Lecture Notes

Generalized Method of Moments

Based on notes and book by Alastair Hall (NCSU)

Lecture outline:

1. **Introduction:**

2. GMM estimation in correctly specified non-linear dynamic models - part I.

3. GMM estimation in correctly specified non-linear dynamic models - part II.

4. Hypothesis testing.

5. Finite sample behaviour and moment selection.

Today:

- Why GMM?
- Statistical antecedents.
- Contemporay example.
- IV estimator in linear regression model.

1. Introduction

Hansen (1982, *Econometrica*) introduced the Generalized Method of Moments (GMM) estimator.

Two advantages:

- Provides a general framework for considering issues of statistical inference because it encompasses many estimators of interest in econometrics.

- Provides a computationally convenient method for the estimation of nonlinear dynamic models without complete specification of the probability distribution of the data.

1.1 Statistical antecedents:

Method of Moments (Pearson,1893, 1894, 1895) Suppose we wish to estimate the population mean, μ, and the population variance, σ^2, of a r.v. v_t.

These two parameters satisfy the *population moment conditions*

$$
\begin{aligned}
E[v_t] - \mu &= 0 \\
E[v_t^2] - (\sigma^2 + \mu^2) &= 0
\end{aligned}
$$

Pearson's method based on analogous sample moment conditions.

$$
\begin{aligned}
T^{-1} \sum_{t=1}^{T} v_t - \hat{\mu} &= 0 \\
T^{-1} \sum_{t=1}^{T} v_t^2 - (\hat{\sigma}^2 + \hat{\mu}^2) &= 0
\end{aligned}
$$

This implies

$$\widehat{\mu} = T^{-1}\sum_{t=1}^{T} v_t$$

$$\widehat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} (v_t - \widehat{\mu})^2$$

Key idea: population moment conditions provide information upon which estimation of parameters can be based.

Minimum chi-square (Neyman and Pearson, 1928)

- Outcome of an experiment lies in one of k mutually exclusive and exhaustive groups.

- $p_i =$ P(outcome lies in the i^{th} group).

- $H_0 : p_i \ = \ h(i, \theta_0)$

For given θ_0, Karl Pearson had shown that under H_0

$$GF_T(\theta_0) \ = \ T \sum_{i=1}^{k} \frac{[\widehat{p}_i \ - h(i;\theta_0)]^2}{\widehat{p}_i} \xrightarrow{d} \chi^2_{k-1}$$

Now suppose θ_0 unknown.

To perform inference need estimate of θ_0

$$\widehat{\theta}_T \ = \ argmin \, GF_T(\theta)$$

- *minimum chi-square estimator*

Connection to moments:

Let $\{D_t(i); i = 1, 2, \ldots k;\ t = 1, 2, \ldots T\}$ satisfy:

$D_t(i) = 1$ if t^{th} outcome in i^{th} group
$D_t(i) = 0$ else

$$\Rightarrow \qquad \begin{aligned} P(D_t(i) = 1) &= h(i; \theta_0) \\ E[D_t(i)] &= h(i; \theta_0) \end{aligned}$$

$\Rightarrow k$ population moment conditions

$$E \begin{bmatrix} D_t(1) - h(1; \theta_0) \\ D_t(2) - h(2; \theta_0) \\ . \\ . \\ D_t(k) - h(k; \theta_0) \end{bmatrix} = 0$$

Sample analogs are given by

$$\begin{bmatrix} \widehat{p}_1 - h(1; \theta) \\ \widehat{p}_2 - h(2; \theta) \\ . \\ . \\ \widehat{p}_k - h(k; \theta) \end{bmatrix} = 0$$

$GF_T(\theta) \;=\; T\times$

$$\begin{bmatrix} \widehat{p}_1 - h(1;\theta) \\ \widehat{p}_2 - h(2;\theta) \\ . \\ . \\ \widehat{p}_k - h(k;\theta) \end{bmatrix}' \begin{bmatrix} \widehat{p}_1^{-1} & 0 & . & . & 0 \\ 0 & \widehat{p}_2^{-1} & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \widehat{p}_k^{-1} \end{bmatrix} \begin{bmatrix} \widehat{p}_1 - h(1;\theta) \\ \widehat{p}_2 - h(2;\theta) \\ . \\ . \\ \widehat{p}_k - h(k;\theta) \end{bmatrix}$$

$\Rightarrow$ $GF_T(\theta) =$ a quadratic form in the sample-moment condition.

minimum chi-square estimator $= \theta$ which is *closest* to solving the sample moment conditions in the metric of $GF_T(\theta)$.

Instrumental variables (Wright, 1925, 1928)
Simple agricultural demand and supply system:

$$\begin{aligned} q_t^D &= \alpha p_t + u_t^D \\ q_t^S &= \beta_1 n_t + \beta_2 p_t + u_t^S \\ q_t^D &= q_t^S = q_t \end{aligned} \tag{1}$$

- q_t^D, q_t^S represent demand and supply in year t.

- p_t is the price of the commodity in year t

- q_t equals quantity produced

Problem: how to estimate α given $\{q_t, p_t\}$.

Ordinary Least Squares (OLS) regression of q_t on p_t runs into problems.

Wright's solution:

Find: z_t^D such that $Cov(z_t^D, u_t^D) = 0$.

Then (1) $\Rightarrow$

$$Cov(z_t^D, q_t) - \alpha_1 Cov(z_t^D, p_t) = 0 \qquad (2)$$

and so if $E[u_t^D] = 0$ then

$$E[z_t^D q_t] - \alpha_1 E[z_t^D p_t] = 0 \qquad (3)$$

Pearson's Method of Moments principle leads to:

$$\hat{\alpha}_1 = \sum_{t=1}^{T} z_t^D q_t / \sum_{t=1}^{T} z_t^D p_t \qquad (4)$$

Instrumental variables (IV) estimator with z_t^D known as "instrument".

Generalized Method of Moments

Population Moment Condition:

$$E[f(v_t, \theta_0)] = 0 \qquad \text{for all } t$$

Generalized Method of Moments Estimator:

$$\widehat{\theta}_T \; = \; argmin_{\theta \in \Theta} Q_T(\theta)$$

where

$$Q_T(\theta) \; = \; T^{-1} \sum_{t=1}^{T} f(v_t, \theta)' W_T T^{-1} \sum_{t=1}^{T} f(v_t, \theta)$$

and W_T is psd and $W_T \xrightarrow{p} W$, pd.

- $W_T \sim$ psd $\Rightarrow Q_T(\theta) \geq 0$ and $Q_T(\widehat{\theta}_T) = 0$ if $T^{-1} \sum_{t=1}^{T} f(v_t, \widehat{\theta}_T) = 0$

- $W \sim$ pd $\Rightarrow Q_T(\widehat{\theta}_T) = 0$ *iff* $T^{-1} \sum_{t=1}^{T} f(v_t, \widehat{\theta}_T) = 0$ in the limit as $T \to \infty$.

1.2 Contemporary example

Consumption based asset pricing model: Hansen & Singleton (1982)

A represenative agent chooses consumption to maximize his/her expected discounted utility

$$E[\sum_{i=0}^{\infty} \delta^i U(c_{t+i})|\text{-}\ _t]$$

subject to

$$c_t + p_t q_t = r_t q_{t-1} + w_t$$

for all t, where

- r_{t+1} is the return on the asset in period $t + 1$

- p_t is the price of the asset in period t

- c_t is consumption in period t

- $-_t$ is the information set available to the agent in period t

Optimal path of consumption and investment satisfies

$$p_t U'(c_t) = \delta E[r_t U'(c_{t+1}) | -_t] \qquad (5)$$

for all t.

Equation (5) can be rewritten as

$$E[\delta(r_{t+1}/p_t)\{U'(c_{t+1})/U'(c_t)\}|\Omega_t] - 1 = 0 \quad (6)$$

Hansen & Singleton (1982) set

$$U(c_t) = c_t^{\gamma}/\gamma$$

and so (6) becomes

$$E[\delta(r_{t+1}/p_t)(c_{t+1}/c_t)^{\gamma-1}|\Omega_t] - 1 = 0 \qquad (7)$$

Two parameters to be estimated: (γ, δ).

Model implies following population moment condition:
Let

$$e_t(\gamma,\delta) \;=\; \delta(r_{t+1}/p_t)(c_{t+1}/c_t)^{\gamma-1} - 1$$

then (7) $\Rightarrow$ for any vector $z_t \in \Omega_t$

$$E[e_t(\gamma,\delta)\, z_t] = E\,[\,E[e_t(\gamma,\delta)\,|\Omega_t]\, z_t] = 0$$

1.3 Instrumental variables estimation of the static linear model

- Population moment condition and identification

- Estimator

- Identifying and overidentifying restrictions

- Asymptotic properties of the estimator

- Covariance matrix estimation

- Two step estimator

- Model specification test

Consider:

$$y_t = x_t'\theta_0 + u_t, \qquad t = 1, 2, \ldots T$$

- x_t is a $(p \times 1)$ vector of observed explanatory variables;

- y_t is scalar, observed;

- u_t is the unobserved error term;

- z_t is a $(q \times 1)$ vector of instruments.

Problem: to estimate θ_0.

(i) Population moment condition and identification:

Assumption 2.1: Strict Stationarity
The random vector $v_t = (x_t', z_t', u_t)'$ is a strictly stationary process.

Let $u_t(\theta) = y_t - x_t'\theta$.

Assumption 2.2: Population Moment Condition
$E[z_t u_t(\theta_0)] = 0$.

Also need: *Identification condition:*
$E[z_t u_t(\theta)] \neq 0$ for all $\theta \neq \theta_0$

$$\begin{aligned} E[z_t u_t(\theta)] &= E[z_t u_t(\theta_0)] + E[z_t x_t'](\theta_0 - \theta) \\ &= E[z_t x_t'](\theta_0 - \theta) \end{aligned}$$

Assumption 2.3: Identification Condition
$rank\{E[z_t x_t']\} = p$.

Relationship between q and p important:

- $q < p$ - θ_0 is under- (or un-) identified

- $q = p$ - θ_0 is just identified

- $q > p$ - θ_0 is over- identified *(ii) The estimator:*

$$\widehat{\theta}_T = \text{argmin}_{\theta \in \Theta} Q_T(\theta)$$

where

$$Q_T(\theta) = \{T^{-1} u(\theta)' Z\} W_T \{T^{-1} Z' u(\theta)\}$$

First order conditions:

$$(T^{-1} X' Z) W_T (T^{-1} Z' y) = (T^{-1} X' Z) W_T (T^{-1} Z' X) \widehat{\theta}_T$$

$\Rightarrow$

$$\widehat{\theta}_T = \{X' Z W_T Z' X\}^{-1} X' Z W_T Z' y$$

(iii) Identifying and overidentifying restrictions:
Return to first order conditions:

$$(T^{-1}X'Z)W_T T^{-1}Z'u(\hat{\theta}_T) = 0$$

$\Rightarrow$ GMM = MM based on:

$$E[x_t z_t']WE[z_t u_t(\theta_0)] = 0$$

- $q = p$: GMM = MM based on $E[z_t u_t(\theta_0)] = 0$

- $q > p$: GMM sets p linear combinations of $E[z_t u_t(\theta_0)]$ equal 0

Rewrite population FOC:

$$F'W^{1/2}E[z_t u_t(\theta_0)] = 0$$

where $F' = E[x_t z_t']W^{1/2'} \Rightarrow$

$$F(F'F)^{-1}F'W^{1/2}E[z_t u_t(\theta_0)] = 0$$

$\sim$ *identifying restrictions*

Remainder is

$$(I_q - F(F'F)^{-1}F')W^{1/2}E[z_t u_t(\theta_0)] = 0$$

$\sim$ *overidentifying restrictions*.

Sample analogs:

- Identifying restrictions satisfied at $\hat{\theta}_T$
- Overidentifying restrictions: $W_T^{1/2}T^{-1}Z'u(\hat{\theta}_T)$

Now,

$$Q_T(\hat{\theta}_T) = \|W_T^{1/2}T^{-1}Z'u(\hat{\theta}_T)\|$$

$\Rightarrow Q_T(\hat{\theta}_T)$ measures how far the sample is from satisfying the overidentifying restrictions.

(iv) Asymptotic properties:

It can be shown that:

- $\widehat{\theta}_T$ is consistent for θ_0

- $(\widehat{\theta}_{T,i} - \theta_{0,i})/\sqrt{\widehat{V}_{T,ii}/T} \overset{a}{\sim} N(0,1)$ where

  - $\widehat{V}_T = (X'ZW_TZ'X)^{-1}X'Z\widehat{S}Z'X(X'ZW_TZ'X)^{-}$

  - $\widehat{S}_T \xrightarrow{p} lim_{T\to\infty}Var[T^{-1/2}Z'u]$

(v) Covariance matrix estimation:

For this model:

$$
\begin{aligned}
Var[T^{-1/2}Z'u] &= E[\{T^{-1/2}\sum_{t=1}^{T} z_t u_t\}\{T^{-1/2}\sum_{t=1}^{T} z_t u_t
\\
&= T^{-1}\sum_{t=1}^{T} E[u_t^2 z_t z_t^{'}]
\end{aligned}
$$

Therefore,

$$
\hat{S}_T = T^{-1}\sum_{t=1}^{T} u(\hat{\theta}_T)^2 z_t z_t^{'}
$$

(vi) Two step estimator:

Notice that asymptotic variance depends on weighting matrix.

Optimal choice: $W_T = \widehat{S}_T^{-1}$

Problem: Need $\widehat{\theta}_T$ to construct $\widehat{S}_T$

Two–step procedure:

1. Estimate with sub–optimal $W_T \rightarrow \widehat{\theta}_T(1) \rightarrow \widehat{S}_T(1)$.

2. Estimate with $W_T = \widehat{S}_T(1)^{-1}$.

(vii) Model specification test:

- Identifying restrictions satisfied in sample regardless of whether model is correct.

- Overidentifying restrictions not imposed in sample.

The *overidentifying restrictions test*

$J_T = TQ_T(\widehat{\theta}_T) = T^{-1/2} u(\widehat{\theta}_T)' Z \, \widehat{S}_T^{-1} \, T^{-1/2} Z' u(\widehat{\theta}_T)$

Under $H_0 : E[z_t u_t(\theta_0)] = 0$

$$J_T \xrightarrow{d} \chi^2_{q-p}$$

Lecture outline:

1. Introduction

2. **GMM estimation in correctly specified nonlinear dynamic models - part I.**

3. GMM estimation in correctly specified nonlinear dynamic models - part II.

4. Hypothesis testing.

5. Finite sample behaviour and moment selection.

Today:

- Identification
- The Estimator
- Identifying and overidentifying restrictions
- Asymptotic properties
- Covariance Matrix estimation

1. Population moment condition and identification

First impose condition on data:

Strict Stationarity
The $(r \times 1)$ random vectors $\{v_t; -\infty < t < \infty\}$ form a strictly stationary process with sample space $\mathbf{V} \subseteq \Re^r$.

Population Moment Condition
Let θ_0 be a vector of unknown parameters which are to be estimated, v_t be a vector of random variables and $f(.)$ a vector of functions then a population moment condition takes the form

$$E[f(v_t, \theta_0)] = 0 \qquad \text{for all } t. \qquad (1)$$

Global Identification

The parameter vector θ_0 is globally identified by the population moment condition in Assumption 3.3 if and only if $E[f(v_t,\ \bar{\theta})] \neq 0$ for all $\bar{\theta} \in \Theta$ such that $\bar{\theta} \neq \theta_0$.

"global" $\Rightarrow$ pmc only holds at one value in the *entire* parameter space.

Identification failures can sometimes be diagnozed, but it is often difficult.

Example: Eichenbaum's (1989) Model for Inventory Holdings by Firms

Population moment condition takes form

$$E[z_t h_{t+2}(\theta_0)] = 0$$

where

$$\begin{aligned} h_{t+1}(\theta) &= I_{t+1} - \{\lambda + (\lambda\beta)^{-1}\}I_t + \beta^{-1}I_{t-1} \\ &\quad + S_{t+1} - (1 - \delta\gamma/\alpha)\beta^{-1}S_t \end{aligned}$$

and $\theta = (\lambda, \beta, \delta, \gamma, \alpha)$.

(δ, γ, α) are globally unidentified.

Reason: they only appear in pmc via

$$\phi(\delta, \gamma, \alpha) = 1 - \delta\gamma/\alpha$$

and

$$\phi(\delta, \gamma, \alpha) = \phi(k\delta, \gamma, k\alpha) = \phi(\delta, k\gamma, k\alpha)$$

Notice problem goes away if we set $\phi = 1 - \delta\gamma/\alpha$ and set $\theta = (\lambda, \beta, \phi)$.

In general identification failures arise due to nonlinearity, properties of data and interaction of both.

This can be hard to spot via condition for "global identification" $\rightarrow$ *local identification*.

This concept is based on restricting attention to sufficiently small neighbourhood of θ_0 such that first order Taylor series expansion holds

$$f(v_t, \theta) \approx f(v_t, \theta_0) + \{\partial f(v_t, \theta_0)/\partial \theta'\}(\theta - \theta_0)$$

Note that $f(v_t, \theta)$ can be approximated by a linear function of $\theta - \theta_0$.

Taking expectations and using $E[f(v_t, \theta_0)] = 0$ yields

$$E[f(v_t, \theta)] \approx \{E[\partial f(v_t, \theta_0)/\partial \theta']\}(\theta - \theta_0)$$

This gives the following condition:

Local Identification

The parameter vector θ_0 is locally identified by the population moment condition in Assumption 3.3 if and only if $E[\partial f(v_t, \theta_0)/\partial\theta'\}]$ is of rank p.

Note:

(i) Condition implies identification fails if there are fewer moment conditions than parameters, i.e. $q < p$.

(ii) Identification in nonlinear models may be sensitive to the value of θ_0.

(iii) No local id $\Rightarrow$ No global id.

This condition is still far from transparent.

Example: Eichenbaum's (1989) Model for Inventory Holdings by Firms (Continued)
First consider version with δ, γ, α included.

$$E[\partial f(v_t, \theta_0)/\partial \theta'] = E[z_t \tilde{x}_t'] M(\theta_0)$$

where $\tilde{x}_t = (I_t, I_{t-1}, S_t)'$ and for some (3×5) matrix $M(\theta_0)$.

Therefore,

$$rank(M(\theta)) \leq 3 \;\Rightarrow\; rank\{E[\partial f(v_t, \theta_0)/\partial \theta']\} \leq 3 < p$$

and so θ_0 is locally unidentified in this model.

Note that if we set $\phi = 1 - \delta\gamma/\alpha$ and $\theta = (\lambda, \beta, \phi)'$ then

$$E[\partial f(v_t, \theta_0)/\partial\theta'] = E[z_t \tilde{x}_t'] M(\theta_0)$$

where

$$M(\theta) = \left[\begin{array}{ccc} (\lambda^2\beta)^{-1} - 1 & +(\lambda\beta^2)^{-1} & 0 \\ 0 & -\beta^{-2} & 0 \\ 0 & \phi\beta^{-2} & -\beta^{-1} \end{array} \right]$$

This time $rank\{M(\theta)\} = 3$.

Note that local identification depends on whether $rank\{E[z_t \tilde{x}_t']\} = 3$.

Terminology:

- If $p > q$ then θ_0 is said to be *un* − identified

- If $p = q$ and θ_0 is local id. then said to be *just* − identified

- If $q > p$ and θ_0 is local id. then said to be *over* − identified.

2. GMM estimation

GMM minimand is:

$$Q_T(\theta) = \{T^{-1}\sum_{t=1}^{T} f(v_t,\theta)\}' W_T \{T^{-1}\sum_{t=1}^{T} f(v_t,\theta)\}$$

where W_T is the weighting matrix.

GMM estimator of θ_0 is

$$\hat{\theta}_T \ = \ argmin_{\theta\in\Theta}\, Q_T(\theta)$$

First order conditions for this minimization imply

$$\partial Q_T(\hat{\theta}_T)/\partial\theta = 0$$

and so

$$\{T^{-1}\sum_{t=1}^{T} \frac{\partial f(v_t,\hat{\theta}_T)}{\partial\theta'}\}' W_T \{T^{-1}\sum_{t=1}^{T} f(v_t,\hat{\theta}_T)\} = 0$$

3. Identifying and overidentifying restrictions

GMM based on $E[f(v_t, \theta_0)] = 0$ is MM based on

$$F(\theta_0)' W^{1/2} E[f(v_t, \theta_0)] = 0$$

where $F(\theta_0) = W^{1/2} E[\partial f(v_t, \theta_0)/\partial\theta']$.

If $p = q$ then GMM = MM based on

$$E[f(v_t, \theta_0)] = 0$$

If $q > p$ then GMM is MM based on:

$$F(\theta_0)[F(\theta_0)'F(\theta_0)]^{-1}F(\theta_0)'W^{1/2}E[f(v_t, \theta_0)] = 0$$

$\sim$ *identifying restrictions*

Remainder is:

$$\{I_q - F(\theta_0)[F(\theta_0)'F(\theta_0)]^{-1}F(\theta_0)'\}W^{1/2}E[f(v_t, \theta_0)] =$$

$\sim$ *overidentifying restrictions*

So GMM effects a decomposition on the population moment condition:

$$\text{pmc} = \text{id.res} + \text{overid. res}$$

id.res $\rightarrow$ estimation
overid.res $\rightarrow$ remainder

Roles reflected in sample analogs.

Sample analog to id.res are satisfied at $\hat{\theta}_T$ by construction.

Sample analog to overid. res not satisfied but

$$W_T^{1/2} g_T(\hat{\theta}_T) = \{I_q - P_T(\hat{\theta}_T)\} W_T^{1/2} g_T(\hat{\theta}_T)$$

$\Rightarrow Q_T(\hat{\theta}_T)$ measures of how far the sample is from satisfying the overidentifying restrictions.

4. Asymptotic properties of $\widehat{\theta}_T$

Consistency: $\widehat{\theta}_T \xrightarrow{p} \theta_0$

Proof based on population analog to $Q_T(\theta)$,

$$Q_0(\theta) = \{E[f(v_t, \theta)]\}' \, W \, \{E[f(v_t, \theta)]\}$$

- $\sup_{\theta \in \Theta} |Q_T(\theta) - Q_0(\theta)| \xrightarrow{p} 0$.

- *pmc* $\Rightarrow Q_0(\theta_0) = 0$

- W = p.d + ident. $\Rightarrow Q_0(\theta) > 0$ for all $\theta \neq \theta_0$.

Asymptotic normality: $T^{1/2}(\widehat{\theta}_T - \theta_0) \xrightarrow{d} N(0, MSM')$ where

- $M = (G_0' W G_0)^{-1} G_0' W$.

- $G_0 = E[\partial f(v_t, \theta_0)/\partial \theta']$

- $S = \lim_{T \to \infty} Var[T^{1/2} g_T(\theta_0)]$

MVT: $g_T(\widehat{\theta}_T) = g_T(\theta_0) + G_T(\widehat{\theta}_T, \theta_0, \lambda_T)(\widehat{\theta}_T - \theta_0)$
Premultiply both sides by $G_T(\widehat{\theta}_T)' W_T$, use FOC and rearrange to give:

$$T^{1/2}(\widehat{\theta}_T - \theta_0) = M_T T^{1/2} g_T(\theta_0) + o_p(1)$$

5. Covariance matrix estimation

Need estimate of long run variance S.

$$S = \Gamma_0 + \sum_{i=1}^{\infty} (\Gamma_i + \Gamma_i')$$

where $\Gamma_j = E[(f_t - E[f_t])(f_{t-j} - E[f_{t-j}])']$

Consider three cases.

Case 1: $f_t \sim$ serially uncorrelated sequence

$$\widehat{S}_{SU} = T^{-1} \sum_{t=1}^{T} \widehat{f}_t \widehat{f}_t'$$

where $\widehat{f}_t = f(v_t, \widehat{\theta}_T)$.

Case 2: $f_t \sim$ invertible VARMA:
Estimate the model

$$\hat{f}_t = A_1 \hat{f}_{t-1} + \ldots + A_k \hat{f}_{t-k} + e_t(k)$$

by OLS. Then

$$\hat{S}_{VARMA} = \{I_q - \sum_{i=1}^{k} \hat{A}_i(k)\}^{-1} \hat{\Sigma}(k) \{I_q - \sum_{i=1}^{k} \hat{A}_i(k)\}^{-1'}$$

where $\hat{\Sigma}(k) = T^{-1} \sum_{t=1}^{T} \hat{e}_t(k) \hat{e}_t(k)'$.

Choice of k:

- Estimate for $k = 0, 1, \ldots K$ and choose using information criterion.
- $K = O(T^{1/3})$

Case 3: heteroscedasticity autocorrelation covariance estimators

$$\widehat{S}_{HAC} = \widehat{\Gamma}_0 + \sum_{i=1}^{b(T)} \omega_{iT}(\widehat{\Gamma}_i + \widehat{\Gamma}_i')$$

where

$$\widehat{\Gamma}_i = T^{-1} \sum_{t=i+1}^{T} \widehat{f}_t \widehat{f}_{t-i}'$$

- choice of kernel: e.g. $\omega_{i,T} = 1 - i/[b(T)+1]$

- choice of bandwidth: $b(T) \rightarrow \infty$ and $b(T) = o(T^{1/2})$

Evidence suggests HAC estimators do not perform well if f_t has slowly decaying autocovariances *i.e.* a strong autoregressive component.

$\rightarrow$ rewhitening and recolouring

Estimate the model

$$\widehat{f}_t = A_1 \widehat{f}_{t-1} + \ldots + A_k \widehat{f}_{t-k} + e_t(k)$$

by OLS. Then

$$\widehat{S}_{PWRC} = \{I_q - \sum_{i=1}^{k} \widehat{A}_i(k)\}^{-1} \widehat{\Sigma}_{HAC} \{I_q - \sum_{i=1}^{k} \widehat{A}_i(k)\}^{-1'}$$

where $\widehat{\Sigma}_{HAC}$ is HAC estimator based on $\widehat{e}_t(k)$.

Numerical optimization
Three important aspects of to numerical optimization routines.

- The *starting value* for θ, $\bar{\theta}(1)$.

- The *iterative search method* by which the candidate value of $\widehat{\theta}$ is updated on the i^{th} step.

- The *convergence criterion* used to judge when the minimum has been reached.

Example of iterative routine: Newton–Raphson (NR) algorithm

$$\widehat{\theta}_T^{(j)} = \widehat{\theta}_T^{(j-1)} - \left[\frac{\partial^2 Q_T^{(2)}(\widehat{\theta}_T^{(j-1)})}{\partial\theta\partial\theta'} \right]^{-1} \frac{\partial Q_T^{(2)}(\widehat{\theta}_T^{(j-1)})}{\partial\theta}$$

Example of Convergence criterion:

$$||\widehat{\theta}_T^{(j)} - \widehat{\theta}_T^{(j-1)}|| < \epsilon$$

Numerical illustration: Hansen & Singleton (1982) CBAPM

- Single asset = equally weighted NYSE index (EWR) or value weighted NYSE index (VWR)

- c_t = aggregate per capita consumption

- $z_t = (1, c_t/c_{t-1}, c_{t-1}/c_{t-2}, r_t/p_{t-1}, r_{t-1}/p_{t-2})'$

- Sample: 1960.1–1991.12

Need to choose W_T: use $10^5 I_5$ and $(T^{-1}Z'Z)^{-1}$.

Notice sensitivity to:

1. starting value
2. convergence criterion
3. W_T

Solutions:

1. Try different starting values
2. Gradually tighten convergence criterion and try different criterion if possible.
3. Characterize "optimal" W_T – to do this, we must first develop a criterion for optimality $\rightarrow$ asymptotic theory.

Lecture outline:

1. Introduction

2. GMM estimation in correctly specified non-linear dynamic models - part I.

3. **GMM estimation in correctly specified nonlinear dynamic models - part II.**

4. Hypothesis testing.

5. Finite sample behaviour and moment selection.

Today:

- Two step and iterated GMM estimation
- Impact of transformations
- Continuous updating GMM estimator

1. Two - step and iterated GMM estimation

Recall that: $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, MSM')$
where

- $M = (G_0' W G_0)^{-1} G_0' W$.

- $G_0 = E[\partial f(v_t, \theta_0)/\partial \theta']$

- $S = \lim_{T \to \infty} Var[T^{1/2} g_T(\theta_0)]$

If $p = q$ then $MSM' = (G_0' S^{-1} G_0)^{-1}$
– independent of W.

If $p < q$ then asymptotic variance depends on W.

Optimal choice of W? Choice that minimizes variance.

Optimal weighting matrix: $W = S^{-1}$

$\rightarrow MSM' = (G_0' S^{-1} G_0)^{-1}$

How can we obtain an estimator that has this variance?

Two–step procedure:

1. Estimate with sub–optimal $W_T \to \widehat{\theta}_T(1) \to \widehat{S}_T(1)$.

2. Estimate with $W_T = \widehat{S}_T(1)^{-1} \to \widehat{\theta}_T(1)$.

Can iterate further

1. $\widehat{\theta}_T(i-1) \to \widehat{S}_T(i-1)$

2. Estimate with $W_T = \widehat{S}_T(i-1)^{-1} \to \widehat{\theta}_T(i)$.

Continue until $\|\widehat{\theta}_T(i-1) - \widehat{\theta}_T(i)\| < \epsilon$ or $i = i_{max}$.

Numerical illustration: Hansen & Singleton (1982) CBAPM

$$\text{pmc}: \; E[e_t(\gamma, \delta)\, z_t] = 0$$

where

$$e_t(\gamma, \delta) \;=\; \delta(r_{t+1}/p_t)(c_{t+1}/c_t)^{\gamma-1} - 1$$

and

- Single asset = equally weighted NYSE index (EWR) or value weighted NYSE index (VWR)
- c_t = aggregate per capita consumption
- $z_t = (1, c_t/c_{t-1}, c_{t-1}/c_{t-2}, r_t/p_{t-1}, r_{t-1}/p_{t-2})'$
- Sample: 1960.1–1991.12

As before for first step use $W_T = 10^5 I_5$ and $(T^{-1}Z'Z)^{-1}$.

2. Impact of transformations

Consider five types of transformation:

- Units of measurement for v_t

- Reparameterization

- Normalization of the parameter vector

- Curvature altering transformations of the population moment condition

- Stationarity inducing transformations of $f(.)$

(i) Units of measurement for v_t:
In general, the GMM estimator is not invariant to changes in the units of measurement of v_t.

Example: $E[v_t] - \theta_0 = 0 \rightarrow \hat{\theta}_T = T^{-1}\sum_{t=1}^{T} v_t$.

Data are $x_t = cv_t$

$E[x_t] - \theta_0 = 0 \rightarrow \tilde{\theta}_T = T^{-1}\sum_{t=1}^{T} x_t$

So $\tilde{\theta}_T = c\hat{\theta}_T$

But interpretation of θ_0 has changed!

(ii) Reparameterization:

θ_0 satisfies:

- globally identified
- can be written as $\theta_0 = h(\gamma_0)$ where $h : \Re^p \to \Re^p$ is a continuous, differentiable bijective mapping.

pmc: $E[f(v_t, h(\gamma_0))] = E[f_\gamma(v_t, \gamma_0)] = 0$

What is relationship between:

- estimate γ_0 based on $E[f_\gamma(v_t, \gamma_0)] = 0$
- estimate of θ_0 based on $E[f(v_t, \theta_0)] = 0$?

The GMM estimator is invariant to reparameterization in the sense that the two parameterizations yield logically consistent estimators.

$Q_{\gamma,T}(\gamma) =$ GMM minimand associated with the reparameterized model, *i.e.* $Q_{\gamma,T}(\gamma) = Q_T(h(\gamma))$.

$\widehat{\gamma}_T = argmin\, Q_{\gamma,T}(\gamma)$

Can calculate $\widehat{\gamma}_T$ as follows.

- min $Q_T(h(\gamma))$ wrt $h(\gamma) \to \widehat{h}_T$
- $\widehat{h}_T = h(\widehat{\gamma}_T) \to \widehat{\gamma}_T$.

But $\widehat{h}_T = \widehat{\theta}_T$ and so by construction

$$\widehat{\theta}_T \;=\; h(\widehat{\gamma}_T)$$

Similar result does not extend to the estimated asymptotic standard errors.

It can be shown that

$$\widehat{V}_{\gamma,T} = [H(\widehat{\gamma}_T)]^{-1}\widehat{V}_{\theta,T}[H(\widehat{\gamma}_T)']^{-1}$$

where $H(.) = \partial h(.)/\partial\gamma'$.

So inferences may be sensitive.

This sensitivity is a potential source of concern, and motivates an alternative method for the construction of confidence intervals covered later.

"natural parameterization" argument

(iii) The GMM estimator and normalization of the parameter vector

In some cases, θ_0 may only be identified up to some scaling factor and so it is necessary to impose some normalization on θ_0, such as $\theta_{0,1} = 1$, in order to achieve identification.

In general, the GMM estimators associated with different normalizations of the parameter vector do not exhibit a logical consistency in finite samples. However, they do exhibit a logical consistency in the limit.

Example: Suppose that

$$e_t(\theta) = \theta_1 R_{1,t} + \theta_2 R_{2,t} + \theta_3 I_t + \theta_4 S_t$$

where

$$E[z_t e_t(\theta_0)] = 0$$

θ_0 is unidentified

Normalize to achieve identification:

- Divide $e_t(\theta_0)$ by $\theta_{0,1}$:

  $\tilde{e}_t(\psi_0) = R_{1,t} + \psi_{0,1} R_{2,t} + \psi_{0,2} I_t + \psi_{0,3} S_t$

  where $\psi_{0,i} = \theta_{0,i+1}/\theta_{0,1}$.

- Divide $e_t(\theta_0)$ by $\theta_{0,4}$:

  $\bar{e}_t(\phi_0) = \phi_{0,1} R_{1,t} + \phi_{0,2} R_{2,t} + \phi_{0,3} I_t + S_t$

  where $\phi_{0,i} = \theta_{0,i}/\theta_{0,4}$.

Both normalizations are logically consistent.

These normalizations lead to pmc's:

$$\begin{aligned} E[z_t \tilde{e}_t(\psi_0)] &= 0 \\ E[z_t \bar{e}_t(\phi_0)] &= 0 \end{aligned}$$

Corresponding estimators are:

$$\begin{aligned} \hat{\psi}_T &= \left[(T^{-1} \sum_{t=1}^{T} x_{1,t} z_t^{'}) W_T (T^{-1} \sum_{t=1}^{T} z_t x_{1,t}^{'}) \right]^{-1} \\ &\quad \times (T^{-1} \sum_{t=1}^{T} x_{1,t} z_t^{'}) W_T (T^{-1} \sum_{t=1}^{T} z_t R_{1,t}) \\ \hat{\phi}_T &= \left[(T^{-1} \sum_{t=1}^{T} x_{2,t} z_t^{'}) W_T (T^{-1} \sum_{t=1}^{T} z_t x_{2,t}^{'}) \right]^{-1} \\ &\quad \times (T^{-1} \sum_{t=1}^{T} x_{2,t} z_t^{'}) W_T (T^{-1} \sum_{t=1}^{T} z_t S_t) \end{aligned}$$

$\hat{\psi}_T$ and $\hat{\phi}_T$ are not logically consistent but exhibit this property in the limit.

(iv) Curvature altering transformations of the population moment condition

Model implies $E[f(v_t, \theta_0)] = 0$ but estimation is based on $c(\theta_0)E[f(v_t, \theta_0)] = 0$.

GMM is invariant to curvature altering transformations if $p = q$ but only in the limit if $p < q$.

$p = q$: $\hat{\theta}_T$ solves $c(\hat{\theta}_T)T^{-1}\sum_{t=1}^{T} f(v_t, \hat{\theta}_T) = 0$.

$p < q$: FOC are

$$\left\{ \left[\frac{\partial c(\hat{\theta}_T)}{\partial \theta}\right] T^{-1} \sum_{t=1}^{T} f(v_t, \hat{\theta}_T)' + c(\hat{\theta}_T) G_T(\hat{\theta}_T)' \right\} \times W_T T^{-1} \sum_{t=1}^{T} f(v_t, \hat{\theta}_T) = 0$$

(v) Stationarity inducing transformations of $f(.)$
Model implies $E[f(\tilde{v}_t, \theta_0)] = 0$ but $\tilde{v}_t$ is nonstationary.

Seek $H(\tilde{v}_{t-1}, \theta_0)$ such that:

- $H(\tilde{v}_{t-1}, \theta_0) f((\tilde{v}_t, \theta_0) = h(v_t, \theta_0)$ where v_t is stationary

- $E[h(v_t, \theta_0)] = 0$.

So that GMM estimation can be based on the scaled moment condition $E[h(v_t, \theta_0)] = 0$.

If one stationarity inducing transformation of $f(.)$ then many.

GMM estimator is sensitive to the choice of transformation in finite samples, but is consistent no matter which transformation is used.

Example: Consumption based asset pricing model

Recall that FOC

$$p_t c_t^{\gamma_0 - 1} = \delta_0 E[r_{t+1} c_{t+1}^{\gamma_0 - 1} | \mathcal{F}_t]$$

Since $p_t c_t^{\gamma_0 - 1} \in \mathcal{F}_t$, both sides of this equation were divided by $c_t^{\gamma_0 - 1} p_t$ to give

$$E[\delta_0 (r_{t+1}/p_t)(c_{t+1}/c_t)^{\gamma_0 - 1} - 1 | \mathcal{F}_t] = 0$$

However FOC also implies

$$E[\delta_0 r_{t+1} c_{t+1}^{\gamma_0 - 1} - p_t c_t^{\gamma_0 - 1} | \mathcal{F}_t] = 0$$

Take first approach because $x_{1,t+1} = c_{t+1}/c_t$ and $x_{2,t+1} = r_{t+1}/p_t$ are stationary but (c_t, r_t, p_t) are not.

Transformation not unique.

If $w_t \in -_t$ is a stationary random variable, then division of FOC by $w_t r_t c_t^{\gamma_0 - 1}$ yields

$$E[\delta_0 w_t^{-1}(r_{t+1}/p_t)(c_{t+1}/c_t)^{\gamma_0 - 1} - w_t^{-1} | -_t] = 0 \quad (1)$$

Now consider pmc: $E[u_t(\theta_0) z_t] = 0$. If apply similar argument to (1) then obtain

$$E[u_t(\theta_0)\tilde{z}_t] = 0$$

where $\tilde{z}_t = w_t^{-1} z_t$.

3. Continuous updating GMM estimator
Recall optimal choice of weighting matrix $W = S^{-1}$.

Recall also that $S = \lim_{T\to\infty} Var[T^{1/2} g_T(\theta_0)]$ and so write $S = S(\theta_0)$.

Population analog to the GMM minimand is

$$Q_{pop}(\theta) \ = \ E[f(v_t, \theta)]' S(\theta)^{-1} E[f(v_t, \theta)]$$

In the iterated estimation,

$$Q_{iter,T}(\theta) \ = \ g_T(\theta)' \widehat{S}_T(i-1)^{-1} g_T(\theta)$$

Alternative is to minimize

$$Q_{cont,T}(\theta) \ = \ g_T(\theta)' S_T(\theta)^{-1} g_T(\theta)$$

The *continuous updating GMM estimator* is defined to be,

$$\widehat{\theta}_{cont,T} \ = \ argmin_{\theta \in \Theta} Q_{cont,T}(\theta)$$

Properties of Continuous updating GMM estimator:

- numerically diferent from two-step or iterated estimator
- has same limiting distribution as two-step or iterated estimator
- invariant to curvature altering transformations

Confidence sets that are invariant to reparameterization.

$$TQ_{cont,T}(\theta_0) \xrightarrow{d} \chi^2_q$$

Therefore an asymptotically valid $100(1-\alpha)\%$ confidence set for θ_0 is then given by

$$\{\, \theta \,:\, TQ_{cont,T}(\theta) < c_q(\alpha) \,\}$$

where $c_q(\alpha)$ is the $100(1-\alpha)\%$ percentile of χ^2_q distribution.