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 $GF_T(\theta)$.

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

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

Hansen & Singleton (1982) set

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

and so (6) becomes

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

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 $z_t \in \text{-}\ _t$

$$E[e_t(\gamma, \delta)\, z_t] = E\,[\,E[e_t(\gamma, \delta)\,|\text{-}\ _t]\, z_t] = 0$$

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:*

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

where

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

First order conditions:

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

$\Rightarrow$

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

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

Now,

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

$\Rightarrow Q_T(\widehat{\theta}_T)$ 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

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

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.

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 $\widehat{\theta}_T$ by construction.

Sample analog to overid. res not satisfied but

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

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

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

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

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

- $M = (G_0' W G_0)^{-1} G_0' W$.
- $G_0 = E[\partial f(v_t, \theta_0)/\partial \theta']$
- $S = \lim_{T\to\infty} Var[T^{1/2} g_T(\theta_0)]$

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

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

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*

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

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

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: $b(T) \to \infty$ 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

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

by OLS. Then

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

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

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:

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

1. Two - step and iterated GMM estimation

Recall that: $T^{1/2}(\hat{\theta}_T - \theta_0) \stackrel{d}{\to} N(0, MSM')$
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 $W_T \rightarrow \widehat{\theta}_T(1) \rightarrow \widehat{S}_T(1)$.

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

Can iterate further

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

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

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 $(T^{-1}Z'Z)^{-1}$.