GMM, HAC estimators, & Standard Errors for Business Cycle Statistics

Wouter J. Den Haan
London School of Economics

Overview

- Generic GMM problem
- Estimation
- Heteroskedastic and Autocorrelation Consistent (HAC) estimators to calcuate optimal weighting matrix and standard errors
- Simple applications
  - OLS with correct standard errors
  - IV with multiple instruments
  - standard errors for business cycle statistics

GMM problem

Underlying true model:

$$\mathsf{E}\left[h(x_t;\theta)\right] = 0_p$$

- $\theta : m \times 1$ vector with parameters
- $x : n \times 1$ vector of observables
- $h(\cdot) : p \times 1$ vector-valued function with $p \geq m$
- $0_p : p \times 1$ vector with zeros

Examples

OLS:

$$q_t = a + bp_t + u_t$$

$$\mathrm{E}\begin{bmatrix} u_t \\ u_t p_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- IV:

$$q_t = a + bp_t + u_t$$

$$\mathrm{E}\begin{bmatrix} u_t \\ u_t z_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Examples

DSGE:

$$\begin{aligned} c_t^{-\gamma} &= \mathsf{E}_t\left[\beta\left(1+r_{t+1}\right)c_{t+1}^{-\gamma}\right] \\ u_{t+1} &= \beta\left(1+r_{t+1}\right)c_{t+1}^{-\gamma} - c_t^{-\gamma} \end{aligned}$$

$$\mathsf{E}_t\left[u_{t+1}f\left(z_t\right)\right] = 0_p \Longrightarrow \mathsf{E}\left[u_{t+1}f\left(z_t\right)\right] = 0_p$$

where

- z_t is a vector with variables in information set in period t and
- $f\left(\cdot\right)$ is a measurable function

GMM estimation

- Let
$$g(\theta; Y_T) = \sum_{t=1}^{T} h\left(x_t; \theta\right)/T,$$
where Y_T contains data

- Idea of estimation: choose θ such that $g(\theta; Y_T)$ is as small as possible

- Throughout the slides, remember that $g(\theta; Y_T)$ is a mean

GMM estimation

$$\widehat{\theta}_T = \arg\min_{\theta \in \Theta} g(\theta; Y_T)' W_T g(\theta; Y_T)$$

- where

$$W_T : p \times p \text{ weighting matrix}$$

- No weighting matrix needed if $p = m$

Asymptotic standard errors

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \longrightarrow N(0, V)$$

$$V = \left(DWD'\right)^{-1} DW\Sigma_0 W'D' \left(DWD'\right)^{-1}$$

Asymptotic standard errors

$$
\begin{aligned}
W &= \underset{T\longrightarrow\infty}{\text{plim}}\, W_T \\
D' &= \underset{T\longrightarrow\infty}{\text{plim}}\, \left.\frac{\partial g\left(\theta; Y_T\right)}{\partial \theta'}\right|_{\theta=\theta_0} \\
\theta_0 &: \quad \text{true } \theta \\
\Sigma_0 &= \sum_{j=-\infty}^{\infty} \mathsf{E}\left[h\left(x_t;\theta_0\right) h\left(x_{t-j};\theta_0\right)'\right]
\end{aligned}
$$

Terms showing up in V

- D measures how sensitive g is to changes in θ_0.
  - less precise estimates of θ_0 if g is not very sensitive to changes in θ.
- Σ_0 is the variance-covariance matrix of $\sqrt{T}g(\theta_0; Y_T)$ as $T \longrightarrow \infty$.
  - if the *mean* underlying the estimation is more volatile $\Longrightarrow$ estimates of θ_0 less precise
- obtaining an estimate for Σ_0 is often the tricky bit (more on this below)

Two cases when formula for V simplifies

1. $p = m$ (no overidentifying restrictions)

$$\begin{aligned} \sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) &\longrightarrow N(0, V) \\ V &= \left(D\Sigma_0^{-1}D'\right)^{-1} \end{aligned}$$

2. using optimal weighting matrix, i.e., the matrix W that minimizes V.

$$\begin{aligned} W^{\text{optimal}} &= \Sigma_0^{-1} \\ \sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) &\longrightarrow N(0, V) \\ V &= \left(D\Sigma_0^{-1}D'\right)^{-1} \end{aligned}$$

Example 1

- $Y_T = \{x_t\}_{t=1}^{T}$ and we want to estimate the mean μ

$$
\begin{aligned}
h(x_t;\mu) &= x_t - \mu \\
g(\mu;Y_T) &= \frac{\sum_{t=1}^{T}(x_t - \mu)}{T} \\
\widehat{\mu}_T &= \frac{\sum_{t=1}^{T} x_t}{T}
\end{aligned}
$$

Example 1

- $$D = 1$$

- $\widehat{\Sigma}_T$ equals the variance of $\sqrt{T}\left(\sum_{t=1}^{T}(x_t - \mu)\right)/T$, which equals variance of $\sqrt{T}\left(\sum_{t=1}^{T} x_t\right)/T$

- Variance of $\sum_{t=1}^{T} x_t$ equals variance of x_1 + covariance of $x_1 \& x_2 + \cdots +$ covariance of $x_1 \& x_T +$ covariance of $x_1 \& x_2 +$ variance $x_2 +$ etc.

- **IF** x_t serially uncorrelated, then variance of $\sqrt{T}\left(\sum_{t=1}^{T} x_t\right)/T$ equals variance of x_t

Example 2

- $x_{1,t}$ and $x_{2,t}$ have the same mean
- for simplicity assume that $x_{1,t}$ and $x_{2,t}$ are not correlated with each other and are not serially correlated
- $Y_T = \{x_{1,t}, x_{2,t}\}_{t=1}^{T}$

$$h(x_t;\mu) = \begin{bmatrix} x_{1,t}-\mu \\ x_{2,t}-\mu \end{bmatrix} \; \& \; D = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$\Sigma_0 = \begin{bmatrix} \sigma^2_{x_1} & 0 \\ 0 & \sigma^2_{x_2} \end{bmatrix}$$

$$W = \begin{bmatrix} 1/\sigma^2_{x_1} & 0 \\ 0 & 1/\sigma^2_{x_2} \end{bmatrix}$$

Estimating variance-covariance matrix

- $g(\theta;Y_T)$ is the mean of $h(x_t;\theta)$
- variance of $g(\theta;Y_T)$ is easy to calculate if $h\left(x_t;\theta\right)$ is serially uncorrelated
- but in general it is difficult

Estimate variance of a mean

- The (limit of the) variance-covariance of $\sqrt{T}g(\theta; Y_T)$ equals

$$\Sigma_0 = \sum_{j=-\infty}^{\infty} \mathsf{E}\left[h\left(x_t; \theta_0\right) h\left(x_{t-j}; \theta_0\right)'\right]$$

  This is the spectral density of $h\left(x_t; \theta_0\right)$ at frequency 0

- Since Σ_0 is a variance-covariance matrix, it should be positive semi-definite (PSD), that is, $z'\Sigma_0 z$ should be non-negative for any non-zero column vector z

HAC estimators

1. Kernel-based (truncated, Newey-West, Andrews)
2. Parametric (Den Haan & Levin)

Estimate variance of a mean

$$\widehat{\Sigma}_T = \sum_{j=-J}^{J} \kappa(j;J) \frac{\sum_{t=\max\{1,j+1\}}^{\min\{T,T+j\}} \left[h\left(x_t;\theta_0\right) h\left(x_{t-j};\theta_0\right)' \right]}{T}$$

where $\kappa(\cdot;J)$ is the kernel with bandwidth parameter J.

Truncated kernel

- Truncated

$$\kappa(j;J) = \begin{cases} 1 & \text{if } |j| \leq J \\ 0 & \text{o.w.} \end{cases}$$

- disadvantages
  - potential bias because autocorrelation at $j > J$ is ignored
  - answer not necessarily positive semi-definite

Newey-West (Bartlett) kernel

$$\kappa(j;J) = 1 - \frac{j}{J+1} \text{ for } |j| \leq J$$

- Newey-West kernel always gives PSD
- disadvantages
  - potential bias because autocorrelation at $j > J$ is ignored
  - bias because $k\,(j;J) < 1$ for included terms

Choice of bandwidth parameter

- a high (low) J leads to
  - less (more) bias
  - more (less) sampling variability
- J should be higher if $h(\cdot; \theta_0)$ is more persistent
- $J^{\text{optimal}} \longrightarrow \infty$ as $T \longrightarrow \infty$ but at a slower rate

Choice of bandwidth parameter

- There are papers that give expressions for J^{optimal}, but adding a constant to these expressions does not affect optimality (that is, the analysis only gives an optimal *rate*).
- Thus, in a finite sample you simply have to experiment and hope your answer is robust

Cost of imposing PSD

- Suppose $h(x_t;\theta_0)$ is an MA(1)
  $\Longrightarrow \mathrm{E}\left[h(x_t;\theta_0)\,h\left(x_{t-j};\theta_0\right)\right] = 0$ for $|j| > 1$
  - If $J = 1$ then you have a biased estimate since $k(1;1) = 1/2$
  - If $J > 1$ then you *must* include estimates of expectations that we know are zero
- Suppose $h(x_t;\theta_0)$ has a persistent and not persistent component
  $\Longrightarrow$ you *must* use the same J for both components

VARHAC

Idea:

- Estimate a flexible time-series process (VAR) for h_t, that is,

$$h\left(x_t;\theta_0\right)=\sum_{j=1}^{J}A_j h\left(x_{t-j};\theta_0\right)+\varepsilon_t$$

- Use implied spectrum at frequency zero as the estimate for Σ_0

VARHAC

$$h\left(x_t;\theta_0\right)=\sum_{j=1}^{J}A_j h\left(x_{t-j};\theta_0\right)+\varepsilon_t$$

Then the implied spectrum at frequency zero, i.e., Σ_0, equals

$$\widehat{\Sigma}_T=\left[I_p-\sum_{j=1}^{J}A_j\right]^{-1}\widehat{\Sigma}_{\varepsilon,T}\left[I_p-\sum_{j=1}^{J}A_j'\right]^{-1}$$

with

$$\widehat{\Sigma}_{\varepsilon,T}=\frac{\sum_{j=J+1}^{T}\varepsilon_t\varepsilon_t'}{T}$$

VARHAC

- Estimate is PSD by construction (PSD is obtained without imposing additional bias)
- Only bias is due to lag length potentially being too short
- You can use standard model selection criteria (AIC, BIC) to determine lag length
- You could estimate a VARMA
- VARHAC also gives a consistent estimate for nonlinear processes since Σ_0 only depends on second-order processes which can be captured with VAR

Back to GMM

- Exactly identified case:
  - estimate $\widehat{\theta}_T$ and calculate V_T
- Over-identified case:
  - You need W_T to estimate $\widehat{\theta}_T$
  - W_T^{optimal} depends on Σ_0, which depends on θ_0

Back to GMM

- What to do in practice?
  - obtain (consistent) estimate of θ_0 with simple W_T (identity although scaling is typically a good idea)
  - use this estimate to calculate W_T^{optimal}
  - use W_T^{optimal} to get a more efficient estimate of θ_0
  - calculate variance of $\widehat{\theta}_T$ using $\left(\widehat{D}_T \widehat{\Sigma}_T^{-1} \widehat{D'_T}\right)^{-1}$
  - you could iterate on this

Examples

- OLS with heteroskedastic and serially correlated errors
- IV with multiple instruments
- Business cycle statistics

Key lesson of today's lecture

If you can write the estimation problem as

$$\mathsf{E}\left[h(x_t;\theta)\right] = 0_p$$

then you can use GMM and we know how to calculate standard errors

OLS

$$
\begin{aligned}
q_t &= \theta p_t + u_t \\
h(x_t) &= u_t p_t \\
&= q_t p_t - \theta p_t^2
\end{aligned}
$$

OLS & GMM

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \longrightarrow N(0, V)$$
$$V = \left(D\Sigma_0^{-1}D'\right)^{-1}$$

$$D = \mathsf{E}\left(p_t^2\right)$$

OLS & GMM

$$V = \left(\left(\mathsf{E}\left(p_t^2\right) \right) \Sigma_0^{-1} \left(\mathsf{E}\left(p_t^2\right) \right)' \right)^{-1}$$

- If $\mathsf{E}[u_t u_{t+\tau}] = 0$ for $\tau \neq 0$, then

$$\Sigma_0^{-1} = \mathsf{E}\left[(u_t p_t)^2\right] \text{ and}$$

$$V = \left(\left(\mathsf{E}\left(p_t^2\right) \right) \mathsf{E}\left[(u_t p_t)^2\right] \left(\mathsf{E}\left(p_t^2\right) \right)' \right)^{-1}$$

- If errors are homoskedastic, then

$$\mathsf{E}\left[(u_t p_t)^2\right] = \mathsf{E}\left[u_t^2\right] \mathsf{E}\left[p_t^2\right]$$

$$V = \left(\mathsf{E}\left(p_t^2\right) \right)^{-1} \mathsf{E}\left[u_t^2\right]$$

OLS & GMM

Suppose that

$$\begin{aligned} q_t &= \theta p_t + u_t \\ u_t &= \varepsilon_t p_t \\ \varepsilon_t \text{ i.i.d, } \mathrm{E}_t\left[p_t \varepsilon_t\right] &= 0\text{, and } \varepsilon_t \text{ homoskedastic} \end{aligned}$$

- Then you should do GLS

$$\frac{q_t}{p_t} = \theta + \varepsilon_t$$

- GMM does not by itself do the transformation, i.e., it does not do GLS, but it does give you standard errors that correct for heteroskedasticity

IV

Suppose that

$$\begin{aligned} q_t &= \theta p_t + u_t, \\ \mathsf{E}\left[u_t z_{1,t}\right] &= 0 \text{ and } \mathsf{E}\left[u_t z_{2,t}\right] = 0 \end{aligned}$$

- With GMM you can find the optimal weighting of the two instruments

Business cycles statistics

1.
$$\psi = \frac{\text{standard deviation } (c_t)}{\text{standard deviation } (y_t)}$$

2.
$$\rho = \text{correlation } (c_t, y_t)$$

Statistics are easy to estimate, but what is the standard error?

Business cycles statistics

GMM problem for ψ

$$h\left(x_t;\theta_0\right) = \begin{bmatrix} c_t - \mu_c \\ y_t - \mu_y \\ \left(y_t - \mu_y\right)^2 \psi^2 - \left(c_t - \mu_c\right)^2 \end{bmatrix}$$

Business cycles statistics

corresponding D matrix

$$
\begin{aligned}
D &= \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 2\mathsf{E}\left[(c_t - \mu_c)\right] & 2\mathsf{E}\left[\left(y_t - \mu_y\right)\psi^2\right] & 2\mathsf{E}\left[\left(y_t - \mu_y\right)^2\psi\right] \end{bmatrix} \\
&= \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2\mathsf{E}\left[\left(y_t - \mu_y\right)^2\psi\right] \end{bmatrix}
\end{aligned}
$$

Steps to follow

1. Use standard estimates for μ_c, μ_y, and ψ

2. Obtain estimate for D

3. Obtain estimate for Σ_0 using

$$h\left(x_t;\widehat{\theta}_T\right) = \begin{bmatrix} c_t - \widehat{\mu}_{c,T} \\ y_t - \widehat{\mu}_{y,T} \\ \left(y_t - \widehat{\mu}_{y,T}\right)^2 \widehat{\psi}^2 - \left(c_t - \widehat{\mu}_{c,T}\right)^2 \end{bmatrix}$$

4. Obtain estimate for V using

$$\left(\widehat{D}_T \widehat{\Sigma}_T^{-1} \widehat{D_T'}\right)^{-1}$$

Business cycles statistics

GMM problem for ρ

$$h\left(x_t'; \theta\right) = \begin{bmatrix} c_t - \mu_c \\ y_t - \mu_y \\ \left(y_t - \mu_y\right)^2 \psi^2 - \left(c_t - \mu_c\right)^2 \\ \left(y_t - \mu_y\right)^2 \psi\rho - \left(c_t - \mu_c\right)\left(y_t - \mu_y\right) \end{bmatrix}$$

Business cycles statistics

corresponding D matrix

$$D = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 2\mathsf{E}\left[\left(y_t - \mu_y\right)^2 \psi\right] & 0 \\ 0 & 0 & \mathsf{E}\left[\left(y_t - \mu_y\right)^2 \rho\right] & \mathsf{E}\left[\left(y_t - \mu_y\right)^2 \psi\right] \end{bmatrix}$$

References

- Den Haan, W.J., and A. Levin, 1997, A practioner's guide to robust covariance matrix estimation
    - a detailed survey of all the ways to estimate Σ_0 with more detailed information