Valid Inference in Partially Unstable GMM Models

Hong Li
Brandeis University

Ulrich K. Müller
Princeton University

December 2, 2005

Motivation

- Time series models have potentially time varying parameters

- Recent interest in testing parameter stability
  $\Rightarrow$ Nyblom (1989), Andrews (1993), Andrews and Ploberger (1994), Hansen (2000), Elliott and Müller (2003)

- What to do if instability are found/suspected?
  - Inference on stable subset of parameters
  - Inference on parameter path

Overview

- Generalized Method of Moments (GMM) framework
- Focus on instabilities that are small in the sense that reasonable tests detect them with (possibly large) probability smaller than one in the limit
- Main result: standard GMM inference (ignoring the partial instability) remains asymptotically valid for the subset of stable parameters

Structure of Talk

1. Introduction
2. High Level Assumptions and Main Result
3. Sketch of Proof
4. Contiguity and Its Application
5. Monte Carlo Results
6. Conclusion

GMM Set-up

- Data is $\{y_t\}_{t=1}^{T}$. Model with time invariant parameter $\theta_0 \in \Theta \subset \mathbb{R}^m$ satisfies

$$E[g(y_t, \theta_0)] = 0 \text{ for all } t \leq T.$$

- Let $\{\theta_t\}_{t=1}^{T} \in \Theta^T$ be the parameter path in the unstable model, such that

$$E[g(y_t, \theta_t)] = 0 \text{ for all } t \leq T.$$

- Let $g_t(\theta) = g(y_t, \theta)$ and $G_t(\theta) = \partial g(y_t, \theta)/\partial\theta$.

- We analyze properties of usual GMM estimator

$$\hat{\theta} = \arg\min_{\theta} \left(T^{-1} \sum_{1}^{T} g_t(\theta) \right) Q_T \left(T^{-1} \sum_{1}^{T} g_t(\theta) \right)$$

for sequence of positive definite weighting matrices Q_T

Example

- Linear model $y_t = X_t\beta_t + Z_t\delta + \mu + \varepsilon_t, \;\; \varepsilon_t \sim iid(0, \sigma^2)$

- Rewrite $y_t = W_t'\theta_t + \varepsilon_t$, where $W_t = (X_t, Z_t, 1)$ and $\theta_t = (\beta_t, \delta, \mu)$

- GMM with $g_t(\theta) = W_t(y_t - W_t'\theta)$ and $Q_T = I_3$ equivalent to OLS

- We are interested in conducting inference on δ, μ

- Can't simply run short regression of y_t on $(Z_t, 1)$, since X_t and Z_t might be correlated

High Level Assumptions I

(i) The parameter evolves as $T^{1/2}(\theta_t - \theta_0) = f(t/T)\ \forall t \leq T$ for some nonstochastic, bounded and piece-wise continuous function $f : [0,1] \mapsto \mathbb{R}^m$ with at most a finite number of discontinuities.

Comments

- corresponds to local neighborhood in which tests of parameter stability have nontrivial power
- almost unrestricted otherwise: smooth evolution, single break, multiple breaks, ...

High Level Assumptions II

(ii) In some neighborhood Θ_0 of θ_0, $g_t(\theta)$ is differentiable in θ a.s. for $t \leq T, T \geq 1$.

(iii) $T^{-1/2} \sum_1^T g_t(\theta_t) \Rightarrow \mathcal{N}(0, V)$ for some positive definite $p \times p$ matrix V.
$(T^{-1/2} \sum_1^T W_t \varepsilon_t \Rightarrow \mathcal{N}(0, V))$

(iv) $||\hat{\theta} - \theta_0|| \xrightarrow{p} 0$.

(v) $||Q_T - Q_0|| \xrightarrow{p} 0$ for some positive definite matrix Q_0, and there exist positive definite $p \times p$ matrices $\hat{V}_T$ such that $||\hat{V}_T - V|| \xrightarrow{p} 0$.
$(\hat{V}_T = \hat{\sigma}^2 T^{-1} \sum_1^T W_t W_t')$

High Level Assumptions III

(vi) $T^{-1}\sum_1^T ||G_t(\theta_0)|| = O_p(1)$ $(T^{-1}\sum_1^T ||W_t W_t'|| = O_p(1))$, and for any decreasing neighborhood Θ_T of θ_0 contained in Θ_0, i.e. $\Theta_T = \{\theta : ||\theta - \theta_0|| < c_T\} \subset \Theta_0$ for some sequence of real numbers $c_T \rightarrow 0$, $T^{-1}\sum_1^T \sup_{\theta\in\Theta_T} ||G_t(\theta) - G_t(\theta_0)|| \xrightarrow{p} 0$. $(T^{-1}\sum_1^T ||0|| \xrightarrow{p} 0)$

(vii) $\sup_{0\le\lambda\le1} \left\| T^{-1}\sum_{t=1}^{[\lambda T]} G_t(\theta_0) - \lambda\Gamma \right\| \xrightarrow{p} 0$ for some positive definite $p \times m$ matrix Γ.
$(\sup_{0\le\lambda\le1} \left\| T^{-1}\sum_{t=1}^{[\lambda T]} W_t W_t' - \lambda\Gamma \right\| \xrightarrow{p} 0)$

Main result

Theorem: (i) Under the stated assumption

$$T^{1/2}\hat{\Sigma}_{\theta}^{-1/2}(\hat{\theta} - T^{-1}\sum_{1}^{T}\theta_t) \Rightarrow \mathcal{N}(0, I_m)$$

where $\hat{\Sigma}_{\theta} = (\hat{\Gamma}'Q_T\hat{\Gamma})^{-1}\hat{\Gamma}'Q_T\hat{V}_TQ_T\hat{\Gamma}(\hat{\Gamma}'Q_T\hat{\Gamma})^{-1}$, $\hat{\Gamma} = T^{-1}\sum_1^T G_t(\hat{\theta})$ and $\hat{V}_T$ is a consistent estimator of V, so that standard Student-t and Wald Statistics on stable coefficients have usual asymptotic null distribution.

In OLS example, $Q_T = I_3$, $\hat{\Gamma} = T^{-1}\sum_1^T W_tW_t'$ and $\hat{V}_T = \hat{\sigma}^2\hat{\Gamma}$, so that $\hat{\Sigma}_{\theta} = \hat{\sigma}^2\hat{\Gamma}^{-1}$. Hence

$$T^{1/2}\hat{\sigma}^{-1}\hat{\Gamma}^{1/2}\left(\begin{pmatrix}\hat{\beta}\\ \hat{\delta}\\ \hat{\mu}\end{pmatrix} - \begin{pmatrix}T^{-1}\sum_1^T\beta_t\\ \delta_0\\ \mu_0\end{pmatrix}\right) \Rightarrow \mathcal{N}(0, I_3)$$

Main result

Theorem (ctd): (ii)

$$T^{-1/2}\sum_{1}^{T} g_t(\hat{\theta}) \Rightarrow \mathcal{N}(0, AVA'),$$

where $A = (I_p - \Gamma(\Gamma' Q_0 \Gamma)^{-1}\Gamma' Q_0)$ and $\hat{\Gamma} \xrightarrow{p} \Gamma$.

(iii) Furthermore, if in addition, $T^{-1/2}\sum_{t=1}^{[\cdot T]} g_t(\theta_t) \Rightarrow V^{1/2}W(\cdot)$, then

$$T^{-1/2}\sum_{t=1}^{[\cdot T]} g_t(\hat{\theta}) \Rightarrow \zeta(\cdot)$$

where $\zeta(\lambda) = V^{1/2}W(\lambda) - \lambda\Gamma(\Gamma' Q_0 \Gamma)^{-1}\Gamma' Q_0 V^{1/2} W(1) + \Gamma\left(\int_0^\lambda f(l)dl - \lambda\int_0^1 f(l)dl\right)$ and W is a Wiener process.

Interpretation: (ii) Null distribution of overidentification test unaffected by instability and (iii) null distribution of standard stability tests concerning subset of parameters unaffected by instabilities in other parameters.

Sketch of Proof I

- By a first order Taylor expansion of the first order condition of GMM,

$$
\begin{aligned}
0 &= (T^{-1}\sum_{1}^{T} G_t(\hat{\theta}))' Q_T T^{-1/2}\sum_{1}^{T} g_t(\hat{\theta}) \\
&= \hat{\Gamma}' Q_T T^{-1/2}\sum_{1}^{T} g_t(\theta_t) + \hat{\Gamma}' Q_T (T^{-1}\sum_{1}^{T}\tilde{G}_t) T^{1/2}(\hat{\theta}-\theta_0) \\
&\qquad - \hat{\Gamma}' Q_T T^{-1}\sum_{1}^{T}\tilde{G}_t T^{1/2}(\theta_t-\theta_0) \\
&= \Gamma' Q_T T^{-1/2}\sum_{1}^{T} g_t(\theta_t) + \Gamma' Q_T \Gamma T^{1/2}(\hat{\theta} - T^{-1}\sum_{1}^{T}\theta_t) + o_p(1)
\end{aligned}
$$

  where jth row of $\tilde{G}_t$ is the jth row of G_t evaluated at some $\tilde{\theta}_{t,j}$ that lies on the line segment between θ_t and $\hat{\theta}$.

- Key insight: $T^{-1}\sum_1^T \tilde{G}_t T^{1/2}(\theta_t-\theta_0) = \Gamma T^{-1}\sum_1^T T^{1/2}(\theta_t-\theta_0) + o_p(1)$

Sketch of Proof II

- Special case $\theta_t = \theta_0 + T^{-1/2}\kappa_0 \mathbf{1}[t/T \leq \lambda] + T^{-1/2}\kappa_1 \mathbf{1}[t/T > \lambda]$ for $0 < \lambda < 1$, i.e. $f(s) = \kappa_0 \mathbf{1}[s \leq \lambda] + \kappa_1 \mathbf{1}[s > \lambda]$. Then under the assumption $\sup_{0 \leq \lambda \leq 1} \left\| T^{-1} \sum_{t=1}^{[\lambda T]} \tilde{G}_t - \lambda \Gamma \right\| \xrightarrow{p} 0$

$$
\begin{aligned}
T^{-1} \sum_1^T \tilde{G}_t T^{1/2}(\theta_t - \theta_0) &= T^{-1} \sum_1^T \tilde{G}_t f(t/T) \\
&= T^{-1} \sum_1^{[\lambda T]} \tilde{G}_t \kappa_0 + T^{-1} \sum_{[\lambda T]+1}^{T} \tilde{G}_t \kappa_1 \\
&= \Gamma \lambda \kappa_0 + \Gamma(1-\lambda)\kappa_1 + o_p(1) \\
&= \Gamma T^{-1} \sum_1^T f(t/T) + o_p(1) \\
&= \Gamma T^{-1} \sum_1^T T^{1/2}(\theta_t - \theta_0) + o_p(1)
\end{aligned}
$$

Sketch of Proof III

- Real analysis result: A bounded and piece-wise continuous function $f : [0, 1] \mapsto \mathbb{R}^m$ with at most a finite number of discontinuities can be uniformly approximated by a step function.

- Apply same argument as with single step to multiple step function to obtain

$$T^{-1} \sum_{1}^{T} \tilde{G}_t T^{1/2}(\theta_t - \theta_0) = T^{-1} \sum_{1}^{T} T^{1/2}(\theta_t - \theta_0) + o_p(1)$$

Technical Difficulties

- Models with unstable parameters tend to generate nonstationary data. Think of VAR with time varying parameters.
  $\Rightarrow$ how to argue for the high-level assumptions to hold in the unstable model?

- Ploberger and Kontrus (1989), Sowell (1996), Stock and Watson (1998): Strong assumptions on DGP that rule out VARs.

- Andrews (1993): Highly technical mixing conditions.

- Follow Andrews and Ploberger (1994) and use indirect reasoning via 'Contiguity': Make standard assumptions on likelihood of stable model, and then argue that likelihood of unstable model is close to likelihood of stable model in the limit.

Contiguity

A sequence of densities $\{f_{T,1}(y)\}_T$ is called contiguous to another sequence of densities $\{f_{T,0}(y)\}_T$ when every $o_p(1)$ random variable under the latter sequence of densities is also $o_p(1)$ under the former.

Contiguity II

Theorem (Le Cam): A sequence of densities $\{f_{T,1}\}_T$ is contiguous to a the sequence of densities $\{f_{T,0}\}_T$ if
(1) Under $f_{T,0}$, $LR_T = f_{T,1}/f_{T,0} \Rightarrow LR$
(2) $E[LR] = 1$

Intuition: LR_T describes the reweighting to get from $f_{T,0}$ probability statements to $f_{T,1}$ probability statements:

$$\begin{aligned} P_{T,0}(A_T) &= \int_{A_T} f_{T,0} d\mu_T \\ P_{T,1}(A_T) &= \int_{A_T} f_{T,1} d\mu_T = \int_{A_T} LR_T f_{T,0} d\mu_T \end{aligned}$$

$\Rightarrow$ controlling the asymptotic behavior of LR_T makes sure that whenever $P_{T,0}(A_T) \to 0$, then also $P_{T,1}(A_T) \to 0$.
(Note that $E_0 LR_T = \int LR_T f_{T,0} d\mu_T = \int f_{T,1} d\mu_T = 1$.)

Likelihood Structure

- Density of data $\{y_t\}_{t=1}^T$ is parametrized by time varying $k \times 1$ ($k \geq m$) parameter vector β.

- In unstable model, $T^{1/2}(\beta_t - \beta_0) = B(t/T)$ for some bounded and piecewise continuous vector function $B : [0, 1] \mapsto \mathbb{R}^k$ with at most a finite number of discontinuities.

- Let $\mathfrak{F}_t$ be the σ–field generated by $\{y_s\}_{s=1}^t$, and suppose the conditional density of y_t given $\mathfrak{F}_{t-1}$ with respect to μ_t is given by $f_t(y_t; \beta_t)$, so that density of data is $\prod_{t=1}^T f_t(y_t; \beta_t)$.

- Define $l_t(\beta) = \ln f_t(y_t; \beta)$, the scores $s_t(\beta) = \partial l_t(\beta)/\partial\beta$ and the Hessians $h_t(\beta) = \partial s_t(\beta)/\partial\beta'$.

Likelihood Structure II

- Under weak regularity conditions

$$
\begin{aligned}
E[s_t(\beta_t)|\mathfrak{F}_{t-1}] &= \int s_t(\beta_t) f_t(y_t;\beta_t) d\mu_t \\
&= \int \frac{\partial f_t(y_t;\beta)}{\partial \beta}|_{\beta=\beta_t} d\mu_t \\
&= \frac{\partial}{\partial \beta} \int f_t(y_t;\beta) d\mu_t|_{\beta=\beta_t} = 0
\end{aligned}
$$

  so that $\{s_t(\beta_t), \mathfrak{F}_t\}_{t=1}^T$ is a martingale difference sequence

- Similarly, $\{s_t(\beta_0)s_t(\beta_0)' + h_t(\beta_0), \mathfrak{F}_t\}_{t=1}^T$ is a martingale difference sequence

Assumptions on Likelihood of Stable Model

(i) In some neighborhood $\mathcal{B}_0$ of β_0, $l_t(\beta)$ is twice differentiable a.s. with respect to β for $t = 1, \cdots, T$.

(ii) $\{s_t(\beta_0), \mathfrak{F}_t\}$ is a square-integrable martingale difference array with $\sup_{0 \le \lambda \le 1} ||T^{-1} \sum_{t=1}^{[\lambda T]} E[s_t(\beta_0) s_t(\beta_0)' | \mathfrak{F}_{t-1}] - \int_0^\lambda \Upsilon(l) dl|| \xrightarrow{p} 0$ for some nonstochastic bounded Riemann integrable matrix function $\Upsilon : [0, 1] \mapsto \mathbb{R}^{k \times k}$, and there exists $\epsilon > 0$ such that $\sup_{t \le T, T \ge 1} E[||s_t(\beta_0)||^{2+\epsilon} | \mathfrak{F}_{t-1}] < \infty$ a.s.

(iii) $T^{-1} \sum_1^T ||h_t(\beta_0)|| = O_p(1)$, and for any decreasing neighborhood of β_0 contained in $\mathcal{B}_0$, $T^{-1} \sum_1^T \sup_{\beta \in \mathcal{B}_T} ||h_t(\beta) - h_t(\beta_0)|| \xrightarrow{p} 0$.

(iv) $\sup_{0 \le \lambda \le 1} ||T^{-1} \sum_{t=1}^{[\lambda T]} h_t(\beta_0) + \int_0^\lambda \Upsilon(l) dl|| \xrightarrow{p} 0$.

Contiguity!

Lemma: Under the stated Conditions, the unstable model is contiguous to the stable model.

Sketch of proof: From an exact Taylor expansion, under the stable model

$$\begin{aligned} LR_T &= \exp[\sum_1^T (l_t(\beta_t) - l_t(\beta_0))] \\ &= \exp[\sum_1^T s_t(\beta_0)'(\beta_t - \beta_0) + \tfrac{1}{2}\sum_1^T (\beta_t - \beta_0)' h_t(\tilde{\beta}_t)(\beta_t - \beta_0)] \\ &= \exp[T^{-1/2}\sum_1^T s_t(\beta_0)' B(t/T) + \tfrac{1}{2} T^{-1} \sum_1^T B(t/T)' h_t(\tilde{\beta}_t) B(t/T)] \\ &\Rightarrow \exp[\omega \mathcal{N}(0,1) - \tfrac{1}{2}\omega^2] \end{aligned}$$

where $\omega^2 = \int B(l)'\Upsilon(l)B(l)dl$. But $E\exp[\omega\mathcal{N}(0,1) - \frac{1}{2}\omega^2] = 1$, and contiguity follows by LeCam's Theorem.

Application of Contiguity

- With contiguity, it suffices to establish the high-level assumptions (iv)–(vii) in the stable model.

- Likelihood structure does not need to be known: Assumptions are 'regularity conditions'.

- Example: VAR with Gaussian disturbances $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$

$$y_t = \sum_{i=1}^{\ell} A_{t,i} y_{t-i} + \varepsilon_t$$

- Even under contiguity, assumption (iii) in the unstable model, i.e. $T^{-1/2} \sum_1^T g_t(\theta_t) \Rightarrow \mathcal{N}(0, V)$, does not follow from $T^{-1/2} \sum_1^T g_t(\theta_0) \Rightarrow \mathcal{N}(0, V)$ in the stable model.

Application of Contiguity II

Paper makes two further arguments that facilitate derivation of $T^{-1/2} \sum_1^T g_t(\theta_t) \Rightarrow \mathcal{N}(0, V)$ in the unstable model:

1. Under a martingale difference sequence assumption for $g_t(\theta_t)$ in the unstable model, one can exploit contiguity to establish sufficient conditions for martingale CLT, which take the form of convergences in probability.

2. If $g_t(\theta_0) = F' s_t(\beta_0)\ \forall t$ in stable model for some $k \times p$ matrix F, then CLT in unstable model follows from LeCam's Third Lemma, a change of asymptotic measure. Idea: For finite T, if we know the distribution of (Y_T, LR_T) under $f_{T,0}$, then we can determine the distribution of Y_T also under $f_{T,1}$. Same works asymptotically under contiguity.

Monte Carlo Set-up I

OLS regression $\quad y_t = X_t\beta_t + Z_t\delta + \mu + \varepsilon_t$

- $\begin{pmatrix} X_t \\ Z_t \end{pmatrix} = \rho \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + u_t$, $u_t \sim iid\mathcal{N}\left(0, \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}\right)$, $\varepsilon_t \sim iid\mathcal{N}(0,1)$

- $\beta_t = \mathbf{1}(t \geq T/2) \times hT^{-1/2}$

- 20,000 repetitions

Monte Carlo Results I

$T = 100$	coverage 95% CI		Nyblom tests		
DGP	δ	μ	all	β	δ
$h = 0, \rho = 0$	94.5%	94.3%	4.6%	4.9%	4.8%
$h = 5, \rho = 0$	94.3%	94.3%	34.7%	40.0%	4.1%
$h = 5, \rho = 0.5$	94.0%	93.0%	31.9%	36.2%	4.1%
$h = 10, \rho = 0$	94.1%	94.5%	89.7%	84.8%	2.5%
$h = 10, \rho = 0.5$	92.9%	90.1%	86.9%	81.9%	2.7%

$T = 200$	coverage 95% CI		Nyblom tests		
DGP	δ	μ	all	β	δ
$h = 0, \rho = 0$	95.0%	94.6%	4.4%	5.0%	4.7%
$h = 5, \rho = 0$	94.8%	94.9%	38.9%	43.9%	4.3%
$h = 5, \rho = 0.5$	94.6%	93.7%	38.1%	42.7%	4.4%
$h = 10, \rho = 0$	94.9%	94.5%	94.9%	92.9%	3.2%
$h = 10, \rho = 0.5$	93.7%	92.1%	83.5%	91.5%	3.4%

Monte Carlo Set-up II

Stylized New Keynesian Phillips Curve

$$\begin{aligned} \Delta\pi_t &= \phi E_t \Delta\pi_{t+1} + \kappa s_t + \varepsilon_t \\ s_t &= \rho_{1t} s_{t-1} + \rho_{2t} s_{t-2} + \xi_t \end{aligned}$$

- Driving variable s_t is unemployment gap, specified to be an AR(2).

- Solve forward and use resulting reduced form as data generating process with $(\varepsilon_t, \xi_t) \sim iid\mathcal{N}(0, \Sigma)$, $T = 160$. Unknown parameters estimated from U.S. data (1960:1 to 2000:4) using GMM, with instruments s_{t-1} and s_{t-2}.

- Time varying monetary policy induces instabilities in dynamics of driving variable (ρ_1 and ρ_2), but ϕ and κ remain stable. In Monte Carlo, discrete jumps of ρ_1 and ρ_2 in middle of sample to values estimated over Greenspan period.

Monte Carlo Results II

$T = 160$	coverage 95% CI		Nyblom tests		
DGP	ϕ	κ	all	ρ_1,ρ_2	ϕ, κ
all stable	95.2%	95.2%	5.0%	5.0%	5.0%
half-size	95.7%	94.7%	30.1%	54.4%	4.1%
full-size	95.8%	94.2%	52.2%	95.6%	4.6%

Conclusion

- Standard GMM inference for subset of stable parameters is asymptotically valid if instabilities are local. Result holds for broad range of instabilities and data generating processes.

- Technical arguments for analysis of unstable models might be of independent interest.

- Identification of stable subset often difficult. Possible guidance by economic theory. In any event, results broaden applicability of standard GMM inference to instances with time varying nuisance parameters.