p^*(X \,|\, \Lambda, \theta)$, and marginal densities $p^*(\Lambda)$.

Abstraction, 2 of 3

- $(\mathbb{R}^{\mathsf{dim}(X)} \times \mathbb{R}^{\mathsf{dim}(\Lambda)}, \mathcal{C}, P_\theta) = (\mathbb{R}^{\mathsf{dim}(X)} \times \mathbb{R}^{\mathsf{dim}(\Lambda)}, \mathcal{C}, P_\theta^o)$

- $(\mathbb{R}^{\mathsf{dim}(X)} \times \mathbb{R}^{\mathsf{dim}(\Lambda)}, \mathcal{C}, P_\theta^*) = (\mathbb{R}^{\mathsf{dim}(X)} \times \mathbb{R}^{\mathsf{dim}(\Lambda)}, \mathcal{C}, P_\theta^o)$

- $(\mathbb{R}^{\mathsf{dim}(X)} \times \mathbb{R}^{\mathsf{dim}(\Lambda)}, \mathcal{C}^*, P_\theta^*) = (\mathbb{R}^{\mathsf{dim}(X)} \times \mathbb{R}^{\mathsf{dim}(\Lambda)}, \mathcal{C}^*, P_\theta^o)$

Abstraction, 3 of 3

- If we assume that the union O_λ of all sets in $\mathcal{C}$ that can occur if $\Lambda = \lambda$ is known to have occurred has probability one, then

$$\begin{aligned} p^*(X, \Lambda, \theta) &= p(X, \Lambda, \theta) p^*(\Lambda, \theta) \\ p^*(X \,|\, \Lambda, \theta) &= p(X, \Lambda, \theta) \end{aligned}$$

- And we can recover

$$\int f(x, \lambda)\, p(\lambda \,|\, X, \theta)\, d\lambda$$

via a particle filter as long as we restrict attention to $\mathcal{C}$-measurable f.

Interpretation

- If we assume compact Θ, then

- Chernozukov-Hong are Bayes on

$$(\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}, P_\theta) = (\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}, P_\theta^o)$$

- And we are Bayes on

$$(\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}^*, P_\theta^*) = (\mathbb{R}^{\text{dim}(X)} \times \mathbb{R}^{\text{dim}(\Lambda)}, \mathcal{C}^*, P_\theta^o)$$

Contribution

- The contribution of GMM (Hansen and Singleton, 1982) was to allow frequentist inference regarding the parameters of a nonlinear structural model without having to solve the model.
  - Provided there were no latent variables.

- The contribution of this paper is the same.
  - With latent variables.