A Sieve-SMM Estimator for Dynamic Models

Jean-Jacques Forneron

May 17, 2019

Boston University

Introduction

- Interested in a class of Non-Linear Dynamic Models
  - numerically complex to estimate
  - have to simulate from the model to estimate
  - common in many economic settings:
    DSGE, Dynamic Discrete Choice, Asset Pricing, ...

$\Rightarrow$ this approach is called simulation-based estimation

Introduction

- Interested in a class of Non-Linear Dynamic Models
    - numerically complex to estimate
    - have to simulate from the model to estimate
    - common in many economic settings:
      DSGE, Dynamic Discrete Choice, Asset Pricing, ...

  $\Rightarrow$ this approach is called simulation-based estimation
- Caveat: typically make strong assumptions to simulate
    - assume the distribution of the shocks is known

Introduction

- Interested in a class of Non-Linear Dynamic Models
  - numerically complex to estimate
  - have to simulate from the model to estimate
  - common in many economic settings:
    DSGE, Dynamic Discrete Choice, Asset Pricing, ...

  $\Rightarrow$ this approach is called simulation-based estimation
- Caveat: typically make strong assumptions to simulate
  - assume the distribution of the shocks is known
- This paper: estimate parameters of the model and *flexibly approximate the distribution of shocks*
  - learn about the nature of risks rather than assuming it
  - methodological contribution: simple algorithm
  - theoretical contribution: theory with general dynamics

Introduction

- Interested in a class of Non-Linear Dynamic Models
  - numerically complex to estimate
  - have to simulate from the model to estimate
  - common in many economic settings:
    DSGE, Dynamic Discrete Choice, Asset Pricing, ...

  $\Rightarrow$ this approach is called simulation-based estimation
- Caveat: typically make strong assumptions to simulate
  - assume the distribution of the shocks is known
- This paper: estimate parameters of the model and *flexibly approximate the distribution of shocks*
  - learn about the nature of risks rather than assuming it
  - methodological contribution: simple algorithm
  - theoretical contribution: theory with general dynamics

Introduction

- Interested in a class of Non-Linear Dynamic Models
    - numerically complex to estimate
    - have to simulate from the model to estimate
    - common in many economic settings:
      DSGE, Dynamic Discrete Choice, Asset Pricing, ...

  $\Rightarrow$ this approach is called simulation-based estimation
- Caveat: typically make strong assumptions to simulate
    - assume the distribution of the shocks is known
- This paper: estimate parameters of the model and *flexibly approximate the distribution of shocks*
    - learn about the nature of risks rather than assuming it
    - methodological contribution: simple algorithm
    - theoretical contribution: theory with general dynamics

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$y_t = g_{obs}(y_{t-1}, x_t, \theta, f, u_t) \qquad \text{(observed)}$$

$$u_t = g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f \qquad \text{(latent)}$$

x_t strictly exogenous regressors, $t = 1, \dots, n$

g_{obs}, g_{latent} known up to θ finite dimensional and f

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$y_t = g_{obs}(y_{t-1}, x_t, \theta, f, u_t) \qquad \text{(observed)}$$

$$u_t = g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f \qquad \text{(latent)}$$

x_t strictly exogenous regressors, $t = 1, \ldots, n$

g_{obs}, g_{latent} known up to θ finite dimensional and f

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$y_t = g_{obs}(y_{t-1}, x_t, \theta, f, u_t) \qquad \text{(observed)}$$

$$u_t = g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f \qquad \text{(latent)}$$

x_t strictly exogenous regressors, $t = 1, \dots, n$

g_{obs}, g_{latent} known up to θ finite dimensional and f

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$y_t = g_{obs}(y_{t-1}, x_t, \theta, f, u_t) \qquad \text{(observed)}$$

$$u_t = g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f \qquad \text{(latent)}$$

x_t strictly exogenous regressors, $t = 1, \dots, n$

g_{obs}, g_{latent} known up to θ finite dimensional and f

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$y_t = g_{obs}(y_{t-1}, x_t, \theta, f, u_t) \qquad \text{(observed)}$$

$$u_t = g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f \qquad \text{(latent)}$$

x_t strictly exogenous regressors, $t = 1, \dots, n$

g_{obs}, g_{latent} known up to θ finite dimensional and f

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$
\begin{aligned}
y_t &= g_{obs}(y_{t-1}, x_t, \theta, f, u_t) && \text{(observed)} \\
u_t &= g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f && \text{(latent)}
\end{aligned}
$$

  x_t strictly exogenous regressors, $t = 1, \dots, n$

  g_{obs}, g_{latent} known up to θ finite dimensional and f
- Likelihood and moments intractable and hard to compute:
  - MLE and GMM not feasible

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$\begin{aligned} y_t &= g_{obs}(y_{t-1}, x_t, \theta, f, u_t) && \text{(observed)} \\ u_t &= g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f && \text{(latent)} \end{aligned}$$

  x_t strictly exogenous regressors, $t = 1, \dots, n$

  g_{obs}, g_{latent} known up to θ finite dimensional and f

- Likelihood and moments intractable and hard to compute:
  - MLE and GMM not feasible
- Common in applications with e.g.:
  - unobserved latent variables (e.g. productivity, volatility)
  - and/or intractable policy function (e.g. dynamic programming)

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$y_t = g_{obs}(y_{t-1}, x_t, \theta, f, u_t) \qquad \text{(observed)}$$

$$u_t = g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f \qquad \text{(latent)}$$

  x_t strictly exogenous regressors, $t = 1, \dots, n$

  g_{obs}, g_{latent} known up to θ finite dimensional and f

- Likelihood and moments intractable and hard to compute:
  - MLE and GMM not feasible
- Common in applications with e.g.:
  - unobserved latent variables (e.g. productivity, volatility)
  - and/or intractable policy function (e.g. dynamic programming)

Introduction: Intractable Models

- Non-Linear Dynamic Models:

$$y_t = g_{obs}(y_{t-1}, x_t, \theta, f, u_t) \qquad \text{(observed)}$$

$$u_t = g_{latent}(u_{t-1}, \theta, f, e_t), \quad e_t \overset{iid}{\sim} f \qquad \text{(latent)}$$

  x_t strictly exogenous regressors, $t = 1, \dots, n$

  g_{obs}, g_{latent} known up to θ finite dimensional and f

- Likelihood and moments intractable and hard to compute:
  - MLE and GMM not feasible
- Common in applications with e.g.:
  - unobserved latent variables (e.g. productivity, volatility)
  - and/or intractable policy function (e.g. dynamic programming)

Solution: Simulation-Based Estimation

- Summarize observables (y_t, x_t) with vector of moment $\hat{\psi}_n$
- Use a fully parameterized model (assume f known) to
  i. Simulate $S \geq 1$ samples $(y_t^s(\theta), x_t)$
  ii. Compute simulated moments $\hat{\psi}_n^s(\theta)$
  iii. Match sample with average simulated moment $\hat{\psi}_n^S = \sum_s \hat{\psi}_n^s / S$

$$\hat{\theta}_n^{SMM} = \text{argmin}_\theta \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|$$

Solution: Simulation-Based Estimation

- Summarize observables (y_t, x_t) with vector of moment $\hat{\psi}_n$
- Use a fully parameterized model (assume f known) to
  i. Simulate $S \geq 1$ samples $(y_t^s(\theta), x_t)$
  ii. Compute simulated moments $\hat{\psi}_n^s(\theta)$
  iii. Match sample with average simulated moment $\hat{\psi}_n^S = \sum_s \hat{\psi}_n^s / S$

$$\hat{\theta}_n^{SMM} = \mathrm{argmin}_\theta \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|$$

Solution: Simulation-Based Estimation

- Summarize observables (y_t, x_t) with vector of moment $\hat{\psi}_n$
- Use a fully parameterized model (assume f known) to
  i. Simulate $S \geq 1$ samples $(y_t^s(\theta), x_t)$
  ii. Compute simulated moments $\hat{\psi}_n^s(\theta)$
  iii. Match sample with average simulated moment $\hat{\psi}_n^S = \sum_s \hat{\psi}_n^s / S$

$$\hat{\theta}_n^{SMM} = \text{argmin}_\theta \| \hat{\psi}_n - \hat{\psi}_n^S(\theta) \|$$

Solution: Simulation-Based Estimation

- Summarize observables (y_t, x_t) with vector of moment $\hat{\psi}_n$
- Use a fully parameterized model (assume f known) to
    i. Simulate $S \geq 1$ samples $(y_t^s(\theta), x_t)$
    ii. Compute simulated moments $\hat{\psi}_n^s(\theta)$
    iii. Match sample with average simulated moment $\hat{\psi}_n^S = \sum_s \hat{\psi}_n^s / S$

$$\hat{\theta}_n^{SMM} = \text{argmin}_\theta \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|$$

Solution: Simulation-Based Estimation

- Summarize observables (y_t, x_t) with vector of moment $\hat{\psi}_n$
- Use a fully parameterized model (assume f known) to
  i. Simulate $S \geq 1$ samples $(y_t^s(\theta), x_t)$
  ii. Compute simulated moments $\hat{\psi}_n^s(\theta)$
  iii. Match sample with average simulated moment $\hat{\psi}_n^S = \sum_s \hat{\psi}_n^s / S$

$$\hat{\theta}_n^{SMM} = \text{argmin}_\theta \|\hat{\psi}_n - \hat{\psi}_n^S(\theta)\|$$

- Vast existing literature (SMM, Indirect Inference, EMM)
  - Review, see e.g. Gouriéroux and Monfort (1996); Smith (2016)
- Limitations: requires a fully parametric model
  - ▷ but economic models rarely informative about distribution f
  - ▷ so that $\hat{\theta}_n^{SMM}$ may be sensitive to choice of f
  - ▷ and f may also be of interest

Recent Literature

- Recent work allows for more flexible parametric distributions in SMM estimation
- Flexibly approximate the first 3-4 moments to e.g.:
  - Model asymmetric shocks and their impact on business cycles, the yield curve (Ruge-Murcia, 2012, 2017)
  - Restore identification when first 2 moments are not enough (Gospodinov and Ng, 2015; Gospodinov et al., 2017)

Recent Literature

- Recent work allows for more flexible parametric distributions in SMM estimation
- Flexibly approximate the first 3-4 moments to e.g.:
  - Model asymmetric shocks and their impact on business cycles, the yield curve (Ruge-Murcia, 2012, 2017)
  - Restore identification when first 2 moments are not enough (Gospodinov and Ng, 2015; Gospodinov et al., 2017)
- Pushing this idea further:
  - Look for a flexible parametric family which can approximate a large class of distributions
  - Need moments that are informative enough about θ + the entire distribution f

Recent Literature

- Recent work allows for more flexible parametric distributions in SMM estimation
- Flexibly approximate the first 3-4 moments to e.g.:
  - Model asymmetric shocks and their impact on business cycles, the yield curve (Ruge-Murcia, 2012, 2017)
  - Restore identification when first 2 moments are not enough (Gospodinov and Ng, 2015; Gospodinov et al., 2017)
- Pushing this idea further:
  - Look for a flexible parametric family which can approximate a large class of distributions
  - Need moments that are informative enough about θ + the entire distribution f

Recent Literature

- Recent work allows for more flexible parametric distributions in SMM estimation
- Flexibly approximate the first 3-4 moments to e.g.:
  - Model asymmetric shocks and their impact on business cycles, the yield curve (Ruge-Murcia, 2012, 2017)
  - Restore identification when first 2 moments are not enough (Gospodinov and Ng, 2015; Gospodinov et al., 2017)
- Pushing this idea further:
  - Look for a flexible parametric family which can approximate a large class of distributions
  - Need moments that are informative enough about θ + the entire distribution f

This Paper: A Sieve-SMM Estimator

1. A Gaussian Mixture sieve
   - ✓ To approximate smooth densities:

$$f(e) \simeq \sum_{j=1}^{k(n)} \frac{\omega_j}{\sigma_j} \phi\left(\frac{e-\mu_j}{\sigma_j}\right)$$

   - # of components $k(n) \nearrow$ with sample size n

This Paper: A Sieve-SMM Estimator

1. A Gaussian Mixture sieve
   - ✓ To approximate smooth densities:

$$f(e) \simeq \sum_{j=1}^{k(n)} \frac{\omega_j}{\sigma_j} \phi\left(\frac{e-\mu_j}{\sigma_j}\right)$$

   - $\#$ of components $k(n) \nearrow$ with sample size n

This Paper: A Sieve-SMM Estimator

1. A Gaussian Mixture sieve
   - ✓ To approximate smooth densities:

$$f(e) \simeq \sum_{j=1}^{k(n)} \frac{\omega_j}{\sigma_j} \phi\left(\frac{e-\mu_j}{\sigma_j}\right)$$

   - $\#$ of components $k(n) \nearrow$ with sample size n
2. And a Moment Function: Characteristic Function (CF)

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t,\mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$

This Paper: A Sieve-SMM Estimator

1. A Gaussian Mixture sieve
   - ✓ To approximate smooth densities:
   $$f(e) \simeq \sum_{j=1}^{k(n)} \frac{\omega_j}{\sigma_j}\phi\left(\frac{e-\mu_j}{\sigma_j}\right)$$
   - # of components $k(n) \nearrow$ with sample size n
2. And a Moment Function: Characteristic Function (CF)
   $$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t,\mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$
   - Infinite dim. moments: use all $\tau \in \mathbb{R}^{d_\mathbf{y}+d_\mathbf{x}}$

This Paper: A Sieve-SMM Estimator

1. A Gaussian Mixture sieve
   - ✓ To approximate smooth densities:

$$f(e) \simeq \sum_{j=1}^{k(n)} \frac{\omega_j}{\sigma_j} \phi\left(\frac{e-\mu_j}{\sigma_j}\right)$$

   - # of components $k(n) \nearrow$ with sample size n
2. And a Moment Function: Characteristic Function (CF)

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t,\mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$

   - ✓ Infinite dim. moments: identify the infinite dim. $\beta = (\theta, f)$

Main Results

1. Consistency + Rate of convergence + Asym. normality results
    - under low-level conditions on the data generating process
    - assuming $\beta = (\theta, f)$ identified by $\hat{\psi}_n(\cdot)$

Main Results

1. Consistency + Rate of convergence + Asym. normality results
   - under low-level conditions on the data generating process
   - assuming $\beta = (\theta, f)$ identified by $\hat{\psi}_n(\cdot)$
2. Allows for general dynamics
   - extends the existing sieve estimation theory

Main Results

1. Consistency + Rate of convergence + Asym. normality results
   - under low-level conditions on the data generating process
   - assuming $\beta = (\theta, f)$ identified by $\hat{\psi}_n(\cdot)$
2. Allows for general dynamics
   - extends the existing sieve estimation theory
3. Inference is simple
   - compute SEs using Δ-method on $(\theta, \omega, \mu, \sigma)$ + Newey-West
   - SEs adapt to the rate of convergence

The Sieve-SMM Algorithm

Algorithm: Computing the Sieve-SMM Estimator

Compute $\hat{\psi}_n$, the sample CF of the data:

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t,\mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$

Set $k(n)$ number of mixture components

for $s = 1, \dots, S$ **do**

  Simulate $e_t^s \sim f_{\omega,\mu,\sigma}$ - Gaussian Mixture

  Simulate y_t^s using DGP & shocks e_t^s

  Compute $\hat{\psi}_n^s$, the CF of the simulated $(\mathbf{y}_t^s, \mathbf{x}_t)$

end for

Compute $\hat{Q}_n^S(\theta, f_{\omega,\mu,\sigma}) = \int |\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \theta, f_{\omega,\mu,\sigma})|^2 \pi(\tau) d\tau$

Minimize $\hat{Q}_n^S$ over $(\theta, \omega, \mu, \sigma)$

The Sieve-SMM Algorithm

Algorithm: Computing the Sieve-SMM Estimator

Compute $\hat{\psi}_n$, the sample CF of the data:

$$\hat{\psi}_n(\tau) = \frac{1}{n} \sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t, \mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$

Set $k(n)$ number of mixture components

for $s = 1, \dots, S$ **do**

Simulate $e_t^s \sim f_{\omega,\mu,\sigma}$ - Gaussian Mixture

Simulate y_t^s using DGP & shocks e_t^s

Compute $\hat{\psi}_n^s$, the CF of the simulated $(\mathbf{y}_t^s, \mathbf{x}_t)$

end for

Compute $\hat{Q}_n^S(\theta, f_{\omega,\mu,\sigma}) = \int |\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \theta, f_{\omega,\mu,\sigma})|^2 \pi(\tau) d\tau$

Minimize $\hat{Q}_n^S$ over $(\theta, \omega, \mu, \sigma)$

The Sieve-SMM Algorithm

Algorithm: Computing the Sieve-SMM Estimator

Compute $\hat{\psi}_n$, the sample CF of the data:

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t, \mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$

Set $k(n)$ number of mixture components

for $s = 1, \dots, S$ **do**

  Simulate $e_t^s \sim f_{\omega,\mu,\sigma}$ - Gaussian Mixture

  Simulate y_t^s using DGP & shocks e_t^s

  Compute $\hat{\psi}_n^s$, the CF of the simulated $(\mathbf{y}_t^s, \mathbf{x}_t)$

end for

Compute $\hat{Q}_n^S(\theta, f_{\omega,\mu,\sigma}) = \int |\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \theta, f_{\omega,\mu,\sigma})|^2 \pi(\tau) d\tau$

Minimize $\hat{Q}_n^S$ over $(\theta, \omega, \mu, \sigma)$

The Sieve-SMM Algorithm

Algorithm: Computing the Sieve-SMM Estimator

Compute $\hat{\psi}_n$, the sample CF of the data:

$$\hat{\psi}_n(\tau) = \frac{1}{n} \sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t, \mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$

Set $k(n)$ number of mixture components

for $s = 1, \dots, S$ **do**

  Simulate $e_t^s \sim f_{\omega,\mu,\sigma}$ - Gaussian Mixture

  Simulate y_t^s using DGP & shocks e_t^s

  Compute $\hat{\psi}_n^s$, the CF of the simulated $(\mathbf{y}_t^s, \mathbf{x}_t)$

end for

Compute $\hat{Q}_n^S(\theta, f_{\omega,\mu,\sigma}) = \int |\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \theta, f_{\omega,\mu,\sigma})|^2 \pi(\tau) d\tau$

Minimize $\hat{Q}_n^S$ over $(\theta, \omega, \mu, \sigma)$

The Sieve-SMM Algorithm

Algorithm: Computing the Sieve-SMM Estimator

Compute $\hat{\psi}_n$, the sample CF of the data:

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} e^{i\tau'(\mathbf{y}_t,\mathbf{x}_t)}, \quad \mathbf{y}_t = (y_t, \dots, y_{t-L})$$

Set $k(n)$ number of mixture components

for $s = 1, \dots, S$ **do**

Simulate $e_t^s \sim f_{\omega,\mu,\sigma}$ - Gaussian Mixture

Simulate y_t^s using DGP & shocks e_t^s

Compute $\hat{\psi}_n^s$, the CF of the simulated $(\mathbf{y}_t^s, \mathbf{x}_t)$

end for

Compute $\hat{Q}_n^S(\theta, f_{\omega,\mu,\sigma}) = \int |\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\theta,f_{\omega,\mu,\sigma})|^2 \pi(\tau) d\tau$

Minimize $\hat{Q}_n^S$ over $(\theta, \omega, \mu, \sigma)$

General Dynamics: A Non-Standard Problem

- Rich literature on Sieve-GMM and M-estimation (Chen, 2007)
- No result for Sieve + General Dynamics with Latent Variables
- Moments involve unobserved u_t
- Requires a filtering step: $\hat{u}_1(\hat{u}_0, y_1, \beta), \hat{u}_2(\hat{u}_1, y_2, \beta), \ldots$

General Dynamics: A Non-Standard Problem

- Rich literature on Sieve-GMM and M-estimation (Chen, 2007)
- No result for Sieve + General Dynamics with Latent Variables
- Moments involve unobserved u_t
- Requires a filtering step: $\hat{u}_1(\hat{u}_0, y_1, \beta), \hat{u}_2(\hat{u}_1, y_2, \beta), \dots$
- Non-standard setting because of:
  i. nonstationarity: $\hat{u}_0$ fixed
  ii. dependence properties of $\hat{u}_t$ vary with β ($\Rightarrow$ the results of Doukhan et al., 1995, do not apply)
  iii. propagation and accumulation of the approximation bias

General Dynamics: A Non-Standard Problem

- Rich literature on Sieve-GMM and M-estimation (Chen, 2007)
- No result for Sieve + General Dynamics with Latent Variables
- Moments involve unobserved u_t
- Requires a filtering step: $\hat{u}_1(\hat{u}_0, y_1, \beta), \hat{u}_2(\hat{u}_1, y_2, \beta), \ldots$
- Non-standard setting because of:
  - i. nonstationarity: $\hat{u}_0$ fixed
  - ii. dependence properties of $\hat{u}_t$ vary with β ($\Rightarrow$ the results of Doukhan et al., 1995, do not apply)
  - iii. propagation and accumulation of the approximation bias

$$
\begin{tikzcd}
 & y_1^s \arrow[r, "\textit{bias}"] & y_2^s \arrow[r, "\textit{bias}"] & \cdots \arrow[r, "\textit{bias}"] & y_t^s \\
e_1^s \sim f \arrow[ur, "\textit{bias}"] & e_2^s \sim f \arrow[ur, "\textit{bias}"] & \cdots & e_t^s \sim f \arrow[ur, "\textit{bias}"] &
\end{tikzcd}
$$

Allowing for general dynamics

This paper allows for:

i. nonstationarity due to (y_0, u_0)

ii. moments depend on full history $\psi_t(e_t^s, \ldots, e_1^s, x_t, \ldots, x_1, \beta)$

iii. dependence properties can vary with β

vi. accumulation of the approximation bias

Allowing for general dynamics

This paper allows for:

i. nonstationarity due to (y_0, u_0)

ii. moments depend on full history $\psi_t(e^s_t, \ldots, e^s_1, x_t, \ldots, x_1, \beta)$

iii. dependence properties can vary with β

vi. accumulation of the approximation bias

It relies on

1. An inequality adapted from results of Andrews and Pollard (1994); Ben Hariz (2005) - it requires
    i. bounded moments $|\hat{\psi}_t(\tau)| \leq M$ (CF, CDF)
    ii. and geometric ergodicity (as in Duffie and Singleton, 1993)
2. Conditions on the DGP to control the bias accumulation

Allowing for general dynamics

This paper allows for:

i. nonstationarity due to (y_0, u_0)
ii. moments depend on full history $\psi_t(e_t^s, \ldots, e_1^s, x_t, \ldots, x_1, \beta)$
iii. dependence properties can vary with β
vi. accumulation of the approximation bias

It relies on

1. An inequality adapted from results of Andrews and Pollard (1994); Ben Hariz (2005) - it requires
   i. bounded moments $|\hat{\psi}_t(\tau)| \leq M$ (CF, CDF)
   ii. and geometric ergodicity (as in Duffie and Singleton, 1993)
2. Conditions on the DGP to control the bias accumulation

Allowing for general dynamics

This paper allows for:

i. nonstationarity due to (y_0, u_0)
ii. moments depend on full history $\psi_t(e_t^s, \ldots, e_1^s, x_t, \ldots, x_1, \beta)$
iii. dependence properties can vary with β
vi. accumulation of the approximation bias

It relies on

1. An inequality adapted from results of Andrews and Pollard (1994); Ben Hariz (2005) - it requires
   i. bounded moments $|\hat{\psi}_t(\tau)| \leq M$ (CF, CDF)
   ii. and geometric ergodicity (as in Duffie and Singleton, 1993)
2. Conditions on the DGP to control the bias accumulation

Consistency - Assumptions (Common)

Assumption (Sieves, Identification, Dependence)

- *Sieves:* f *smooth,* Θ *compact, ...*
- *Identification:* $\beta_0 = (\theta_0, f_0)$ *unique minimizer*
- *Dependence: geometric ergodicity*

Consistency - Assumptions (Data Generating Process)

Assumption (DGP)

- *Decay Conditions*
  effect of (y_{t-1}, u_{t-1}) *on* (y_{t+h}, u_{t+h}) *decays geometrically in* $h \geq 1$

Remark: Lipschitz conditions can be replaced with L^2-smoothness

Consistency - Assumptions (Data Generating Process)

Assumption (DGP)

- *Decay Conditions* - **control the accumulation of the bias**
  effect of (y_{t-1}, u_{t-1}) *on* (y_{t+h}, u_{t+h}) *decays geometrically in* $h \geq 1$

Remark: Lipschitz conditions can be replaced with L^2-smoothness

Consistency - Assumptions (Data Generating Process)

Assumption (DGP)

- *Decay Conditions*
  effect of (y_{t-1}, u_{t-1}) *on* (y_{t+h}, u_{t+h}) *decays geometrically in* $h \geq 1$
- *Lipschitz in the Parameters*
  small changes in β *do not affect* (y_t, u_t) *too much*

Remark: Lipschitz conditions can be replaced with L^2-smoothness

Consistency - Assumptions (Data Generating Process)

Assumption (DGP)

- *Decay Conditions*
  effect of (y_{t-1}, u_{t-1}) *on* (y_{t+h}, u_{t+h}) *decays geometrically in* $h \geq 1$
- *Lipschitz in the Parameters* **- simulations are well behaved**
  small changes in β *do not affect* (y_t, u_t) *too much*

Remark: Lipschitz conditions can be replaced with L^2-smoothness

Consistency - Assumptions (Data Generating Process)

Assumption (DGP)

- *Decay Conditions*
  effect of (y_{t-1}, u_{t-1}) *on* (y_{t+h}, u_{t+h}) *decays geometrically in* $h \geq 1$
- *Lipschitz in the Parameters*
  small changes in β *do not affect* (y_t, u_t) *too much*
- *Lipschitz in the Shocks*
  small changes in e_t *do not affect* (y_t, u_t) *too much*

Remark: Lipschitz conditions can be replaced with L^2-smoothness

Consistency - Assumptions (Data Generating Process)

Assumption (DGP)

- *Decay Conditions*
  effect of (y_{t-1}, u_{t-1}) *on* (y_{t+h}, u_{t+h}) *decays geometrically in* $h \geq 1$
- *Lipschitz in the Parameters*
  small changes in β *do not affect* (y_t, u_t) *too much*
- *Lipschitz in the Shocks* - **drawing from a nonparametric density**
  small changes in e_t *do not affect* (y_t, u_t) *too much*

Remark: Lipschitz conditions can be replaced with L^2-smoothness

Consistency - Theorem

Theorem (Consistency)

Suppose Assumptions Sieves, Identification, Dependence and DGP hold + technical conditions then:

$$\|\hat{\beta}_n - \beta_0\|_{TV,\infty} = o_p(1)$$

where $\|\hat{\beta}_n - \beta_0\|_{TV,\infty}$ is one of

$$\|\hat{\beta}_n - \beta_0\|_{TV} = \|\hat{\theta}_n - \theta_0\| + 0.5 \int |\hat{f}_n(e) - f_0(e)| de$$

$$\|\hat{\beta}_n - \beta_0\|_{\infty} = \|\hat{\theta}_n - \theta_0\| + \sup_e |\hat{f}_n(e) - f_0(e)|$$

Decay Conditions and Approximation Bias

- Assumption Sieves implies
$$\text{Bias}_n = \|\beta_0 - \Pi_{k(n)}\beta_0\|_{TV,\infty} = O(\log[k(n)]^{2r} / k(n)^r)$$
  - $\Pi_{k(n)}\beta_0 =$ mixture approximation
  - $r =$ smoothness of f
- Simplify dynamics: $y_t^s = g_{obs,t}(\theta, e_t^s, \ldots, e_1^s)$

Decay Conditions and Approximation Bias

- Assumption Sieves implies

$$\text{Bias}_n = \|\beta_0 - \Pi_{k(n)}\beta_0\|_{TV,\infty} = O(\log[k(n)]^{2r} / k(n)^r)$$

  - $\Pi_{k(n)}\beta_0 =$ mixture approximation
  - $r =$ smoothness of f

- Simplify dynamics: $y_t^s = g_{obs,t}(\theta, e_t^s, \ldots, e_1^s)$

Decay Conditions and Approximation Bias

- Assumption Sieves implies

$$\text{Bias}_n = \|\beta_0 - \Pi_{k(n)}\beta_0\|_{TV,\infty} = O(\log[k(n)]^{2r}/k(n)^r)$$

  - $\Pi_{k(n)}\beta_0 =$ mixture approximation
  - $r =$ smoothness of f

- Simplify dynamics: $y_t^s = g_{obs,t}(\theta, e_t^s, \dots, e_1^s)$
- Moments imply a direct bound

$$\int |\mathbb{E}(\hat{\psi}_n^s(\tau, \beta_0) - \hat{\psi}_n^s(\tau, \pi_{k(n)}\beta_0))|^2 \pi(\tau) d\tau$$

$$= \int \Big| \frac{1}{n} \sum_{t=1}^{n} \int e^{i\tau' g_{obs,t}(\theta_0, e_t, \dots, e_1)} \underbrace{[f_0^{\otimes t} - \Pi_{k(n)} f_0^{\otimes t}]}_{\text{appears t times}} de_t \dots de_1 \Big|^2 \pi(\tau) d\tau$$

$$\leq n^2 \times \|f_0 - \Pi_{k(n)} f_0\|_{TV}^2 \text{ (up to a constant)}$$

Decay Conditions and Approximation Bias

- Assumption Sieves implies

$$\text{Bias}_n = \|\beta_0 - \Pi_{k(n)}\beta_0\|_{TV,\infty} = O(\log[k(n)]^{2r}/k(n)^r)$$

  - $\Pi_{k(n)}\beta_0$ = mixture approximation
  - r = smoothness of f

- Simplify dynamics: $y_t^s = g_{obs,t}(\theta, e_t^s, \dots, e_1^s)$
- Moments imply a direct bound

$$\int |\mathbb{E}(\hat{\psi}_n^s(\tau,\beta_0) - \hat{\psi}_n^s(\tau,\pi_{k(n)}\beta_0))|^2 \pi(\tau) d\tau$$

$$= \int \Big| \frac{1}{n} \sum_{t=1}^{n} \int e^{i\tau' g_{obs,t}(\theta_0, e_t,\dots,e_1)} \underbrace{[f_0^{\otimes t} - \Pi_{k(n)} f_0^{\otimes t}]}_{\text{appears t times}} de_t \dots de_1 \Big|^2 \pi(\tau) d\tau$$

$$\leq n^2 \times \|f_0 - \Pi_{k(n)} f_0\|_{TV}^2 \text{ (up to a constant)}$$

Decay Conditions and Approximation Bias

- Assumption Sieves implies

$$\text{Bias}_n = \|\beta_0 - \Pi_{k(n)}\beta_0\|_{TV,\infty} = O(\log[k(n)]^{2r}/k(n)^r)$$

  - $\Pi_{k(n)}\beta_0$ = mixture approximation
  - r = smoothness of f

- Simplify dynamics: $y_t^s = g_{obs,t}(\theta, e_t^s, \dots, e_1^s)$
- Moments imply a direct bound

$$\int |\mathbb{E}(\hat{\psi}_n^s(\tau, \beta_0) - \hat{\psi}_n^s(\tau, \pi_{k(n)}\beta_0))|^2 \pi(\tau) d\tau$$

$$= \int \Big| \frac{1}{n} \sum_{t=1}^{n} \int e^{i\tau' g_{obs,t}(\theta_0, e_t, \dots, e_1)} \underbrace{[f_0^{\otimes t} - \Pi_{k(n)} f_0^{\otimes t}]}_{\text{appears t times}} de_t \dots de_1 \Big|^2 \pi(\tau) d\tau$$

$$\leq n^2 \times \|f_0 - \Pi_{k(n)} f_0\|_{TV}^2 \text{ (up to a constant)}$$

Decay Conditions and Approximation Bias

- Assumption Sieves implies

$$\text{Bias}_n = \|\beta_0 - \Pi_{k(n)}\beta_0\|_{TV,\infty} = O(\log[k(n)]^{2r} / k(n)^r)$$

  - $\Pi_{k(n)}\beta_0 =$ mixture approximation
  - $r =$ smoothness of f

- Simplify dynamics: $y_t^s = g_{obs,t}(\theta, e_t^s, \dots, e_1^s)$
- Moments imply a direct bound

$$\int |\mathbb{E}(\hat{\psi}_n^s(\tau, \beta_0) - \hat{\psi}_n^s(\tau, \pi_{k(n)}\beta_0))|^2 \pi(\tau) d\tau$$

$$= \int \Big| \frac{1}{n} \sum_{t=1}^{n} \int e^{i\tau' g_{obs,t}(\theta_0, e_t, \dots, e_1)} \underbrace{[f_0^{\otimes t} - \Pi_{k(n)} f_0^{\otimes t}]}_{\text{appears t times}} de_t \dots de_1 \Big|^2 \pi(\tau) d\tau$$

$$\leq n^2 \times \|f_0 - \Pi_{k(n)} f_0\|_{TV}^2 \text{ (up to a constant)}$$

- Bound grows too fast! $n \times \text{Bias}_n \to \infty$

$\Rightarrow$ Use decay conditions to control the bias

Decay Conditions and Approximation Bias

- Sketch of the argument: take a shorter history $m \ll n$

$$\tilde{y}_t^s = g_{obs,t}(e_t^s, \ldots, e_{t-m}^s, 0, \ldots, 0)$$

Two implications:

i. For $\tilde{y}_t^s$, the bias is $m \times \text{Bias}_n \ll n \times \text{Bias}_n$

ii. The decay conditions imply that for some $0 \leq \bar{\rho} < 1$

$$\mathbb{E}(\|y_t^s - \tilde{y}_t^s\|) \leq \bar{\rho}^m \text{ (up to a constant)}$$

Decay Conditions and Approximation Bias

- Sketch of the argument: take a shorter history $m \ll n$

$$\tilde{y}_t^s = g_{obs,t}(e_t^s, \ldots, e_{t-m}^s, 0, \ldots, 0)$$

Two implications:

i. For $\tilde{y}_t^s$, the bias is $m \times \text{Bias}_n \ll n \times \text{Bias}_n$

ii. The decay conditions imply that for some $0 \leq \bar{\rho} < 1$

$$\mathbb{E}(\|y_t^s - \tilde{y}_t^s\|) \leq \bar{\rho}^m \text{ (up to a constant)}$$

Decay Conditions and Approximation Bias

- Sketch of the argument: take a shorter history $m \ll n$

$$\tilde{y}_t^s = g_{obs,t}(e_t^s, \ldots, e_{t-m}^s, 0, \ldots, 0)$$

Two implications:

i. For $\tilde{y}_t^s$, the bias is $m \times \text{Bias}_n \ll n \times \text{Bias}_n$

ii. The decay conditions imply that for some $0 \leq \bar{\rho} < 1$

$$\mathbb{E}(\|y_t^s - \tilde{y}_t^s\|) \leq \bar{\rho}^m \text{ (up to a constant)}$$

Decay Conditions and Approximation Bias

- Sketch of the argument: take a shorter history $m \ll n$

$$\tilde{y}_t^s = g_{obs,t}(e_t^s, \ldots, e_{t-m}^s, 0, \ldots, 0)$$

Two implications:

i. For $\tilde{y}_t^s$, the bias is $m \times \text{Bias}_n \ll n \times \text{Bias}_n$

ii. The decay conditions imply that for some $0 \leq \bar{\rho} < 1$

$$\mathbb{E}(\|y_t^s - \tilde{y}_t^s\|) \leq \bar{\rho}^m \text{ (up to a constant)}$$

Decay Conditions and Approximation Bias

- Sketch of the argument: take a shorter history $m \ll n$

$$\tilde{y}_t^s = g_{obs,t}(e_t^s, \ldots, e_{t-m}^s, 0, \ldots, 0)$$

Two implications:

i. For $\tilde{y}_t^s$, the bias is $m \times \text{Bias}_n \ll n \times \text{Bias}_n$
ii. The decay conditions imply that for some $0 \leq \bar{\rho} < 1$

$$\mathbb{E}(\|y_t^s - \tilde{y}_t^s\|) \leq \bar{\rho}^m \text{ (up to a constant)}$$

- Balance the two with $m = C_{\bar{\rho}} \times |\log[\text{Bias}_n]|$:

$$\bar{\rho}^m + m \times \text{Bias}_n = O(\text{Bias}_n |\log[\text{Bias}_n]|)$$

Decay Conditions and Approximation Bias

- Sketch of the argument: take a shorter history $m \ll n$

$$\tilde{y}_t^s = g_{obs,t}(e_t^s, \ldots, e_{t-m}^s, 0, \ldots, 0)$$

Two implications:

i. For $\tilde{y}_t^s$, the bias is $m \times \mathrm{Bias}_n \ll n \times \mathrm{Bias}_n$

ii. The decay conditions imply that for some $0 \leq \bar{\rho} < 1$

$$\mathbb{E}(\|y_t^s - \tilde{y}_t^s\|) \leq \bar{\rho}^m \text{ (up to a constant)}$$

- Balance the two with $m = C_{\bar{\rho}} \times |\log[\mathrm{Bias}_n]|$:

$$\bar{\rho}^m + m \times \mathrm{Bias}_n = O(\mathrm{Bias}_n |\log[\mathrm{Bias}_n]|)$$

- The dynamics inflate the bias by a log term:

$$\mathrm{Bias}_n |\log[\mathrm{Bias}_n]| = O\left(\log[k(n)]^{2r+1} / k(n)^r\right)$$

Theorem - Rate of Convergence

Theorem (Rate of Convergence)

Under previous assumptions:

$$\|\hat{\beta}_n - \beta_0\|_{weak} = O_p\left(\underbrace{\frac{\log[k(n)]^{2r+1}}{k(n)^r}}_{bias \times |\log(bias)|} + \underbrace{\frac{(k(n)\log[k(n)])^2}{\sqrt{n}}}_{variance}\right)$$

$\|\cdot\|_{weak}$ of Ai and Chen (2003) with continuum of moments

- Empirical Application:
  - if $r = 2$, optimal rate is $n^{-1/4}$ (up to a log term)

  $\Rightarrow$ slower than usual $n^{-2/5}$ rate.

$\Rightarrow$ Require smoother densities - at least $r \geq 3$ cont. derivatives.

Theorem - Asymptotic Normality

Theorem (Asymptotic Normality)

Suppose $\|\hat{\beta}_n - \beta_0\|_{weak} = o_p(n^{-1/4})$ + *undersmoothing* + *additional conditions (common) then:*

$$r_n \left(\phi(\hat{\beta}_n) - \phi(\beta_0)\right) \xrightarrow{d} \mathcal{N}(0, 1)$$

for $r_n = \sqrt{n}/\sigma_n$, ϕ *smooth functional*

Depending on both $\hat{\psi}_n$ and ϕ, two cases:

i. $\sigma_n \to \infty$: slower than $\sqrt{n}$-convergence

ii. $\sigma_n \not\to \infty$: $\sqrt{n}$-convergence

Conclusion

- A Sieve-SMM estimator
  - extends SMM literature (semi-nonparametric)
  - and sieve literature (more general dynamics)

Conclusion

- A Sieve-SMM estimator
  - extends SMM literature (semi-nonparametric)
  - and sieve literature (more general dynamics)
- High-level conditions allow for other moments/bases/functions

Conclusion

- A Sieve-SMM estimator
  - extends SMM literature (semi-nonparametric)
  - and sieve literature (more general dynamics)
- High-level conditions allow for other moments/bases/functions
- Limitations & open questions:
  - bounded moments - need more work for dynamic sieve-GMM
  - identification of (θ, f) - hard to check, impossible to test
  - More efficient simulation methods: control variates and quasi-Monte Carlo integration

Thank You!

References

Ai, C. and Chen, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, 71(6):1795–1843.

Andrews, D. W. K. and Pollard, D. (1994). An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *International Statistical Review / Revue Internationale de Statistique*, 62(1):119.

Ben Hariz, S. (2005). Uniform CLT for empirical process. *Stochastic Processes and their Applications*, 115(2):339–358.

Blundell, R., Chen, X., and Kristensen, D. (2007). Semi-nonparametric IV estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669.

Carrasco, M., Chernov, M., Florens, J.-P., and Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2):529–573.

Chen, X. (2007). Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models. In *Handbook of Econometrics*, volume 6, pages 5549–5632.

Chen, X. and Liao, Z. (2015). Sieve semiparametric two-step GMM under weak dependence. *Journal of Econometrics*, 189(1):163–186.

Chen, X. and Pouzo, D. (2015). Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models. *Econometrica*, 83(3):1013–1079.

Doukhan, P., Massart, P., and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Annales de l'Institut Henri Poincaré, section B*, tome 31(2):393–427.

Duffie, D. and Singleton, K. J. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61(4):929.

Gospodinov, N., Komunjer, I., and Ng, S. (2017). Simulated minimum distance estimation of dynamic models with errors-in-variables. *Journal of Econometrics*, 200(2):181–193.

Gospodinov, N. and Ng, S. (2015). Minimum Distance Estimation of Possibly Noninvertible Moving Average Models. *Journal of Business & Economic Statistics*, 33(3):403–417.

Gouriéroux, C. and Monfort, A. (1996). *Simulation-based Econometric Methods*. CORE lectures. Oxford University Press.

References iii

Ruge-Murcia, F. (2012). Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles. *Journal of Economic Dynamics and Control*, 36(6):914–938.

Ruge-Murcia, F. (2017). Skewness Risk and Bond Prices. *Journal of Applied Econometrics*, 32(2):379–400.

Smith, A. A. (2016). Indirect Inference. In *The New Palgrave Dictionary of Economics*, pages 1–6. Palgrave Macmillan UK, London.

Drawing from a Gaussian Mixtures

- Draw $\nu_t^s \overset{iid}{\sim} \mathcal{U}_{[0,1]}$ and $Z_{t,1}^s, \ldots, Z_{t,k(n)}^s \overset{iid}{\sim} \mathcal{N}(0,1)$
- Compute

$$e_t^s = \sum_{j=1}^{k(n)} \mathbb{1}_{\nu_t^s \in [\sum_{l=0}^{j-1} \omega_l, \sum_{l=0}^{j} \omega_l]} \left(\mu_j + \sigma_j Z_{t,j}^s\right)$$

- Then

$$e_t^s \sim \sum_{j=1}^{k(n)} \frac{\omega_j}{\sigma_j} \phi\left(\frac{e - \mu_j}{\sigma_j}\right) = f_{\omega,\mu,\sigma}$$

- Keep $\nu_t^s, Z_{t,j}^s$ fixed during optimization!
- Discontinuities due to $\mathbb{1}_{(\cdot)}$ but standard derivative-free optimizers (Nelder-Mead) work well

Main

Approximation Properties of Gaussian Mixtures

Lemma (Approximation Rate)

Suppose $f = f_1 \times \cdots \times f_{d_e}$ *with each* f_j *such that:*

- *Smoothness:* f_j *is* $r \geq 2$ *times cont. differentiable*
- *Tails:* $f_j(e) \leq e^{-a|e|^b}$ *(up to a constant) for some* $a, b > 0$

There exists a mixture $\Pi_{k(n)} f = \Pi_{k(n)} f_1 \times \cdots \times \Pi_{k(n)} f_{d_e}$ *with*

- *Bandwith:* $\sigma_j \geq \sigma_{k(n)} = O(\log[k(n)]^{2/b} / k(n))$
- *Bounds:* $\mu_j \in [-\mu_{k(n)}, \mu_{k(n)}]$, $\mu_{k(n)} = O(\log[k]^{1/b})$

such that

$$\|f - \Pi_{k(n)} f\|_{TV,\infty} = O(\log[k(n)]^{2r/b} / k(n)^r)$$

Main

Allowing for Student Like Tails

- Add two tail components to the density: f_L, f_R such that:

$$f_L(e,\xi_L) = (2+\xi_L)\frac{-e^{1+\xi_L}}{[1-e^{2+\xi_L}]^2} \quad \text{for } e \leq 0$$
$$f_R(e,\xi_R) = (2+\xi_R)\frac{e^{1+\xi_R}}{[1+e^{2+\xi_R}]^2} \quad \text{for } e \geq 0$$

- To simulate, draw $\nu_L, \nu_R \sim \mathcal{U}_{[0,1]}$:

$$Z_L = -(1/\nu_L - 1)^{1/(2+\xi_L)}, \quad Z_R = (1/\nu_R - 1)^{1/(2+\xi_R)}$$

- The left and right tail indices are $3+\xi_L$ and $3+\xi_R$

Main

L^2-Smoothness of Mixture Draws

Lemma (L^2-Smoothness)

For $|\mu_j|$ and $|\tilde{\mu}_j| \leq \bar{\mu}_{k(n)}$, $|\sigma_j|$ and $|\tilde{\sigma}_j| \leq \bar{\sigma}$, let

$$e_t^s = \sum_{j=1}^{k(n)} \mathbb{1}_{\nu_t^s \in [\sum_{l=0}^{j-1} \omega_l, \sum_{l=0}^{j} \omega_l]} \left(\mu_j + \sigma_j Z_{t,j}^s \right)$$

$$\tilde{e}_t^s = \sum_{j=1}^{k(n)} \mathbb{1}_{\nu_t^s \in [\sum_{l=0}^{j-1} \tilde{\omega}_l, \sum_{l=0}^{j} \tilde{\omega}_l]} \left(\tilde{\mu}_j + \tilde{\sigma}_j Z_{t,j}^s \right)$$

If $\mathbb{E}(|Z_{t,j}^s|^2) \leq C_Z^2 < \infty$ then there exists a finite constant C which only depends on C_Z such that:

$$\mathbb{E}\left(\sup_{\|(\omega,\mu,\sigma)-(\tilde{\omega},\tilde{\mu},\tilde{\sigma})\|_2 \leq \delta} \left| e_t^s - \tilde{e}_t^s \right|^2 \right)^{1/2} \leq C \left(1 + \bar{\mu}_{k(n)} + \bar{\sigma} + k(n) \right) \delta^{1/2}.$$

Identification of the SV Model

$$
\begin{aligned}
y_t &= \mu_y + \rho_y y_{t-1} + \sigma_t e_{t,1}, \quad e_{t,1} \overset{iid}{\sim} f \\
\sigma_t^2 &= \mu_\sigma + \rho_\sigma \sigma_{t-1}^2 + \kappa_\sigma e_{t,2}
\end{aligned}
$$

- Suppose $\mathbf{y}_t = (y_t, y_{t-1}, y_{t-2})$ $(L \geq 2)$
- If $\mathbb{E}(\hat{\psi}_n^s(\tau, \beta_1)) = \mathbb{E}(\hat{\psi}_n^s(\tau, \beta_2))$ for all τ
- Then for any $\ell_1, \ell_2, \ell_3 \geq 0$

$$
\mathbb{E}_{\beta_1}(y_t^{\ell_1} y_{t-1}^{\ell_2} y_{t-2}^{\ell_3}) = \mathbb{E}_{\beta_2}(y_t^{\ell_1} y_{t-1}^{\ell_2} y_{t-2}^{\ell_3})
$$

- Using $(\ell_1, \ell_2, \ell_3) = (1,0,0), (2,0,0), (1,1,0), (2,2,0), (2,0,2)$ implies θ identified if $\rho_\sigma \neq 0$
- Then for $\ell_2, \ell_3 = 0$, pick $\ell_1 = 3, 4, 5, \dots$ if f_y, f have analytic MGF then f identified

Main

Weak Norm of Ai and Chen (2003)

- Infinite dim. space: not all norms are equivalent
- Easier to derive results in norm related to $\hat{Q}_n^s$ around β_0:

$$\|\beta_1 - \beta_2\|_{weak}^2 = \int \left| \frac{d\mathbb{E}(\hat{\psi}_n^s(\tau, \beta_0))}{d\beta} [\beta_1 - \beta_2] \right|^2 \pi(\tau) d\tau$$

- Then derive cv. rate in strong norm $\|\cdot\|_{TV,\infty}$ using:

$$\|\hat{\beta}_n - \beta_0\|_{TV,\infty} \leq \underbrace{\|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_{TV,\infty}}_{\text{approximation bias}}$$

$$+ \underbrace{\|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_{weak}}_{\text{cv. rate in weak norm}} \times \underbrace{\sup_{\mathcal{B}_{k(n)}, \|\beta - \Pi_{k(n)}\beta_0\|_{weak} \neq 0} \frac{\|\beta - \Pi_{k(n)}\beta_0\|_{TV,\infty}}{\|\beta - \Pi_{k(n)}\beta_0\|_{weak}}}_{\text{measure of ill-posedness}}$$

Main

Measure of Ill-Posedness

- Local measure of ill-posedness of Blundell et al. (2007)

$$\tau_n = \sup_{\mathcal{B}_{k(n)}, \|\beta - \Pi_{k(n)}\beta_0\|_{weak} \neq 0} \frac{\|\beta - \Pi_{k(n)}\beta_0\|_{TV,\infty}}{\|\beta - \Pi_{k(n)}\beta_0\|_{weak}}$$

- In the paper: simple example

$$\tau_{TV,n} = O\left(k(n)\right), \quad \tau_{TV,n} = O\left(k(n)^2\right)$$

- Also derive upper-bounds in the general case:

$$\tau_{TV,n} \leq \lambda_n^{-1/2} k(n), \quad \tau_{\infty,n} \leq \lambda_n^{-1/2} k(n)^2$$

- Where λ_n is the smallest eigenvalue of:

$$\int \frac{d\mathbb{E}(\hat{\psi}_n^s(\tau, \Pi_{k(n)}\beta_0))}{d(\theta, \omega, \mu, \sigma)'} \overline{\frac{d\mathbb{E}(\hat{\psi}_n^s(\tau, \Pi_{k(n)}\beta_0))}{d(\theta, \omega, \mu, \sigma)}} \pi(\tau) d\tau$$

- It can be computed numerically using $\hat{\beta}_n$ and $\hat{\psi}_n^s$

Faster Rate of cv. with $S \to \infty$

- If can simulate a long sequence $y_1^s, \ldots, y_{n\times S}^s$ then variance becomes:

$$\max\left(\frac{1}{\sqrt{n}}, \frac{(k(n)\log[k(n)])^2}{\sqrt{nS}}\right)$$

- For $S = k(n)^4$ and $k(n) = O(n^{1/2r})$ get:

$$\|\hat{\beta}_n - \beta_0\|_{weak} = O(1/\sqrt{n})$$

- <u>But in practice</u>
  - $\|\hat{\beta}_n - \beta_0\|_{TV,\infty}$ slower
  - Computationally demanding: $n = 1,000$ and $r = 2$ $\Rightarrow k(n) \simeq 5$ and $S > 600$
- In general: $S \geq 1$ can shift the bias/variance tradeoff

Main

Optimal Number of Mixture Components $k(n)$

- In theory optimal $k(n)$:

$$\text{Bias}_n \times |\log(\text{Bias}_n)| \asymp \text{Variance}$$

- Suggests $k(n) \asymp n^{-1/(4+2r)}$
    - e.g. $r = 2, k(n) \asymp n^{-1/8}$
- In practice:
    i. Monte-Carlo Simulations
    ii. Information Criterion (AIC/BIC):

$$\hat{Q}_n^S(\hat{\beta}_n) + \frac{(k(n)\log[k(n)])^4}{n} \simeq \text{Bias}_n^2 \times |\log(\text{Bias}_n)|^2 + \text{Variance}^2$$

Main

Standard Errors σ_n

- Standard Errors σ_n formula

$$\sigma_n^2 = n\mathbb{E}\left(\left[\int Real\left(\psi_\beta(\tau, v_n^*)\overline{\left[\psi_n^s(\tau,\beta_0) - \hat{\psi}_n(\tau)\right]}\right)\right]^2\right)$$

where v_n^*: sieve repr. computed using $\hat{\psi}_n^s$ and $\hat{\beta}_n$ (as in Chen and Pouzo, 2015). $\psi_\beta = d\hat{\psi}_n^s / d\beta$

- LR variance computed using either
  i. Newey-West estimator
  ii. or the Block Bootstrap

(Chen and Liao, 2015; Carrasco et al., 2007)

Main

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions

  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_\mathbf{y}}\right]\right)$$

Main

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions

  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t

  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f

  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_\mathbf{y}}\right]\right)$$

Main

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions
  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_\mathbf{y}}\right]\right)$$

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions
  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_\mathbf{y}}\right]\right)$$

Main

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions
  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_\mathbf{y}}\right]\right)$$

- e.g. Stochastic Volatility model (estimate only ρ_y, ρ_σ)
  i. holds by construction
  ii. holds by restriction: $\mathbb{E}(e_{t,1}) = 0, \mathbb{V}(e_{t,1}) = 1$
  iii. holds if $\mu_\sigma \neq 0$ and $\rho_\sigma \neq 0$

Main

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions
  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_\mathbf{y}}\right]\right)$$

- e.g. Stochastic Volatility model (estimate only ρ_y, ρ_σ)
  i. holds by construction
  ii. holds by restriction: $\mathbb{E}(e_{t,1}) = 0, \mathbb{V}(e_{t,1}) = 1$
  iii. holds if $\mu_\sigma \neq 0$ and $\rho_\sigma \neq 0$

Main

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions
  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_{\mathbf{y}}}\right]\right)$$

- e.g. Stochastic Volatility model (estimate only ρ_y, ρ_σ)
  i. holds by construction
  ii. holds by restriction: $\mathbb{E}(e_{t,1}) = 0, \mathbb{V}(e_{t,1}) = 1$
  iii. holds if $\mu_\sigma \neq 0$ and $\rho_\sigma \neq 0$

Main

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions
    i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
    ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
    iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_\mathbf{y}}\right]\right)$$

- e.g. Stochastic Volatility model (estimate only ρ_y, ρ_σ)
    i. holds by construction
    ii. holds by restriction: $\mathbb{E}(e_{t,1}) = 0, \mathbb{V}(e_{t,1}) = 1$
    iii. holds if $\mu_\sigma \neq 0$ and $\rho_\sigma \neq 0$

More on Asymptotic Normality

- More specific result:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V)$$

- Under additional restrictions
  i. g_{obs}, g_{latent} only depend on f via e_t + smooth in θ + no x_t
  ii. $\mathbb{E}(\mathbf{y_t}), \mathbb{V}(\mathbf{y_t})$ do not depend on f
  iii. The following matrix has full rank

$$\mathbb{E}_{\theta_0, f_0}\left(\frac{d\mathbf{y}_t^s}{d\theta'}\left[\begin{pmatrix} 1, & \mathbf{y}_t^{s\prime} \end{pmatrix} \otimes I_{d_{\mathbf{y}}}\right]\right)$$

- e.g. Stochastic Volatility model (estimate only ρ_y, ρ_σ)
  i. holds by construction
  ii. holds by restriction: $\mathbb{E}(e_{t,1}) = 0, \mathbb{V}(e_{t,1}) = 1$
  iii. holds if $\mu_\sigma \neq 0$ and $\rho_\sigma \neq 0$

Main

Short Panels

Theorem (Short Panels)

Suppose y_t is generated by either

$$\begin{aligned} y_{j,t} &= g_{obs}(x_{j,t}, \beta, u_{j,t}) \\ u_{j,t} &= g_{latent}(u_{j,t-1}, \beta, e_{j,t}) \end{aligned} \tag{1}$$

or

$$y_{j,t} = g_{obs}(y_{j,t-1}, x_{j,t}, \beta, e_{j,t}) \tag{2}$$

where $e_{j,t} \overset{iid}{\sim} f$ + technical cond. Then the main results hold with

$$\|\hat{\beta}_n - \beta_0\|_{weak} = O_p\left(\frac{\log[k(n)]^{2r+1}}{k(n)^r} + \frac{\sqrt{k(n)\log[k(n)]}}{\sqrt{n}}\right)$$

Main

Short Panels: Example

- Dynamic Tobit Model

$$y_{j,t} = (x_{j,t}'\theta_1 + u_{j,t})\mathbb{1}_{x_{j,t}'\theta_1 + u_{j,t} \geq 0}$$
$$u_{j,t} = \rho_u u_{j,t-1} + e_{j,t}, \quad e_{j,t} \overset{iid}{\sim} f$$

- Applications: Labor supply
- Quantities of interest include e.g.
  - $\mathbb{P}(y_{j,t} > 0 | x, y_{j,t-1} = 0)$ - (re)-entering the labor force
  - $\partial_{x_{j,t}}\mathbb{P}(y_{j,t} > 0 | x, y_{j,t-1} = 0)$ - marginal effects of covariates on (re)-entering the labor force

Main

Auxiliary Variables

Theorem (Using Auxiliary Variables)

Suppose

$$z_t^{aux} = g_{aux,t}(y_t, \dots, y_1, x_t, \dots, x_1, \hat{\eta}_n^{aux})$$
$$z_t^{s,aux} = g_{aux,t}(y_t^s, \dots, y_1^s, x_t, \dots, x_1, \hat{\eta}_n^{aux})$$

with $\hat{\eta}_n^{aux}$ *CAN + computed from* (y_t, x_t) *with* $g_{aux,t}$ *Lipschitz in* η^{aux} *and*

$$\|g_{aux,t}(y, x, \eta^{aux}) - g_{aux,t}(\tilde{y}, x, \eta^{aux})\| \leq \sum_{j=1}^{t} \rho_j \|y_j - \tilde{y}_j\|$$

and (Summability) $\sum_{j=1}^{\infty} \rho_j < \infty$ *+ additional cond. then the main results hold.*

Main

Auxiliary Variables: Example

- GARCH(1,1) filtered volatility for the SV model:

$$\sigma_t^{2,aux} = \mu^{aux} + \alpha_1^{aux}\sigma_{t-1}^{2,aux} + \alpha_2^{aux} y_t^2$$

- Summability cond. implied by $|\alpha_1^{aux}| \leq \overline{\alpha} < 1$.
- Instead of just lags $\mathbf{y}_t = (y_t, \ldots, y_{t-L})$, $\hat{\sigma}_t^{2,aux}$ add information in CF from the whole history:

$$\hat{\sigma}_t^{2,aux} = g_{aux,t}(y_t, \ldots, y_1; \hat{\eta}^{aux})$$

Main

Second Empirical Application: USD/GBP Exchange Rate Data

- January 2000 - December 2016: 5,447 daily observations
- Data Generating Process:

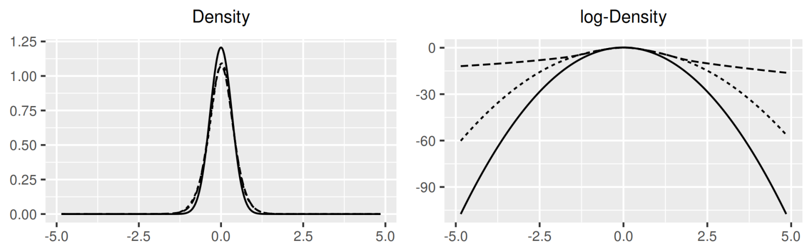
$$\begin{aligned} y_t &= \mu_y + \sigma_t e_{t,1}, & e_{t,1} &\overset{iid}{\sim} f \\ \log(\sigma_t) &= \rho_\sigma \log(\sigma_{t-1}) + \kappa_\sigma e_{t,2}, & e_{t,2} &\overset{iid}{\sim} \mathcal{N}(0,1) \end{aligned}$$

- with the restriction $e_{t,1} \sim (0, \sigma^2)$
- Estimates: parametric Gibbs Sampling, Sieve-SMM Gaussian Mixture and Gaussian Mixture + Tails

Main

Second Empirical Application: Estimates

Exchange Rate: Density and log-Density Estimates



Note: solid line: Gaussian density, dotted line: Gaussian mixture, dashed: Gaussian and tails mixture.

Main

Second Empirical Application: Estimates

		ρ_z	σ_z
Bayesian	Estimate	0.24	1.31
	95% CI	$[0.16, 0.34]$	$[1.21, 1.41]$
Sieve-SMM	Estimate	0.96	0.22
	95% CI	$[0.59, 0.99]$	$[0.06, 0.83]$
Sieve-SMM Tails	Estimate	0.97	0.19
	95% CI	$[0.62, 0.99]$	$[0.05, 0.79]$

Note: CI is the credible interval for the Bayesian and the confidence interval for the frequentist estimates.

Main