Lecture 6A: General Method of Moments (GMM) — Asymptotic Theory

Vitor Possebom

EESP-FGV

Econometrics 2

Administrative

- **Recommended Reading: Hamilton's Chapter 14**
- Soft Reading: Chaussé (2021)
- Optional Reading: Hayashi's Chapters 3 and 4
- Problem Set 5 - Deadline: June 20th at 9:00 am (Be careful with deadlines here.)

Outline

Motivation

Motivation

We want to estimate a vector of parameters $\theta_0 \in \mathbb{R}^a$ from a model based on the following $r \times 1$ vector of unconditional moment conditions

$$\mathbb{E}\left[h\left(\theta_0, W_t\right)\right] = 0,$$

where W_t is $h \times 1$ vector of variables that are observed at time t.

Motivation

- GMM encompasses OLS, IV and Maximum Likelihood models.
- It also encompasses Nonlinear LS and Quantile Regression.
- It can also be used in many other cases:
  - Heteroscedasticity as a Source of Identification [Lewbel, 2012]
  - Misclassification Models [Haider and Stephens, 2020]
  - Lee Bounds [Lee, 2009]: Identifying Treatment Effects with Sample Selection
  - Demand Estimation: Berry et al. [1995]
  - R&D Investment Models: Chen and Xu [2022].
  - Macroeconomic Structural Models
  - CAPM Models
- Examples will be discussed in detail in the next lecture.

Definition

Outline

Definition

- W_t: $h \times 1$ vector of variables observed at date t.
- θ: unknown $a \times 1$ vector of coefficients
- $h : \mathbb{R}^a \times \mathbb{R}^h \to \mathbb{R}^r$: $h(\theta, W_t)$ is a vector-valued function.

Definition

- W_t: $h \times 1$ vector of variables observed at date t.
- θ: unknown $a \times 1$ vector of coefficients
- $h : \mathbb{R}^a \times \mathbb{R}^h \to \mathbb{R}^r$: $h(\theta, W_t)$ is a vector-valued function.

True value θ_0 satisfies the ortogonality conditions

$$\mathbb{E}\left[h\left(\theta_0, W_t\right)\right] = 0.$$

- $Y_T := \left(W_T', W_{T-1}', \ldots, W_1'\right)'$ is a $Th \times 1$ vector with all the observations in a sample of size T.

Definition

- $Y_T := \left(W_T', W_{T-1}', \ldots, W_1'\right)'$ is a $Th \times 1$ vector with all the observation in a sample of size T.
- $g : \mathbb{R}^a \to \mathbb{R}^r$ such that $g\left(\theta, Y_T\right) = \frac{\sum_{t=1}^{T} h\left(\theta, W_t\right)}{T}$

Definition

- $Y_T := \left(W_T', W_{T-1}', \ldots, W_1'\right)'$ is a $Th \times 1$ vector with all the observation in a sample of size T.
- $g : \mathbb{R}^a \to \mathbb{R}^r$ such that $g\left(\theta, Y_T\right) = \dfrac{\sum_{t=1}^{T} h\left(\theta, W_t\right)}{T}$

GMM estimator $\hat{\theta}_T$ satisfies

$$\hat{\theta}_T := \underset{\theta \in \mathbb{R}^a}{\operatorname{argmin}} \left[g\left(\theta, Y_T\right)\right]' W_T \left[g\left(\theta, Y_T\right)\right],$$

where $\{W_T\}_{T=1}^{+\infty}$ is a sequence of $(r \times r)$ positive definite weighting matrices.

Consistency

Outline

Consistency

If W_t is strictly stationary and weakly dependent and $h(\cdot,\cdot)$ is continuous, a Law of Large Number holds:

$$g(\theta, Y_T) \xrightarrow{p} \mathbb{E}\left[h(\theta, W_t)\right].$$

Consistency

If W_t is strictly stationary and weakly dependent and $h(\cdot,\cdot)$ is continuous, a Law of Large Number holds:

$$g(\theta, Y_T) \xrightarrow{p} \mathbb{E}\left[h(\theta, W_t)\right].$$

Additionally, if θ_0 is the only value that satisfies the ortogonality conditions, then

$$\hat{\theta}_T \xrightarrow{p} \theta_0.$$

Consistency

If W_t is strictly stationary and weakly dependent and $h(\cdot,\cdot)$ is continuous, a Law of Large Number holds:

$$g(\theta, Y_T) \xrightarrow{p} \mathbb{E}\left[h(\theta, W_t)\right].$$

Additionally, if θ_0 is the only value that satisfies the ortogonality conditions, then

$$\hat{\theta}_T \xrightarrow{p} \theta_0.$$

OBS: If you are interested in the formal assumptions behind this result and the other results in this lecture, I can send you some extra material.

Optimal Weighting Matrix

Outline

Optimal Weighting Matrix

- Asymptotic variance of the sample mean of $h(\theta_0, W_t)$:

$$S = \lim_{T \to +\infty} \left\{ T \cdot \mathbb{E}\left[\left(g\left(\theta_0, Y_T\right)\right) \left(g\left(\theta_0, Y_T\right)\right)' \right] \right\}$$

By choosing W_T smartly, we can minimize the asymptotic variance of the GMM estimator.

Optimal Weighting Matrix

The unfeasible optimally weighted GMM estimator $\hat{\theta}_T$ satisfies

$$\hat{\theta}_T := \underset{\theta \in \mathbb{R}^a}{\operatorname{argmin}} \left[g\left(\theta, Y_T\right)\right]' S^{-1} \left[g\left(\theta, Y_T\right)\right].$$

OBS: In his Chapter 14, Wooldridge has a detailed discussion on the GMM estimator's asymptotic variance.

Optimal Weighting Matrix

Assuming that $h\left(\theta_0, W_t\right)$ is serially uncorrelated and that $\hat{\theta}_T$ is any consistent estimator of θ_0, we can estimate the optimal weighting matrix:

$$\hat{S}_T := \frac{\sum_{t=1}^{T}\left\{\left[h\left(\hat{\theta}_T, W_t\right)\right]\left[h\left(\hat{\theta}_T, W_t\right)\right]'\right\}}{T} \xrightarrow{p} S.$$

Optimal Weighting Matrix

Assuming that $h(\theta_0, W_t)$ is serially uncorrelated and that $\hat{\theta}_T$ is any consistent estimator of θ_0, we can estimate the optimal weighting matrix:

$$\hat{S}_T := \frac{\sum_{t=1}^{T} \left\{ \left[h\left(\hat{\theta}_T, W_t\right)\right] \left[h\left(\hat{\theta}_T, W_t\right)\right]' \right\}}{T} \xrightarrow{p} S.$$

Problem! The optimally weighted GMM estimator's definition seems circular!

Optimal Weighting Matrix

Assuming that $h(\theta_0, W_t)$ is serially uncorrelated and that $\hat{\theta}_T$ is any consistent estimator of θ_0, we can estimate the optimal weighting matrix:

$$\hat{S}_T := \frac{\sum_{t=1}^{T} \left\{ \left[h\left(\hat{\theta}_T, W_t\right) \right] \left[h\left(\hat{\theta}_T, W_t\right) \right]' \right\}}{T} \xrightarrow{p} S.$$

Problem! The optimally weighted GMM estimator's definition seems circular!

Solution:

Optimal Weighting Matrix

Assuming that $h(\theta_0, W_t)$ is serially uncorrelated and that $\hat{\theta}_T$ is any consistent estimator of θ_0, we can estimate the optimal weighting matrix:

$$\hat{S}_T := \frac{\sum_{t=1}^{T} \left\{ \left[h\left(\hat{\theta}_T, W_t\right) \right] \left[h\left(\hat{\theta}_T, W_t\right) \right]' \right\}}{T} \xrightarrow{p} S.$$

Problem! The optimally weighted GMM estimator's definition seems circular!

Solution: Iterative Process

Two-step and Iterative GMM Estimators

Two-step and Iterative GMM Estimators

1. Start with $\hat{\theta}_T^{(0)} := \text{argmin}_{\theta \in \mathbb{R}^a} \left[g\left(\theta, Y_T\right) \right]' \left[g\left(\theta, Y_T\right) \right].$

Two-step and Iterative GMM Estimators

1. Start with $\hat{\theta}_T^{(0)} := \text{argmin}_{\theta \in \mathbb{R}^a} \left[g\left(\theta, Y_T\right) \right]' \left[g\left(\theta, Y_T\right) \right].$

2. Compute $\hat{S}_T^{(0)} := \dfrac{\sum_{t=1}^{T} \left\{ \left[h\left(\hat{\theta}_T^{(0)}, W_t \right) \right] \left[h\left(\hat{\theta}_T^{(0)}, W_t \right) \right]' \right\}}{T}.$

Two-step and Iterative GMM Estimators

1. Start with $\hat{\theta}_T^{(0)} := \text{argmin}_{\theta \in \mathbb{R}^a} \left[g\left(\theta, Y_T\right)\right]' \left[g\left(\theta, Y_T\right)\right].$

2. Compute $\hat{S}_T^{(0)} := \dfrac{\sum_{t=1}^{T} \left\{ \left[h\left(\hat{\theta}_T^{(0)}, W_t\right)\right] \left[h\left(\hat{\theta}_T^{(0)}, W_t\right)\right]' \right\}}{T}.$

3. Two-Step GMM estimator:
$\hat{\theta}_T^{(1)} := \text{argmin}_{\theta \in \mathbb{R}^a} \left\{ \left[g\left(\theta, Y_T\right)\right]' \left[\hat{S}_T^{(0)}\right]^{-1} \left[g\left(\theta, Y_T\right)\right] \right\}.$

Two-step and Iterative GMM Estimators

1. Start with $\hat{\theta}_T^{(0)} := \text{argmin}_{\theta \in \mathbb{R}^a} [g(\theta, Y_T)]' [g(\theta, Y_T)]$.

2. Compute $\hat{S}_T^{(0)} := \dfrac{\sum_{t=1}^T \left\{ \left[h\left(\hat{\theta}_T^{(0)}, W_t \right) \right] \left[h\left(\hat{\theta}_T^{(0)}, W_t \right) \right]' \right\}}{T}$.

3. Two-Step GMM estimator:
$\hat{\theta}_T^{(1)} := \text{argmin}_{\theta \in \mathbb{R}^a} \left\{ [g(\theta, Y_T)]' \left[\hat{S}_T^{(0)} \right]^{-1} [g(\theta, Y_T)] \right\}$.

4. Iterative GMM estimator: Iterate until $\hat{\theta}_T^{(j)} \approx \hat{\theta}_T^{(j+1)}$

Two-step and Iterative GMM Estimators

Both estimators have the same asymptotic distribution.

Iterative GMM estimator is invariant with respect to the scale of the data and to the initial weighting matrix for W_T.

Serial Correlation

Serial Correlation

So far, we have assumed that $\{h(\theta_0, W_t)\}_{t=-\infty}^{+\infty}$ is serially uncorrelated.

Serial Correlation

So far, we have assumed that $\{h(\theta_0, W_t)\}_{t=-\infty}^{+\infty}$ is serially uncorrelated.

But, frequently, $\{h(\theta_0, W_t)\}_{t=-\infty}^{+\infty}$ is serially correlated.

Serial Correlation

So far, we have assumed that $\{h(\theta_0, W_t)\}_{t=-\infty}^{+\infty}$ is serially uncorrelated.

But, frequently, $\{h(\theta_0, W_t)\}_{t=-\infty}^{+\infty}$ is serially correlated.

We can use variance estimators that take serial correlation into account: Newey-West (1987), Gallant (1987), Andrews (1991) and Andrews & Monahan (1992).

`R`'s default is Andrews' (1991) estimator.

Asymptotic Distribution of the GMM Estimator

Outline

Asymptotic Distribution of the GMM Estimator

Define $\hat{\theta}_T := \text{argmin}_{\theta \in \mathbb{R}^a} \left\{ [g(\theta, Y_T)]' \left[\hat{S}_T\right]^{-1} [g(\theta, Y_T)] \right\}$ with $\hat{S}_T$ regarded as fixed with respect to θ and $\hat{S}_T \xrightarrow{p} S$.

Define $\hat{\theta}_T := \text{argmin}_{\theta \in \mathbb{R}^a} \left\{ [g(\theta, Y_T)]' \left[\hat{S}_T\right]^{-1} [g(\theta, Y_T)] \right\}$ with $\hat{S}_T$ regarded as fixed with respect to θ and $\hat{S}_T \xrightarrow{p} S$.

Consequently, $\hat{\theta}_T$ is the solution to the following system of nonlinear equations:

$$\underbrace{\left\{ \left. \frac{\partial g(\theta, Y_T)}{\partial \theta'} \right|_{\theta = \hat{\theta}_T} \right\}'}_{a \times r} \times \underbrace{\left[\hat{S}_T\right]^{-1}}_{r \times r} \times \underbrace{g(\theta, Y_T)}_{r \times 1} = \underbrace{0}_{a \times 1} \tag{1}$$

Asymptotic Distribution of the GMM Estimator

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

[1] Strict stationarity, weak dependence, continuity and restriction on higher moments is enough to ensure that a Central Limit Theorem holds.

Asymptotic Distribution of the GMM Estimator

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

Let $g(\theta, Y_T)$ be differentiable in θ for all Y_T and let $\hat{\theta}_T$ be the GMM estimator satisfying Equation (1) with $r \geq a$.

[1] Strict stationarity, weak dependence, continuity and restriction on higher moments is enough to ensure that a Central Limit Theorem holds.

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

Let $g(\theta, Y_T)$ be differentiable in θ for all Y_T and let $\hat{\theta}_T$ be the GMM estimator satisfying Equation (1) with $r \geq a$. Let $\left\{\hat{S}_T\right\}_{T=1}^{+\infty}$ be a sequence of positive definite $r \times r$ matrices such that $\hat{S}_T \xrightarrow{p} S$, with S positive definite.

[1]Strict stationarity, weak dependence, continuity and restriction on higher moments is enough to ensure that a Central Limit Theorem holds.

Asymptotic Distribution of the GMM Estimator

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

Let $g(\theta, Y_T)$ be differentiable in θ for all Y_T and let $\hat{\theta}_T$ be the GMM estimator satisfying Equation (1) with $r \geq a$. Let $\left\{\hat{S}_T\right\}_{T=1}^{+\infty}$ be a sequence of positive definite $r \times r$ matrices such that $\hat{S}_T \xrightarrow{p} S$, with S positive definite. Suppose, further, that the following hold:

1. $\hat{\theta}_T \xrightarrow{p} \theta_0$;

[1] Strict stationarity, weak dependence, continuity and restriction on higher moments is enough to ensure that a Central Limit Theorem holds.

Asymptotic Distribution of the GMM Estimator

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

Let $g(\theta, Y_T)$ be differentiable in θ for all Y_T and let $\hat{\theta}_T$ be the GMM estimator satisfying Equation (1) with $r \geq a$. Let $\left\{\hat{S}_T\right\}_{T=1}^{+\infty}$ be a sequence of positive definite $r \times r$ matrices such that $\hat{S}_T \xrightarrow{p} S$, with S positive definite. Suppose, further, that the following hold:

1. $\hat{\theta}_T \xrightarrow{p} \theta_0$;
2. $\sqrt{T} \cdot g(\theta_0, Y_T) \xrightarrow{d} N(0, S)$;[1] and

[1] Strict stationarity, weak dependence, continuity and restriction on higher moments is enough to ensure that a Central Limit Theorem holds.

Asymptotic Distribution of the GMM Estimator

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

3. for any sequence $\{\theta_T^*\}_{T=1}^{+\infty}$ satisfying $\theta_T^* \xrightarrow{p} \theta_0$, it is the case that

$$\text{plim}\left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\theta_T^*} \right\} = \text{plim}\left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\theta_0} \right\} =: \underbrace{D'}_{r\times a},$$

with the columns of D' linearly independent.

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

3. for any sequence $\{\theta_T^*\}_{T=1}^{+\infty}$ satisfying $\theta_T^* \xrightarrow{p} \theta_0$, it is the case that

$$\text{plim}\left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\theta_T^*} \right\} = \text{plim}\left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\theta_0} \right\} =: \underbrace{D'}_{r \times a},$$

with the columns of D' linearly independent.

Then,

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{L} N\left(0, \left[D\, S^{-1}\, D'\right]^{-1}\right).$$

Asymptotic Distribution of the GMM Estimator

Proposition 1 (Asymptotic Distribution of the GMM Estimator)

3. for any sequence $\{\theta_T^*\}_{T=1}^{+\infty}$ satisfying $\theta_T^* \xrightarrow{p} \theta_0$, it is the case that

$$\operatorname{plim}\left\{ \left.\frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'}\right|_{\theta=\theta_T^*} \right\} = \operatorname{plim}\left\{ \left.\frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'}\right|_{\theta=\theta_0} \right\} =: \underbrace{D'}_{r \times a},$$

   with the columns of D' linearly independent.

   Then,

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{L} N\left(0, \left[D\, S^{-1}\, D'\right]^{-1}\right).$$

Proof

Asymptotic Distribution of the GMM Estimator

We can estimate D' using

$$\underbrace{\hat{D}'_T}_{r \times a} := \left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta = \hat{\theta}_T} \right\}.$$

Testing Overidentifying Restrictions

Outline

Testing Overidentifying Restrictions

When the number of orthogonality conditions exceeds the number of parameters to be estimated ($r > a$), the model is overidentified.[2]

[2] It is easy to understand overidentification in a parametric model. However, it is possible to discuss overidentification in nonparametric models too: Chen and Santos [2018].

Testing Overidentifying Restrictions

When the number of orthogonality conditions exceeds the number of parameters to be estimated ($r > a$), the model is overidentified.[2]

In this case, we can test whether all of the sample moments $g\left(\hat{\theta}_T, Y_T\right)$ are as close to zero as would be expected if the corresponding population moments $\mathbb{E}\left[h\left(\theta_0, W_t\right)\right]$ were truly zero.

[2] It is easy to understand overidentification in a parametric model. However, it is possible to discuss overidentification in nonparametric models too: Chen and Santos [2018].

Testing Overidentifying Restrictions

We have that

$$\left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right]' \hat{S}_T^{-1} \left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right] \xrightarrow{d} \chi^2_{r-a}.$$

Testing Overidentifying Restrictions

We have that

$$\left[\sqrt{T}\cdot g\left(\hat{\theta}_T, Y_T\right)\right]' \hat{S}_T^{-1}\left[\sqrt{T}\cdot g\left(\hat{\theta}_T, Y_T\right)\right] \xrightarrow{d} \chi^2_{r-a}.$$

Is this test the solution to all identification problems?

Testing Overidentifying Restrictions

We have that

$$\left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right]' \hat{S}_T^{-1} \left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right] \xrightarrow{d} \chi^2_{r-a}.$$

Is this test the solution to all identification problems?

Of course not! This test is underpowered against many types of misspecification.

Testing Overidentifying Restrictions

We have that

$$\left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right]' \hat{S}_T^{-1} \left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right] \xrightarrow{d} \chi^2_{r-a}.$$

Is this test the solution to all identification problems?

Of course not! This test is underpowered against many types of misspecification.

What can we do?

Testing Overidentifying Restrictions

We have that

$$\left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right]' \hat{S}_T^{-1} \left[\sqrt{T} \cdot g\left(\hat{\theta}_T, Y_T\right)\right] \xrightarrow{d} \chi^2_{r-a}.$$

Is this test the solution to all identification problems?

Of course not! This test is underpowered against many types of misspecification.

What can we do? Robustness checks, placebo tests, theoretical models, mechanisms discussion etc.

Thank you!

Contact Information:

Vitor Possebom

E-mail: vitor.possebom@fgv.br

Website: `sites.google.com/site/vitorapossebom/`

References

S. Berry, J. Levinsohn, and A. Pakes. Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841–890, 1995. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2171802.

X. Chen and A. Santos. Overidentification in Regular Models. *Econometrica*, 86(5):pp. 1771–1817, 2018.

Y. Chen and D. Xu. A Structural Empirical Model of R&D Investment, Firm Heterogeneity, and Industry Evolution. NBER Working Paper n. 29733. Available at https://www.nber.org/papers/w29733., Feb. 2022.

S. Haider and M. Stephens. Correcting for Misclassified Binary Regressors using Instrumental Variables. NBER Working Paper n. 27797. Available at `nber.org/papers/w27797.`, Sept. 2020.

D. S. Lee. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, 76:pp. 1071–1102, 2009.

A. Lewbel. Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. *Journal of Business and Economic Statistics*, 30(1):p. 67–80, 2012.

Proof of Proposition 1

Let $g_i(\theta, Y_T)$ denote the i-th element of $g(\theta, Y_T)$, so that $g_i : \mathbb{R}^a \to \mathbb{R}$. By the mean-value theorem,

$$g_i\left(\hat{\theta}_T, Y_T\right) = g_i(\theta_0, Y_T) + \left[d_i\left(\theta^*_{i,T}, Y_T\right)\right]'\left(\hat{\theta}_T - \theta_0\right) \tag{2}$$

for some $\theta^*_{i,T}$ between θ_0 and $\hat{\theta}_T$, where $d_i\left(\theta^*_{i,T}, Y_T\right) := \left.\dfrac{\partial g_i(\theta, Y_T)}{\partial \theta}\right|_{\theta=\theta^*_{i,T}}$ is a $a \times 1$ vector.

Proof of Proposition 1

Note that $D_T' := \begin{bmatrix} \left[d_1\left(\theta_{i,T}^*, Y_T\right)\right]' \\ \left[d_2\left(\theta_{i,T}^*, Y_T\right)\right]' \\ \vdots \\ \left[d_r\left(\theta_{i,T}^*, Y_T\right)\right]' \end{bmatrix}$ is a $r \times a$ matrix and stack all the scalars in Equation (2) to produce the following $r \times 1$ vector:

$$g\left(\hat{\theta}_T, Y_T\right) = g\left(\theta_0, Y_T\right) + D_T'\left(\hat{\theta}_T - \theta_0\right). \tag{3}$$

Proof of Proposition 1

Now, let's pre-multiply everything by the following $a \times r$ matrix,

$$\left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\hat{\theta}_T} \right\}' \times \hat{S}_T^{-1},$$

because we want something that looks like Equation (1).

Proof of Proposition 1

We have that

$$\left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\hat{\theta}_T} \right\}' \times \hat{S}_T^{-1} \times g\left(\hat{\theta}_T, Y_T\right)$$

$$= \left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\hat{\theta}_T} \right\}' \times \hat{S}_T^{-1} \times g\left(\theta_0, Y_T\right)$$

$$+ \left\{ \left. \frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'} \right|_{\theta=\hat{\theta}_T} \right\}' \times \hat{S}_T^{-1} \times D_T' \left(\hat{\theta}_T - \theta_0\right).$$

Proof of Proposition 1

According to Equation (1), the left-hand side of this equation is zero. So we can rearrange everything to get

$$\left(\hat{\theta}_T - \theta_0\right) = -\left[\left\{ \left.\frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'}\right|_{\theta=\hat{\theta}_T} \right\}' \times \hat{S}_T^{-1} \times D_T'\right]^{-1}$$

$$\times \left\{ \left.\frac{\partial g\left(\theta, Y_T\right)}{\partial \theta'}\right|_{\theta=\hat{\theta}_T} \right\}' \times \hat{S}_T^{-1} \times g\left(\theta_0, Y_T\right).$$

Now, note that $\theta_{i,T}^*$ inside D_T' is between θ_0 and $\hat{\theta}_T$, implying, by condition 1, that $\theta_{i,T}^* \xrightarrow{p} \theta_{0,i}$ for each i. Thus, condition 3 ensures that each row of D_T' converges in probability to the corresponding row of D'.

Consequently, the last equation implies that

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{p} -\left\{DS^{-1}D'\right\}^{-1} \times DS^{-1}\sqrt{T} \cdot g\left(\theta_0, Y_T\right).$$

Proof of Proposition 1

Define $C := -\left\{DS^{-1}D'\right\}^{-1} \times DS^{-1}$ and rewrite the last equation as

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{p} C\sqrt{T} \cdot g\left(\theta_0, Y_T\right).$$

Since condition 2 states that $\sqrt{T} \cdot g\left(\theta_0, Y_T\right) \xrightarrow{d} N\left(0, S\right)$, Slutsky Theorem implies that

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left(0, V\right),$$

where $V := CSC' = \left[D\, S^{-1}\, D'\right]^{-1}$.

Back