GMM with latent variables

Raffaella Giacomini (UCL/Cemmap/CEPR)
Ron Gallant (Penn State) Giuseppe Ragusa (Luiss)

Giannini conference, 26/3/14

Contribution

- Frequentist inference in models defined by nonlinear moment conditions that depend on dynamic latent variables (time-varying parameters, structural shocks, factors)

Contribution

- Frequentist inference in models defined by nonlinear moment conditions that depend on dynamic latent variables (time-varying parameters, structural shocks, factors)
- Uses MCMC methods $\longrightarrow$ Bayesian inference is a trivial extension

Contribution

- Frequentist inference in models defined by nonlinear moment conditions that depend on dynamic latent variables (time-varying parameters, structural shocks, factors)
- Uses MCMC methods $\longrightarrow$ Bayesian inference is a trivial extension
- It's like Hansen and Singleton (1982) with latent variables

Contribution

- Frequentist inference in models defined by nonlinear moment conditions that depend on dynamic latent variables (time-varying parameters, structural shocks, factors)
- Uses MCMC methods $\longrightarrow$ Bayesian inference is a trivial extension
- It's like Hansen and Singleton (1982) with latent variables
- Two main applications:

Contribution

- Frequentist inference in models defined by nonlinear moment conditions that depend on dynamic latent variables (time-varying parameters, structural shocks, factors)
- Uses MCMC methods $\longrightarrow$ Bayesian inference is a trivial extension
- It's like Hansen and Singleton (1982) with latent variables
- Two main applications:
  1. Moment condition models with time-varying parameters

Contribution

- Frequentist inference in models defined by nonlinear moment conditions that depend on dynamic latent variables (time-varying parameters, structural shocks, factors)
- Uses MCMC methods $\longrightarrow$ Bayesian inference is a trivial extension
- It's like Hansen and Singleton (1982) with latent variables
- Two main applications:
  1. Moment condition models with time-varying parameters
  2. Estimating Dynamic Stochastic General Equilibrium models without solving the model

Do we really need a new estimation method?

- Dynamic latent variables typically handled with state-space methods

Do we really need a new estimation method?

- Dynamic latent variables typically handled with state-space methods
- Problem: they assume that the model defines a likelihood. However:

Do we really need a new estimation method?

- Dynamic latent variables typically handled with state-space methods
- Problem: they assume that the model defines a likelihood. However:
- Moment condition models with time-varying parameters

Do we really need a new estimation method?

- Dynamic latent variables typically handled with state-space methods
- Problem: they assume that the model defines a likelihood. However:
- Moment condition models with time-varying parameters
    - Limited information so no likelihood without auxiliary assumptions

Do we really need a new estimation method?

- Dynamic latent variables typically handled with state-space methods
- Problem: they assume that the model defines a likelihood. However:
- Moment condition models with time-varying parameters
    - Limited information so no likelihood without auxiliary assumptions
    - No estimation method currently exists

Do we really need a new estimation method?

- DSGE models

Do we really need a new estimation method?

- DSGE models
  - In principle can write a likelihood by first solving the model

Do we really need a new estimation method?

- DSGE models
  - In principle can write a likelihood by first solving the model
  - In practice this involves approximation

Do we really need a new estimation method?

- DSGE models
  - In principle can write a likelihood by first solving the model
  - In practice this involves approximation
    - Numerical approximation $\rightarrow$ only handle small models

Do we really need a new estimation method?

- DSGE models
  - In principle can write a likelihood by first solving the model
  - In practice this involves approximation
    - Numerical approximation → only handle small models
    - Linearization (majority of literature) or higher-order Taylor expansions (small literature) → effect of approximation on inference?

Do we really need a new estimation method?

- DSGE models
  - In principle can write a likelihood by first solving the model
  - In practice this involves approximation
    - Numerical approximation → only handle small models
    - Linearization (majority of literature) or higher-order Taylor expansions (small literature) → effect of approximation on inference?
  - Also need to deal with stochastic singularity and multiplicity of solutions

Do we really need a new estimation method?

- DSGE models
  - In principle can write a likelihood by first solving the model
  - In practice this involves approximation
    - Numerical approximation $\longrightarrow$ only handle small models
    - Linearization (majority of literature) or higher-order Taylor expansions (small literature) $\longrightarrow$ effect of approximation on inference?
  - Also need to deal with stochastic singularity and multiplicity of solutions
  - Desirable to have estimation methods that only use the information in equilibrium conditions

Basic idea

- Construct an approximate density based on the moment condition

Basic idea

- Construct an approximate density based on the moment condition
  - It's like Chernozhukov and Hong (2003) with latent variables

Basic idea

- Construct an approximate density based on the moment condition
  - It's like Chernozhukov and Hong (2003) with latent variables
- Apply a nonlinear filtering method to handle the latent variables

Basic idea

- Construct an approximate density based on the moment condition
  - It's like Chernozhukov and Hong (2003) with latent variables
- Apply a nonlinear filtering method to handle the latent variables
- Show that the use of an approximate density does not matter asymptotically

Assumption 1 - the model

- There exists a dynamic structural model but we have incomplete information in the form of m moment conditions

$$E\left[g\left(X_{t+1}, \Lambda_{t+1}, \theta_0\right) \middle| I_t\right] = 0$$

Assumption 1 - the model

- There exists a dynamic structural model but we have incomplete information in the form of m moment conditions

$$E\left[g\left(X_{t+1},\Lambda_{t+1},\theta_0\right)|I_t\right]=0$$

- We observe a sample $X=(X_1,...,X_T)$

Assumption 1 - the model

- There exists a dynamic structural model but we have incomplete information in the form of m moment conditions

$$E\left[g\left(X_{t+1}, \Lambda_{t+1}, \theta_0\right) | I_t\right] = 0$$

- We observe a sample $X = (X_1, ..., X_T)$
- The remaining variables are latent, $\Lambda = (\Lambda_1, ..., \Lambda_T)$

Assumption 2 - the latent variables

- Assume we can draw from the transition density of the dynamic latent variables

$$\Lambda_{t+1} \sim P(\Lambda_{t+1}|\Lambda_t, \theta_0)$$

where P is known and assumed to be ergodic

Assumption 2 - the latent variables

- Assume we can draw from the transition density of the dynamic latent variables

$$\Lambda_{t+1} \sim P(\Lambda_{t+1}|\Lambda_t, \theta_0)$$

  where P is known and assumed to be ergodic
  - Note that θ_0 includes structural parameters of the model and parameters of the law of motion of latent variables (more on this later)

Assumption 2 - the latent variables

- Assume we can draw from the transition density of the dynamic latent variables
$$\Lambda_{t+1} \sim P(\Lambda_{t+1}|\Lambda_t, \theta_0)$$
where P is known and assumed to be ergodic
  - Note that θ_0 includes structural parameters of the model and parameters of the law of motion of latent variables (more on this later)
- Different notion of latent variables than in microeconometrics

Examples of latent variables

- Time-varying parameters, structural shocks, dynamic factors: $\Lambda_{t+1} = \Phi\Lambda_t + \varepsilon_{t+1}$, $\varepsilon_{t+1} \sim iidN(0, \Sigma)$

Examples of latent variables

- Time-varying parameters, structural shocks, dynamic factors: $\Lambda_{t+1} = \Phi\Lambda_t + \varepsilon_{t+1}$, $\varepsilon_{t+1} \sim iidN(0, \Sigma)$
- Λ_{t+1} could also contain endogenous latent variables (e.g., capital stock in RBC models, $k_{t+1} = (1-\delta)k_t + i_t$, i_t observable), so that $\Lambda_{t+1} \sim P(\Lambda_{t+1}|\Lambda_t, X_t, \theta_0)$

Examples of latent variables

- Time-varying parameters, structural shocks, dynamic factors: $\Lambda_{t+1} = \Phi\Lambda_t + \varepsilon_{t+1}$, $\varepsilon_{t+1} \sim iidN(0, \Sigma)$
- Λ_{t+1} could also contain endogenous latent variables (e.g., capital stock in RBC models, $k_{t+1} = (1-\delta)k_t + i_t$, i_t observable), so that $\Lambda_{t+1} \sim P(\Lambda_{t+1}|\Lambda_t, X_t, \theta_0)$
- This is different from the usual distinction between state and control variables when solving a DSGE model

Assumption 3 - identification

- Assume moment conditions would identify θ if Λ_t were observable (standard conditions for identification in GMM)

Assumption 3 - identification

- Assume moment conditions would identify θ if Λ_t were observable (standard conditions for identification in GMM)
- Add moment conditions that identify parameters of latent variables' transition density to the model's equilibrium conditions

Assumption 3 - identification

- Assume moment conditions would identify θ if Λ_t were observable (standard conditions for identification in GMM)
- Add moment conditions that identify parameters of latent variables' transition density to the model's equilibrium conditions
- Identification problems in DSGE models even if likelihood known $\longrightarrow$ standard procedure calibrates some parameters and estimates the rest (discuss later in example)

Assumption 3 - identification

- Assume moment conditions would identify θ if Λ_t were observable (standard conditions for identification in GMM)
- Add moment conditions that identify parameters of latent variables' transition density to the model's equilibrium conditions
- Identification problems in DSGE models even if likelihood known $\longrightarrow$ standard procedure calibrates some parameters and estimates the rest (discuss later in example)
  - Identification problems could be due to linearization $\longrightarrow$ we might be better off

Assumption 3 - identification

- Assume moment conditions would identify θ if Λ_t were observable (standard conditions for identification in GMM)
- Add moment conditions that identify parameters of latent variables' transition density to the model's equilibrium conditions
- Identification problems in DSGE models even if likelihood known $\longrightarrow$ standard procedure calibrates some parameters and estimates the rest (discuss later in example)
  - Identification problems could be due to linearization $\longrightarrow$ we might be better off
  - Identification harder to discuss in nonlinear models + weak identification due to latent variables $\longrightarrow$ we might be worse off

Assumption 4 - asymptotic normality of sample moment

- DSGE model implies conditional moments $\longrightarrow$ transform into unconditional moments

Assumption 4 - asymptotic normality of sample moment

- DSGE model implies conditional moments $\longrightarrow$ transform into unconditional moments
- In practice: choice of moments matters (example later)

Assumption 4 - asymptotic normality of sample moment

- DSGE model implies conditional moments $\longrightarrow$ transform into unconditional moments
- In practice: choice of moments matters (example later)
- Primitive assumption: sample moment condition asymptotically normal

$$Z_T = \left[\Sigma\left(X, \Lambda, \theta_0\right)\right]^{-1/2} g_T\left(X, \Lambda, \theta_0\right) \longrightarrow^d N(0, I)$$

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

$\Sigma\left(X, \Lambda, \theta\right)$ asymptotic variance (maybe HAC)

Assumption 5 - Chernozhukov and Hong (2003)

- Consider the approximate density induced by GMM

$$p(X, \Lambda, \theta) =$$
$$(2\pi)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2} g_T\left(X, \Lambda, \theta\right) \Sigma\left(X, \Lambda, \theta\right)^{-1} g_T\left(X, \Lambda, \theta\right)\right\}$$

Assumption 5 - Chernozhukov and Hong (2003)

- Consider the approximate density induced by GMM

$$p(X,\Lambda,\theta) =$$
$$(2\pi)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2} g_T\left(X,\Lambda,\theta\right)\Sigma\left(X,\Lambda,\theta\right)^{-1} g_T\left(X,\Lambda,\theta\right)\right\}$$

- If Λ were observable, the Chernozhukov and Hong (2003) result would hold:

Assumption 5 - Chernozhukov and Hong (2003)

- Consider the approximate density induced by GMM

$$p(X, \Lambda, \theta) =$$
$$(2\pi)^{-\frac{m}{2}} \exp\left\{ -\frac{1}{2} g_T \left(X, \Lambda, \theta\right) \Sigma \left(X, \Lambda, \theta\right)^{-1} g_T \left(X, \Lambda, \theta\right) \right\}$$

- If Λ were observable, the Chernozhukov and Hong (2003) result would hold:
  - equivalent to estimate θ by GMM or to draw from $p(X, \Lambda, \theta)$ using MCMC methods

The estimation method

- Numerical method that samples $\left\{\theta^{(i)}, \Lambda^{(i)}\right\}_{i=1}^{R}$ by combining

The estimation method

- Numerical method that samples $\left\{\theta^{(i)}, \Lambda^{(i)}\right\}_{i=1}^{R}$ by combining
  - Step 1. Modified particle filter $\longrightarrow$ draw Λ given θ and X and previous draw of Λ

The estimation method

- Numerical method that samples $\left\{\theta^{(i)}, \Lambda^{(i)}\right\}_{i=1}^{R}$ by combining
  - Step 1. Modified particle filter $\longrightarrow$ draw Λ given θ and X and previous draw of Λ
  - Step 2. Metropolis $\longrightarrow$ draw θ given Λ and X and previous θ

The estimation method

- Numerical method that samples $\left\{\theta^{(i)}, \Lambda^{(i)}\right\}_{i=1}^{R}$ by combining
  - Step 1. Modified particle filter $\longrightarrow$ draw Λ given θ and X and previous draw of Λ
  - Step 2. Metropolis $\longrightarrow$ draw θ given Λ and X and previous θ
  - Iterate

The estimation method

- Numerical method that samples $\left\{\theta^{(i)}, \Lambda^{(i)}\right\}_{i=1}^{R}$ by combining
    - Step 1. Modified particle filter $\longrightarrow$ draw Λ given θ and X and previous draw of Λ
    - Step 2. Metropolis $\longrightarrow$ draw θ given Λ and X and previous θ
    - Iterate
- What's new here: at both steps, use GMM density $p(X, \Lambda, \theta)$ instead of true density

The estimation method

- Numerical method that samples $\left\{\theta^{(i)}, \Lambda^{(i)}\right\}_{i=1}^{R}$ by combining
  - Step 1. Modified particle filter $\rightarrow$ draw Λ given θ and X and previous draw of Λ
  - Step 2. Metropolis $\rightarrow$ draw θ given Λ and X and previous θ
  - Iterate
- What's new here: at both steps, use GMM density $p(X, \Lambda, \theta)$ instead of true density
- Technical challenge: show that it doesn't matter asymptotically

Theoretical properties

- Note that the method uses $p(X, \Lambda, \theta)$, even though in principle we need $p(\Lambda | X, \theta)$ for the particle filter and $p(\theta | \Lambda, X)$ for the Metropolis

Theoretical properties

- Note that the method uses $p\left(X, \Lambda, \theta\right)$, even though in principle we need $p\left(\Lambda | X, \theta\right)$ for the particle filter and $p\left(\theta | \Lambda, X\right)$ for the Metropolis
- This is because $p\left(\Lambda | X, \theta\right), p\left(\theta | \Lambda, X\right) \propto p\left(X, \Lambda, \theta\right)$ and the proportionality constant

Theoretical properties

- Note that the method uses $p(X, \Lambda, \theta)$, even though in principle we need $p(\Lambda|X, \theta)$ for the particle filter and $p(\theta|\Lambda, X)$ for the Metropolis
- This is because $p(\Lambda|X, \theta), p(\theta|\Lambda, X) \propto p(X, \Lambda, \theta)$ and the proportionality constant
    - doesn't matter in Metropolis because it cancels out in the acceptance prob $\alpha = \min\left[1, \frac{p(X,\Lambda,\theta_{new})T(\theta_{new},\theta_{old})}{p(X,\Lambda,\theta_{old})T(\theta_{old},\theta_{new})}\right]$

Theoretical properties

- Note that the method uses $p(X, \Lambda, \theta)$, even though in principle we need $p(\Lambda | X, \theta)$ for the particle filter and $p(\theta | \Lambda, X)$ for the Metropolis
- This is because $p(\Lambda | X, \theta), p(\theta | \Lambda, X) \propto p(X, \Lambda, \theta)$ and the proportionality constant
    - doesn't matter in Metropolis because it cancels out in the acceptance prob $\alpha = \min\left[1, \frac{p(X,\Lambda,\theta_{new})T(\theta_{new},\theta_{old})}{p(X,\Lambda,\theta_{old})T(\theta_{old},\theta_{new})}\right]$
    - is assumed to equal 1 in particle filter

Theoretical properties

- Note that the method uses $p(X, \Lambda, \theta)$, even though in principle we need $p(\Lambda | X, \theta)$ for the particle filter and $p(\theta | \Lambda, X)$ for the Metropolis
- This is because $p(\Lambda | X, \theta), p(\theta | \Lambda, X) \propto p(X, \Lambda, \theta)$ and the proportionality constant
  - doesn't matter in Metropolis because it cancels out in the acceptance prob $\alpha = \min\left[1, \frac{p(X,\Lambda,\theta_{new}) T(\theta_{new},\theta_{old})}{p(X,\Lambda,\theta_{old}) T(\theta_{old},\theta_{new})}\right]$
  - is assumed to equal 1 in particle filter
    - this is generally satisfied (primitive condition: $g(\cdot)$ unbounded wrt X) but, if not, one could compute it

Theoretical properties - key idea

- We show that $p(\Lambda|X,\theta)$ assigns the same probability as $f(\Lambda|X,\theta)$ to the pairs of (X,Λ) that are made possible by the moment condition

Theoretical properties - key idea

- We show that $p(\Lambda|X,\theta)$ assigns the same probability as $f(\Lambda|X,\theta)$ to the pairs of (X,Λ) that are made possible by the moment condition
- Intuition from Gallant and Hong (2007), using arguments from fiducial probability (Fisher, 1930)

Theoretical properties - key idea

- We show that $p(\Lambda|X,\theta)$ assigns the same probability as $f(\Lambda|X,\theta)$ to the pairs of (X,Λ) that are made possible by the moment condition
- Intuition from Gallant and Hong (2007), using arguments from fiducial probability (Fisher, 1930)
- Consider simple example where draw scalar Λ from $N(0,1)$ then draw $X_1, ..., X_T$ from $N(\Lambda,1)$

Theoretical properties - key idea

- We show that $p\left(\Lambda|X,\theta\right)$ assigns the same probability as $f\left(\Lambda|X,\theta\right)$ to the pairs of $\left(X,\Lambda\right)$ that are made possible by the moment condition
- Intuition from Gallant and Hong (2007), using arguments from fiducial probability (Fisher, 1930)
- Consider simple example where draw scalar Λ from $N(0,1)$ then draw $X_1, ..., X_T$ from $N(\Lambda, 1)$
- The approximate density induced by the moment condition $E\left[X-\Lambda\right]=0$ is

$$p\left(X,\Lambda,\theta\right)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}Z_T^2\right\}=$$

where $Z_T=\sqrt{T}\left(\overline{X}-\Lambda\right)\approx N(0,1)$

Theoretical properties - key idea

- We show that $p\left(\Lambda|X,\theta\right)$ assigns the same probability as $f\left(\Lambda|X,\theta\right)$ to the pairs of (X,Λ) that are made possible by the moment condition
- Intuition from Gallant and Hong (2007), using arguments from fiducial probability (Fisher, 1930)
- Consider simple example where draw scalar Λ from $N(0,1)$ then draw $X_1, ..., X_T$ from $N(\Lambda, 1)$
- The approximate density induced by the moment condition $E\left[X-\Lambda\right]=0$ is

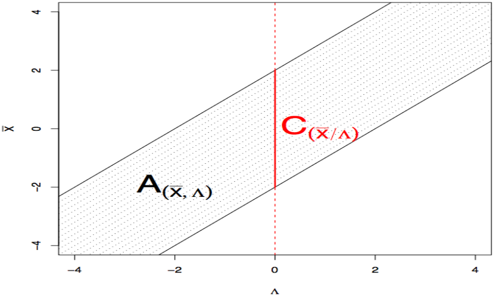
$$p\left(X,\Lambda,\theta\right)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}Z_T^2\right\}=$$

where $Z_T=\sqrt{T}\left(\overline{X}-\Lambda\right)\approx N(0,1)$
- Suppose Z_T exactly normal

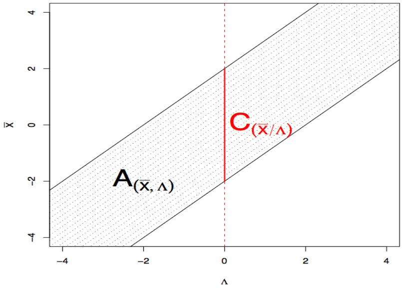
Theoretical properties - key idea

- Exact density $f(X, \Lambda, \theta)$ assigns probability to rectangles (X, Λ), whereas $p(X, \Lambda, \theta)$ assigns probability to sets of the form

$$A = \left\{ \left(\overline{X}, \Lambda\right) : Z_T \in (a, b) \right\} = \left\{ \left(\overline{X}, \Lambda\right) : \frac{a}{\sqrt{T}} + \Lambda < \overline{X} < \frac{b}{\sqrt{T}} + \Lambda) \right\}$$

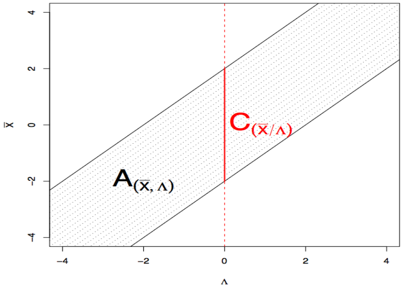


Theoretical properties - key idea



- By the change of measure formula, $p(X, \Lambda, \theta)$ also assigns probability to sets $A^T = \{(X_1, ..., X_T, \Lambda) : Z_T \in (a, b)\}$

Theoretical properties - key idea



- By the change of measure formula, $p(X, \Lambda, \theta)$ also assigns probability to sets $A^T = \{(X_1, ..., X_T, \Lambda) : Z_T \in (a, b)\}$
- Similarly, $p(X|\Lambda, \theta)$ assigns probability to sets of the form C

Theoretical properties - key idea

- Implications of the above:

Theoretical properties - key idea

- Implications of the above:
  - information is obviously lost relative to the true density **but**

Theoretical properties - key idea

- Implications of the above:
  - information is obviously lost relative to the true density **but**
  - for sets of the form A, A^T, C the probability assigned by $p(\cdot)$ is the same as the probability assigned by the true density $f(\cdot)$

Theoretical properties - key idea

- Implications of the above:
  - information is obviously lost relative to the true density **but**
  - for sets of the form A, A^T, C the probability assigned by $p(\cdot)$ is the same as the probability assigned by the true density $f(\cdot)$
- Our main result uses this intuition + asymptotic normality of Z_T

The method in practice - particle filter

- Goal: Draw Λ given θ and X and previous draw of Λ

The method in practice - particle filter

- Goal: Draw Λ given θ and X and previous draw of Λ
- Use Andrieu, Douced and Holenstein (2010): "modified particle filter"

The method in practice - particle filter

- Goal: Draw Λ given θ and X and previous draw of Λ
- Use Andrieu, Douced and Holenstein (2010): "modified particle filter"
- STEP1: Use particle filter to obtain first "particle" $\Lambda_{1:T}^{(1)}$

The modified particle filter

- STEP2: Initialization

The modified particle filter

- STEP2: Initialization
  - Set $1 \leq T_0 \leq T$ to min sample required to compute sample moment

The modified particle filter

- STEP2: Initialization
  - Set $1 \leq T_0 \leq T$ to min sample required to compute sample moment
  - For $i = 2, ..., N$ sample particles $\left(\Lambda_{1:T_0}^{(i)}\right)$ from transition density $p\left(\Lambda_t | \Lambda_{t-1}, \theta\right)$

The modified particle filter

- STEP2: Initialization
  - Set $1 \leq T_0 \leq T$ to min sample required to compute sample moment
  - For $i = 2, ..., N$ sample particles $\left(\Lambda_{1:T_0}^{(i)}\right)$ from transition density $p\left(\Lambda_t | \Lambda_{t-1}, \theta\right)$
  - Set $t = T_0 + 1$

The modified particle filter

- STEP 3: Importance sampling

The modified particle filter

- STEP 3: Importance sampling
  - For $i = 2, ..., N$, sample next observation $\Lambda_t^{(i)}$ from $p\left(\Lambda_t | \Lambda_{t-1}^{(i)}, \theta\right)$

The modified particle filter

- STEP 3: Importance sampling
  - For $i = 2, ..., N$, sample next observation $\Lambda_t^{(i)}$ from $p\left(\Lambda_t | \Lambda_{t-1}^{(i)}, \theta\right)$
  - For $i = 1, ..., N$, compute weights using the **GMM density** $w_t^{(i)} = p\left(X_{1:t}, \Lambda_{1:t}^{(i)}, \theta\right)$

The modified particle filter

- STEP 3: Importance sampling
  - For $i = 2, ..., N$, sample next observation $\Lambda_t^{(i)}$ from $p\left(\Lambda_t | \Lambda_{t-1}^{(i)}, \theta\right)$
  - For $i = 1, ..., N$, compute weights using the **GMM density** $w_t^{(i)} = p\left(X_{1:t}, \Lambda_{1:t}^{(i)}, \theta\right)$
  - Sample with replacement the particles from $\left\{\Lambda_{1:t}^{(i)}\right\}_{i=1}^{N}$ according to the weights

The modified particle filter

- STEP 3: Importance sampling
  - For $i = 2, ..., N$, sample next observation $\Lambda_t^{(i)}$ from $p\left(\Lambda_t | \Lambda_{t-1}^{(i)}, \theta\right)$
  - For $i = 1, ..., N$, compute weights using the **GMM density** $w_t^{(i)} = p\left(X_{1:t}, \Lambda_{1:t}^{(i)}, \theta\right)$
  - Sample with replacement the particles from $\left\{\Lambda_{1:t}^{(i)}\right\}_{i=1}^{N}$ according to the weights
  - Increase t and repeat until reaching T

The modified particle filter

- STEP 3: Importance sampling
  - For $i = 2, ..., N$, sample next observation $\Lambda_t^{(i)}$ from $p\left(\Lambda_t | \Lambda_{t-1}^{(i)}, \theta\right)$
  - For $i = 1, ..., N$, compute weights using the **GMM density** $w_t^{(i)} = p\left(X_{1:t}, \Lambda_{1:t}^{(i)}, \theta\right)$
  - Sample with replacement the particles from $\left\{\Lambda_{1:t}^{(i)}\right\}_{i=1}^{N}$ according to the weights
  - Increase t and repeat until reaching T
  - At T, output the particle $\Lambda_{1:T}^{(N)}$

Intuition

- For a fixed θ and X, the algorithm generates sequence of latent variables "most compatible" with the moment conditions

Metropolis

- Sample $\theta^{(i)}$ from $p(\theta|X, \Lambda^{(i-1)})$ knowing $\theta^{(i-1)}$

Metropolis

- Sample $\theta^{(i)}$ from $p(\theta|X, \Lambda^{(i-1)})$ knowing $\theta^{(i-1)}$
  - Set $\theta_{old} = \theta^{(i-1)}$

Metropolis

- Sample $\theta^{(i)}$ from $p(\theta|X, \Lambda^{(i-1)})$ knowing $\theta^{(i-1)}$
  - Set $\theta_{old} = \theta^{(i-1)}$
  - Propose: draw θ_{new} given θ_{old} using a proposal density $T(\theta_{old}, \theta_{new})$ (e.g., random walk)

Metropolis

- Sample $\theta^{(i)}$ from $p(\theta|X, \Lambda^{(i-1)})$ knowing $\theta^{(i-1)}$
  - Set $\theta_{old} = \theta^{(i-1)}$
  - Propose: draw θ_{new} given θ_{old} using a proposal density $T(\theta_{old}, \theta_{new})$ (e.g., random walk)
  - Accept θ_{new} with probability that depends on the **GMM density**

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

otherwise keep θ_{old}

Metropolis

- Sample $\theta^{(i)}$ from $p(\theta|X, \Lambda^{(i-1)})$ knowing $\theta^{(i-1)}$
  - Set $\theta_{old} = \theta^{(i-1)}$
  - Propose: draw θ_{new} given θ_{old} using a proposal density $T(\theta_{old}, \theta_{new})$ (e.g., random walk)
  - Accept θ_{new} with probability that depends on the **GMM density**

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

  otherwise keep θ_{old}
  - Iterate K times and set $\theta^{(i)}$ = last value of the chain

Why this works

- Theorem 1: The particle filter works because the method generates draws from the true conditional density (for large T)

$$f\left(\Lambda|X,\theta_0\right)$$

Why this works

- Theorem 1: The particle filter works because the method generates draws from the true conditional density (for large T)

$$f\left(\Lambda | X, \theta_0\right)$$

  - BUT, only for pairs of Λ, X that are "allowed" by the structural model

Why this works

- Theorem 1: The particle filter works because the method generates draws from the true conditional density (for large T)

$$f\left(\Lambda | X, \theta_0\right)$$

  - BUT, only for pairs of Λ, X that are "allowed" by the structural model

- The Metropolis works because, once conditioning on Λ, it's Chernozhukov and Hong (2003)

Why this works

- Theorem 1: The particle filter works because the method generates draws from the true conditional density (for large T)

$$f\left(\Lambda | X, \theta_0\right)$$

  - BUT, only for pairs of Λ, X that are "allowed" by the structural model
- The Metropolis works because, once conditioning on Λ, it's Chernozhukov and Hong (2003)
- That iterating the particle filter and Metropolis works follows from Andrieu, Douced and Holenstein (2010)

Does the method work in practice?

- Two examples

Does the method work in practice?

- Two examples
  - Stochastic volatility

Does the method work in practice?

- Two examples
  - Stochastic volatility
  - A simplified DSGE

Does the method work in practice?

- Two examples
  - Stochastic volatility
  - A simplified DSGE
- In both cases we have a likelihood $\longrightarrow$ see what we lose by using GMM

Stochastic volatility example

- Data-generating process

$$
\begin{aligned}
X_t &= \rho X_{t-1} + \exp(\Lambda_t) u_t \\
\Lambda_t &= \phi \Lambda_{t-1} + \sigma e_t \\
e_t, u_t &\sim N(0,1) \text{ independent}
\end{aligned}
$$

Stochastic volatility example

- Data-generating process

$$\begin{aligned} X_t &= \rho X_{t-1} + \exp(\Lambda_t) u_t \\ \Lambda_t &= \phi \Lambda_{t-1} + \sigma e_t \\ e_t, u_t &\sim N(0,1) \text{ independent} \end{aligned}$$

- Best existing method is Flury-Shepherd particle filter

Choice of moment conditions

$$
\begin{aligned}
g_1 &= \left(X_t - \rho X_{t-1}\right) X_{t-1} \\
g_2 &= \left(\Lambda_t - \phi \Lambda_{t-1}\right) \Lambda_{t-1} \\
g_3 &= \left(\Lambda_t - \phi \Lambda_{t-1}\right)^2 - \sigma^2 \\
g_4 &= \left|X_t - \rho X_{t-1}\right| - \left(\frac{2}{\pi}\right)^2 \left[\exp\left(\Lambda_t\right)\right]^2 \\
g_5 &= \left|X_t - \rho X_{t-1}\right| \left|X_{t-1} - \rho X_{t-2}\right| \\
&\quad - \left(\frac{2}{\pi}\right)^2 \exp\left(\Lambda_t\right) \exp\left(\Lambda_{t-1}\right) \\
&\dots \\
g_{4+L} &= \left|X_t - \rho X_{t-1}\right| \left|X_{t-L} - \rho X_{t-L-1}\right| \\
&\quad - \left(\frac{2}{\pi}\right)^2 \exp\left(\Lambda_t\right) \exp\left(\Lambda_{t-L}\right)
\end{aligned}
$$

"Simulation" results

- Estimates of θ from our GMM method and the Flury-Shepherd ML

"Simulation" results

- Estimates of θ from our GMM method and the Flury-Shepherd ML
- Scatter plot of filtered latent variables against the true ones

Table 1. Parameter Estimates for the SV Model Moment Conditions (23) through (28) at both the Metropolis and Gibbs Steps.

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 $T = 250$ was generated by simulating the model of Subsection 6.1 at the parameter values shown in the column labeled "True Value". In the first two panels the model was estimated by using the Metropolis within Gibbs methods described in Section 2 with a one-lag HAC weighting matrix using $N = 1000$ particles for Gibbs and $K = 50$ draws for Metropolis. In the third panel the estimator is the Bayesian estimator proposed by Flury and Shepard (2010) with a flat prior. It is a standard maximum likelihood particle filter estimator except that the seed changes every time a new θ is proposed with N increased as necessary to control the rejection rate of the MCMC chain. The columns labeled mean, mode, and standard deviation are the mean, mode, and standard deviations of a Metropolis within Gibbs chain of length $R = 9637$ for the first two panels and the same from an MCMC chain of length $R = 500000$ with a stride of 5 for the third.

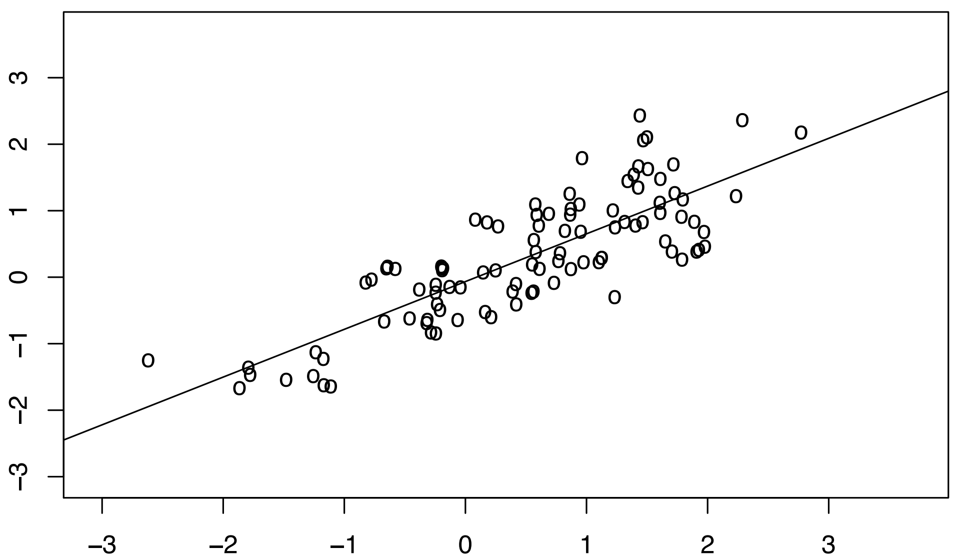


Figure 5. PF for Λ, without Jacobian, Scatter Plot, SV Model. As for Figure 4 except that plotted is the mean of the particles vs. the simulated Λ.

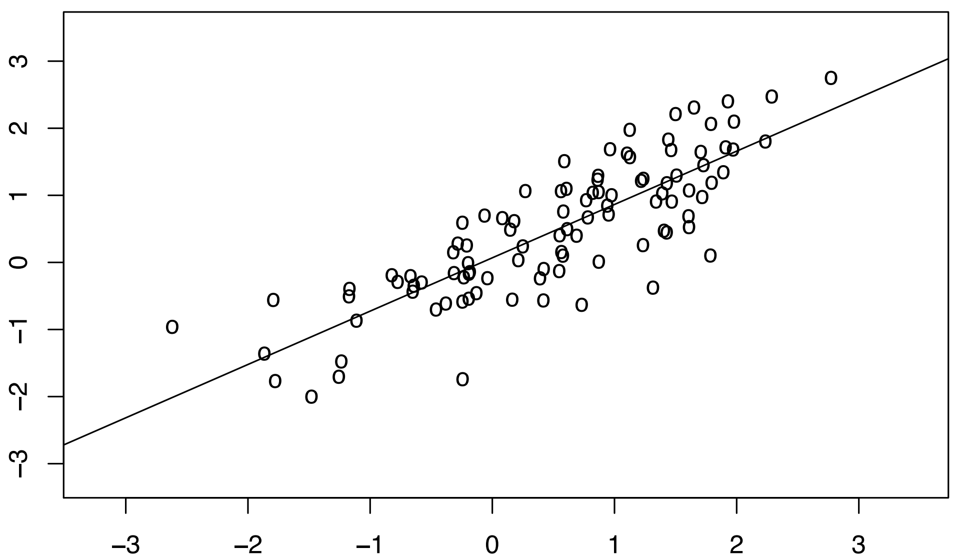


Figure 7. PF for Λ, Flurry-Shephard Method, Scatter Plot. As for Figure 6 except that plotted is the mean of the particles vs. the simulated Λ.

DSGE example

- Simplified version of Del Negro and Schorfheide (2008). First order conditions

$$
\begin{aligned}
E_t\left[y_{t+1}+\pi_{t+1}+z_{t+1}\right]-y_t-\frac{1}{\beta}\pi_t &= 0\\
\lambda_t+w_t &= 0\\
w_t-(1+v)y_t-\phi_t &= 0
\end{aligned}
$$

DSGE example

- Simplified version of Del Negro and Schorfheide (2008). First order conditions

$$
\begin{aligned}
E_t \left[y_{t+1} + \pi_{t+1} + z_{t+1} \right] - y_t - \frac{1}{\beta}\pi_t &= 0 \\
\lambda_t + w_t &= 0 \\
w_t - (1+v) y_t - \phi_t &= 0
\end{aligned}
$$

- Outputs: y_t output, w_t wages, π_t inflation

DSGE example

- Simplified version of Del Negro and Schorfheide (2008). First order conditions

$$
\begin{aligned}
E_t\left[y_{t+1}+\pi_{t+1}+z_{t+1}\right]-y_t-\frac{1}{\beta}\pi_t &= 0\\
\lambda_t+w_t &= 0\\
w_t-(1+v)y_t-\phi_t &= 0
\end{aligned}
$$

- Outputs: y_t output, w_t wages, π_t inflation
- Shocks

$$
\begin{aligned}
z_t &= \rho_z z_{t-1}+\sigma_z\varepsilon_{z,t} \text{ factor productivity}\\
\lambda_t &= \rho_\lambda z_{t-1}+\sigma_\lambda\varepsilon_{\lambda,t} \text{ consumption/leisure preference}\\
\phi_t &= \rho_\phi z_{t-1}+\sigma_\phi\varepsilon_{\phi,t} \text{ price elasticity}
\end{aligned}
$$

DSGE example

- λ_t pinned down by model so we have

$$\begin{aligned} \text{Observable variables } X_t &= (y_t, w_t, \pi_t) \\ \text{Latent variables } \Lambda_t &= (z_t, \phi_t) \end{aligned}$$

DSGE example

- λ_t pinned down by model so we have

$$\begin{aligned} \text{Observable variables } X_t &= (y_t, w_t, \pi_t) \\ \text{Latent variables } \Lambda_t &= (z_t, \phi_t) \end{aligned}$$

- Identification problems: likelihood reveals that only one of $\sigma_z, \sigma_\phi, v, \beta$ can be identified $\longrightarrow$ calibrate σ_z, σ_ϕ, v

DSGE example

- λ_t pinned down by model so we have

$$\begin{aligned}\text{Observable variables } X_t &= (y_t, w_t, \pi_t) \\ \text{Latent variables } \Lambda_t &= (z_t, \phi_t)\end{aligned}$$

- Identification problems: likelihood reveals that only one of $\sigma_z, \sigma_\phi, v, \beta$ can be identified $\longrightarrow$ calibrate σ_z, σ_ϕ, v
- In general, identification problems will cause the MCMC chain not to mix

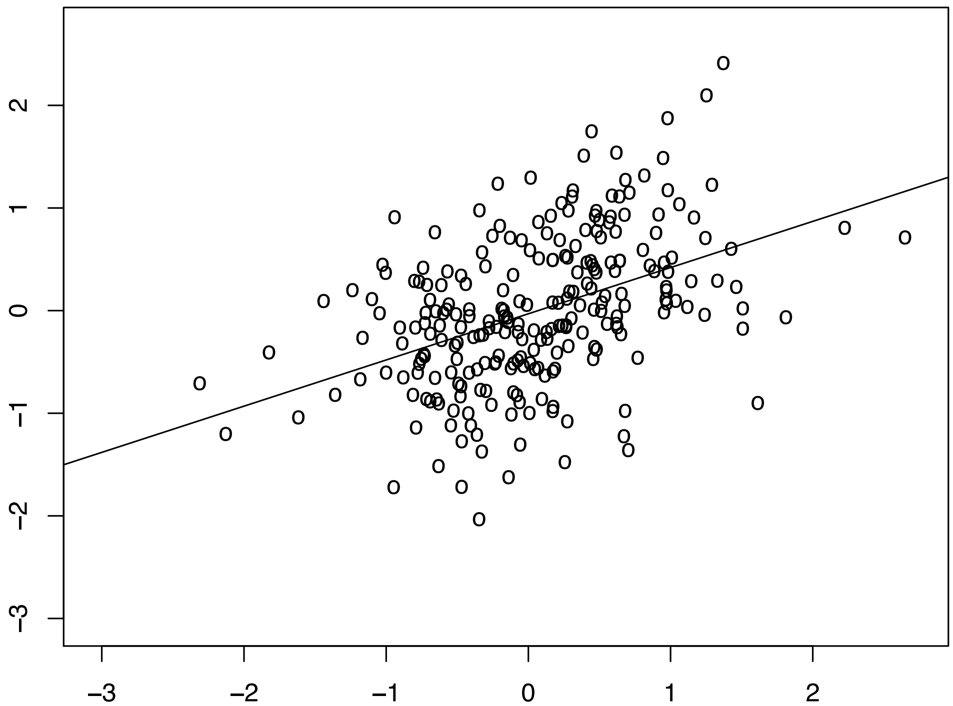
Choice of moment conditions

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

Table 2. Parameter Estimates for the DSGE Model Using Moment Conditions (32) through (40) at Both the Metropolis and Gibbs Steps.

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 of length $T = 250$ was generated by simulating the model of Subsection 6.2 at the parameter values shown in the column labeled "True Value". In the first two panels the model was estimated by using the Metropolis within Gibbs method described in Section 2 with a two-lag HAC weighting matrix using $N = 1000$ particles for Gibbs and $K = 50$ draws for Metropolis. In the third panel the model was estimated by maximum likelihood. The columns labeled mean, mode, and standard deviation are the mean, mode, and standard deviations of a Metropolis within Gibbs chain of length $R = 9637$ for the first two panels and the same from an MCMC chain of length $R = 500000$ with a stride of 5 for the third.



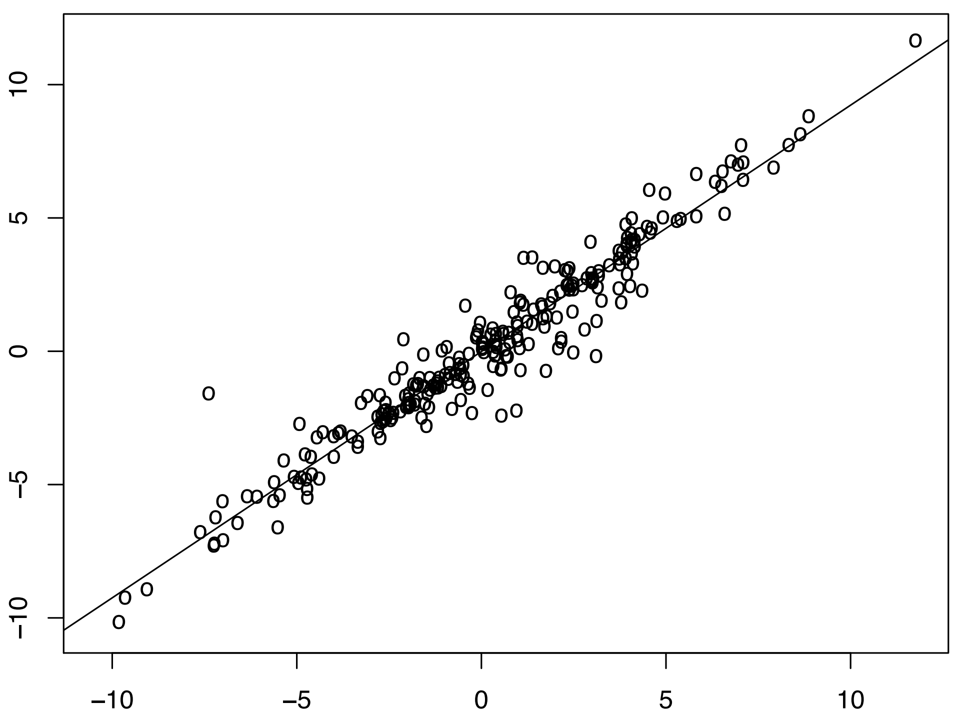


Figure 9. PF for Λ with Jacobian, Scatter Plot, DSGE Model. As for Figure 8 except that plotted is the mean of the particles vs. the simulated Λ for all 250 time points.

Conclusion

- Estimation method for moment-condition models with dynamic latent variables

Conclusion

- Estimation method for moment-condition models with dynamic latent variables
- Implemented by a Metropolis within a particle filter algorithm

Conclusion

- Estimation method for moment-condition models with dynamic latent variables
- Implemented by a Metropolis within a particle filter algorithm
- Works in our limited experience

Conclusion

- Estimation method for moment-condition models with dynamic latent variables
- Implemented by a Metropolis within a particle filter algorithm
- Works in our limited experience
- Must better investigate effect of choice of moment conditions