Single Equation Linear GMM with Serially Correlated Moment Conditions

Eric Zivot

Winter 2013

Univariate Time Series

Let $\{y_t\}$ be an ergodic-stationary time series with $E[y_t] = \mu$ and $\mathrm{var}(y_t) < \infty$. A fundamental decomposition result is the following:

Wold Representation Theorem. y_t has the representation

$$
\begin{aligned}
y_t &= \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \\
&= \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots \\
\psi_0 &= 1, \ \sum_{j=0}^{\infty} \psi_j^2 < \infty \\
& \varepsilon_t \sim \mathrm{MDS}(0, \sigma^2)
\end{aligned}
$$

Remarks

1. The Wold representation shows that y_t has a linear structure. As a result, the Wold representation is often called the *linear process representation* of y_t

2. $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ is called *square-summability* and controls the memory of the process. It implies that $|\psi_j| \to 0$ as $j \to \infty$ at a sufficiently fast rate.

3. $\varepsilon_t \sim \mathrm{MDS}(0, \sigma^2)$ which is weaker than $\varepsilon_t \sim WN(0, \sigma^2)$.

Variance

$$
\begin{aligned}
\gamma_0 &= \mathsf{var}(y_t) \\
&= \mathsf{var}\left(\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right) \\
&= \sum_{k=0}^{\infty} \psi_k^2 \mathsf{var}(\varepsilon_t) \\
&= \sigma^2 \sum_{k=0}^{\infty} \psi_k^2 < \infty
\end{aligned}
$$

Autocovariances

$$\begin{aligned}
\gamma_j &= E[(y_t - \mu)(y_{t-j} - \mu)] \\
&= E\left[\left(\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right)\left(\sum_{h=0}^{\infty} \psi_h \varepsilon_{t-h-j}\right)\right] \\
& \quad E[(\psi_0 + \psi_1 \varepsilon_{t-1} + \cdots + \psi_j \varepsilon_{t-j} + \cdots) \\
& \quad \times(\psi_0 \varepsilon_{t-j} + \psi_1 \varepsilon_{t-j-1} + \cdots)] \\
&= \sigma^2 \sum_{k=0}^{\infty} \psi_{j+k} \psi_k, \ j = 0, 1, 2, \ldots
\end{aligned}$$

Ergodicity

Ergodicity requires

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty$$

It can be shown that

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty$$

implies $\sum_{j=0}^{\infty} |\gamma_j| < \infty$.

Asymptotic Properties of Linear Processes

LLN for Linear Processes (Phillips and Solo, *Annals of Statistics* 1992). Assume

$$
\begin{aligned}
y_t &= \mu + \psi(L)\varepsilon_t,\ \varepsilon_t \sim \mathsf{MDS}(0,\sigma^2) \\
&= \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},\ \psi(L) = \sum_{j=0}^{\infty} \psi_j L^j \\
& \quad \psi(L) \text{ is 1-summable; i.e,} \\
\sum_{j=0}^{\infty} j|\psi_j| &= 1|\psi_1| + 2|\psi_2| + \cdots < \infty
\end{aligned}
$$

Note: $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ implies $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Then

$$
\begin{aligned}
\hat{\mu} &= \frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{p} E[y_t] = \mu \\
\hat{\gamma}_j &= \frac{1}{T}\sum_{t=1}^{T} (y_t - \hat{\mu})(y_t - \hat{\mu}) \xrightarrow{p} \mathrm{cov}(y_t, y_{t-j}) = \gamma_j,\ \ j \geq 0
\end{aligned}
$$

CLT for Linear Processes (Phillips and Solo, *Annals of Statistics* 1992)

$$\begin{aligned} y_t &= \mu + \psi(L)\varepsilon_t,\ \varepsilon_t \sim \mathsf{MDS}(0, \sigma^2) \\ &= \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},\ \psi(L) = \sum_{j=0}^{\infty} \psi_j L^j \\ &\quad \psi(L) \text{ is 1-summable} \\ \psi(1) &= \sum_{j=0}^{\infty} \psi_j \neq 0 \end{aligned}$$

Then

$$\begin{aligned}
\sqrt{T}(\hat{\mu}-\mu) &\xrightarrow{d} N(0, \mathrm{LRV}) \\
\mathrm{LRV} &= \text{long-run variance} \\
&= \sum_{j=-\infty}^{\infty} \gamma_j \\
&= \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j, \text{ since } \gamma_j = \gamma_{-j} \\
&= \sigma^2 \psi(1)^2 = \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j \right)^2
\end{aligned}$$

Intuition behind LRV formula

Consider

$$\text{var}(\sqrt{T}\bar{y}) = \text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t\right) = \frac{1}{T}\text{var}\left(\sum_{t=1}^{T} y_t\right)$$

Using the fact that

$$\sum_{t=1}^{T} y_t = \mathbf{1}'\mathbf{y},\ \mathbf{1} = (1, \ldots, 1)',\ \ \mathbf{y} = (y_1, \ldots, y_T)'$$

it follows that

$$\text{var}\left(\sum_{t=1}^{T} y_t\right) = \text{var}(\mathbf{1}'\mathbf{y}) = \mathbf{1}'\text{var}(\mathbf{y})\mathbf{1}$$

Now

$$\begin{aligned}
\mathsf{var}(\mathbf{y}) &= E[(\mathbf{y}-\mu\mathbf{1})(\mathbf{y}-\mu\mathbf{1})'] \\
&= \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \cdots & \gamma_0 \end{pmatrix} \\
&= \Gamma
\end{aligned}$$

where $\gamma_j =$cov(y_t, y_{t-j}) and $\gamma_j = \gamma_{-j}$. Therefore,

$$\mathsf{var}\left(\sum_{t=1}^{T} y_t\right) = \mathbf{1}'\mathsf{var}(\mathbf{y})\mathbf{1} = \mathbf{1}'\Gamma\mathbf{1}$$

Now,

$\mathbf{1}'\Gamma\mathbf{1} =$ sum of all elements in the $T \times T$ matrix Γ.

This sum may be computed by summing across the rows, or down the columns, or along the diagonals.

Given the band diagonal structure of Γ, it is most convenient to sum along the diagonals. Doing so gives

$$\begin{aligned}
\text{var}\left(\sum_{t=1}^{T} y_t\right) &= \mathbf{1}'\Gamma\mathbf{1} \\
&= T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\gamma_{T-1}
\end{aligned}$$

Then

$$\begin{aligned}
\text{var}(\sqrt{T}\bar{y}) &= \frac{1}{T}\mathbf{1}'\Gamma\mathbf{1} \\
&= \gamma_0 + 2\frac{(T-1)}{T}\gamma_1 + 2\frac{(T-2)}{T}\gamma_2 + \cdots + 2\frac{1}{T}\gamma_{T-1} \\
&= \gamma_0 + 2 \cdot \sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\gamma_j
\end{aligned}$$

As $T \to \infty$, it can be shown that

$$\text{var}(\sqrt{T}\bar{y}) = \frac{1}{T}\mathbf{1}'\Gamma\mathbf{1} \rightarrow \gamma_0 + 2 \cdot \sum_{j=1}^{\infty}\gamma_j = \text{LRV}$$

Remark

Since $\gamma_j = \gamma_{-j}$, we may re-express $T^{-1}\mathbf{1}'\Gamma\mathbf{1}$ as

$$\frac{1}{T}\mathbf{1}'\Gamma\mathbf{1} = \gamma_0 + \sum_{j=-(T-1)}^{T-1}\left(1 - \frac{|j|}{T}\right)\gamma_j$$

Then, an alternative representation for LRV is

$$\mathrm{LRV} = \sum_{j=-\infty}^{\infty}\gamma_j$$

Example: MA(1) Process

$$\begin{aligned} Y_t &= \mu + \varepsilon_t + \theta\varepsilon_{t-1},\ |\theta| < 1 \\ & \varepsilon_t \sim \text{iid } (0, \sigma^2) \end{aligned}$$

Recall

$$\begin{aligned} \psi(L) &= 1 + \theta L,\ \psi(1) = 1 + \theta \\ \gamma_0 &= \sigma^2(1 + \theta^2),\ \gamma_1 = \sigma^2\theta \end{aligned}$$

Then

$$\begin{aligned} \text{LRV} &= \gamma_0 + 2 \cdot \sum_{j=1}^{\infty} \gamma_j \\ &= \sigma^2(1 + \theta^2) + 2\sigma^2\theta = \sigma^2(1 + \theta)^2 \\ &= \sigma^2\psi(1)^2 \end{aligned}$$

Remarks

1. If $\theta = 0$ then LRV $= \sigma^2(1+\theta)^2 = \sigma^2\psi(1)^2 = \sigma^2$

2. If $\theta = -1$ then $\psi(1) = 0 \Rightarrow$ LRV $= \sigma^2(1+\theta)^2 = \sigma^2\psi(1)^2 = 0$

   (a) This motivates the condition $\psi(1) \neq 0$ in the CLT for stationary and ergodic linear processes

Example: AR(1) Process

$$\begin{aligned} Y_t - \mu &= \phi(Y_{t-1} - \mu) + \varepsilon_t,\ \varepsilon_t \sim \mathsf{WN}(0, \sigma^2),\ |\phi| < 1 \\ E[Y_t] &= \mu \end{aligned}$$

Recall,

$$\begin{aligned} \psi(L) &= (1 - \phi L)^{-1} \\ \gamma_0 &= \frac{\sigma^2}{1 - \phi^2},\ \gamma_j = \phi^j \gamma_0 \end{aligned}$$

Then

$$\mathsf{LRV} = \sigma^2 \psi(1)^2 = \frac{\sigma^2}{(1 - \phi)^2}$$

Straightforward algebra gives

$$\begin{aligned} \mathsf{LRV} &= \gamma_0 + 2 \cdot \sum_{j=1}^{\infty} \gamma_j \\ &= \gamma_0 + 2 \cdot \gamma_0 \sum_{j=1}^{\infty} \phi^j = \frac{\sigma^2}{(1-\phi)^2} \end{aligned}$$

Remarks

1. $\mathsf{LRV} = 0$ if $\phi = 0$

2. $\mathsf{LRV} \to \infty$ as $\phi \to 1$

Estimating Long-Run Variance

$$
\begin{aligned}
y_t &= \mu + \psi(L)\varepsilon_t, \ \varepsilon_t \sim \mathrm{MDS}(0, \sigma^2) \\
\mathrm{LRV} &= \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j \\
&= \sigma^2 \psi(1)^2
\end{aligned}
$$

There are two types of estimators

- Parametric (assume a parametric model for y_t)

- Nonparametric (do not assume a parametric model for y_t)

Example: MA(1) process

$$\begin{aligned}
Y_t &= \mu + \varepsilon_t + \theta\varepsilon_{t-1},\ |\theta| < 1 \\
&= \mu + \psi(L)\varepsilon_t,\ \psi(L) = 1 + \theta L \\
&\quad \varepsilon_t \sim \text{iid } (0, \sigma^2) \\
\text{LRV} &= \sigma^2(1+\theta)^2
\end{aligned}$$

A parametric LRV estimate is

$$\widehat{\text{LRV}} = \hat{\sigma}^2(1+\hat{\theta})^2,\ \hat{\sigma}^2 \xrightarrow{p} \sigma^2,\ \hat{\theta} \xrightarrow{p} \theta$$

Example: AR(p) process

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t,\ \phi_1 + \cdots + \phi_p < 1,\ \varepsilon_t \sim \text{iid } (0, \sigma^2)$$

In lag operator notation we have

$$\begin{aligned} \phi(L) Y_t &= c + \varepsilon_t,\ \phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p,\ \phi(1) < 1 \\ Y_t &= \mu + \psi(L)\varepsilon_t,\ \mu = \phi(1)^{-1} c,\ \psi(L) = \phi(L)^{-1} \end{aligned}$$

Then

$$\text{LRV} = \sigma^2 \psi(1)^2 = \frac{\sigma^2}{\left(1 - \phi_1 - \cdots - \phi_p\right)^2}$$

A parametric LRV estimate is

$$\widehat{\text{LRV}} = \frac{\hat{\sigma}^2}{\left(1 - \hat{\phi}_1 - \cdots - \hat{\phi}_p\right)^2}$$

where $\hat{\phi}_1, \ldots, \hat{\phi}_p$ and $\hat{\sigma}^2$ are the least squares estimates of $\phi_1, \ldots, \phi_p$ and σ^2, respectively.

Remarks

1. Parametric estimates require us to specify a model for y_t (e.g. ARMA(p,q) model)

2. For the parametric estimate, we only need consistent estimates for σ^2 and θ (e.g., GMM estimates or ML estimates) in order for $\widehat{\text{LRV}}$ to be consistent

3. Parametric estimates are generally more efficient than non-parametric estimates provided they are based on the correct model

4. AR(p) models with $p \to \infty$ as $T \to \infty$ can appoximate ARMA(p,q) model

For the general linear process

$$y_t = \mu + \psi(L)\varepsilon_t,\ \varepsilon_t \sim \mathsf{MDS}(0, \sigma^2)$$

a parametric estimator is not possible because $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ has an infinite number of parameters.

A natural nonparametric estimator is the truncated sum of sample autocovariances

$$\begin{aligned} \widehat{\mathsf{LRV}}_{\mathsf{q}} &= \hat{\gamma}_0 + 2\sum_{j=1}^{q} \hat{\gamma}_j \\ q &= \text{truncation lag} \end{aligned}$$

Problems

1. How to pick q?

2. q must grow with T in order for $\widehat{\mathsf{LRV}}_{\mathbf{q}} \xrightarrow{p} \mathsf{LRV}$

3. $\widehat{\mathsf{LRV}}_{\mathbf{q}}$ can be negative in finite samples, particularly if q is close to T. This is due to Perceval's result

$$\sum_{j=1}^{T} \hat{\gamma}_j = 0$$

Kernel Based Estimators

These are nonparametric estimators of the form (Hayashi notation)

$$\widehat{\mathsf{LRV}}_{\mathsf{ker}} = \sum_{j=-(T-1)}^{T-1} \kappa\left(\frac{j}{q(T)}\right)\hat{\gamma}_j$$

where

$$\begin{aligned}
\kappa(\cdot) &= \text{kernel weight function} \\
q(T) &= \text{bandwidth (lag truncation) parameter}
\end{aligned}$$

The kernel estimators are motivated by the result

$$\mathsf{var}\left(\sqrt{T}\bar{y}\right) = \sum_{j=-(T-1)}^{T-1} \left(1 - \frac{|j|}{T}\right) \hat{\gamma}_j$$

which suggests

$$\kappa\left(\frac{j}{q(T)}\right) = \left(1 - \frac{|j|}{T}\right),\ q(T) = T$$

However, kerner estimators with bandwidth $=$ sample size, $q(T) = T$, have non-standard asymptotic behavior (see recent papers by Vogelsang) and typically $q(T) << T$ in most kernels.

Examples of Kernel Weight Functions

Truncated kernel

$$
\begin{aligned}
q(T) &= q < T \\
\kappa(x) &= \begin{cases} 1 & \text{for } |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases}
\end{aligned}
$$

Note

$$
\kappa\left(\left|\frac{j}{q}\right|\right) = \begin{cases} 1 & \text{for } |j| \leq q \\ 0 & \text{for } |j| > q \end{cases}
$$

Example: $q = 2$

$$
\begin{aligned}
\widehat{\mathsf{LRV}}_{\mathsf{ker}}^{Trunc} &= \sum_{j=-(T-1)}^{T-1} \kappa\left(\frac{j}{q(T)}\right)\hat{\gamma}_j = \sum_{j=-2}^{2} \hat{\gamma}_j \\
&= \hat{\gamma}_{-2} + \hat{\gamma}_{-1} + \hat{\gamma}_0 + \hat{\gamma}_1 + \hat{\gamma}_2 \\
&= \hat{\gamma}_0 + 2(\hat{\gamma}_1 + \hat{\gamma}_2)
\end{aligned}
$$

Remarks:

1. The bandwidth parameter $q(T) = q$ acts as a lag truncation parameter

2. Here, $\widehat{\mathsf{LRV}}_{\mathsf{ker}}^{Trunc}$ is not guaranteed to be positive if q is close to T

Bartlett kernel

$$\begin{aligned} q(T) &< T \\ \kappa(x) &= \begin{cases} 1-|x| & \text{for } |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases} \end{aligned}$$

$\widehat{\mathsf{LRV}}_{\mathsf{ker}}$ computed with the Bartlet kernel gives the so-called Newey-West estimator.

Example: $q(T) = 3$

$$\kappa\left(\frac{j}{q(T)}\right) = \kappa\left(\frac{j}{3}\right) = \begin{cases} 1-\left|\frac{j}{3}\right| & \text{for } |j| \leq 2 \\ 0 & \text{for } |j| > 2 \end{cases}$$

Then

$$\begin{aligned}
\widehat{\mathsf{LRV}}_{\mathsf{ker}}^{Bartlett} &= \sum_{j=-(T-1)}^{T-1} \kappa\left(\frac{j}{q(T)}\right)\hat{\gamma}_j = \sum_{j=-2}^{2}\left(1-\frac{|j|}{3}\right)\hat{\gamma}_j \\
&= \frac{1}{3}\hat{\gamma}_{-2} + \frac{2}{3}\hat{\gamma}_{-1} + \hat{\gamma}_0 + \frac{2}{3}\hat{\gamma}_1 + \frac{1}{3}\hat{\gamma}_2 \\
&= \hat{\gamma}_0 + 2\cdot\left[\frac{2}{3}\hat{\gamma}_1 + \frac{1}{3}\hat{\gamma}_2\right] \\
&= \hat{\gamma}_0 + 2\cdot\sum_{j=1}^{2}\left[1-\frac{j}{3}\right]\hat{\gamma}_j
\end{aligned}$$

Remarks

1. Use of the Bartlett kernel guarantees that $\widehat{\mathsf{LRV}}^{Bartlett}_{\mathsf{ker}} > 0$ (See Newey and West, 1987 Ecta for a proof)

2. In order for $\widehat{\mathsf{LRV}}^{Bartlett}_{\mathsf{ker}} \xrightarrow{p} \mathsf{LRV}$, it must be the case that $q(T) \to \infty$ as $T \to \infty$. There is a bias-variance tradeoff. If $q(T) \to \infty$ too slowly then there will be bias; if $q(T) \to \infty$ too fast then there will be an increase in variance.

3. Andrews (1991) Ecta proved that if $q(T) \to \infty$ at rate $T^{1/3}$ then

$\mathsf{MSE}(\widehat{\mathsf{LRV}}^{Bartlett}_{\mathsf{ker}}, \mathsf{LRV})$ will be minimized asymptotically.

Remarks continued

4. Andrews' result $q(T) \to \infty$ at rate $T^{1/3}$ suggests setting

$$q(T) = cT^{1/3}, \ c = \text{ some constant}$$

However, Andrews' result says nothing about what c should be. Hence, Andrews' result is not practically useful.

5. Many software programs use the following rule due to Newey and West (1987)

$$q(T) = \text{int}\left[4\left(\frac{T}{100}\right)^{1/4}\right]$$

6. Andrews (1991) studied two other kernels - the Parzen kernel and the quadratic spectral (QS) kernel (these are commonly programmed into software).

He shows that $q(T) \to \infty$ at rate $T^{1/5}$ in order to asymptotically minimize $\mathsf{MSE}(\widehat{\mathsf{LRV}}_{\mathsf{ker}}, \mathsf{LRV})$. He also showed that $\mathsf{MSE}(\widehat{\mathsf{LRV}}^{QS}_{\mathsf{ker}}, \mathsf{LRV})$ is the smallest among the different kernels.

Remarks continued

7. Newey and West (1994) Ecta gave Monte Carlo evidence that the choice of bandwidth, $q(T)$, is more important than the choice of kernel.

They suggested some data-based methods for automatically choosing $q(T)$. However, as discussed in den Haan and Levin (1997) these methods are not really automatic because they depend on an initial estimate of LRV and some pre-specified weights.

See Hall (2005), section 3.5 for more details.

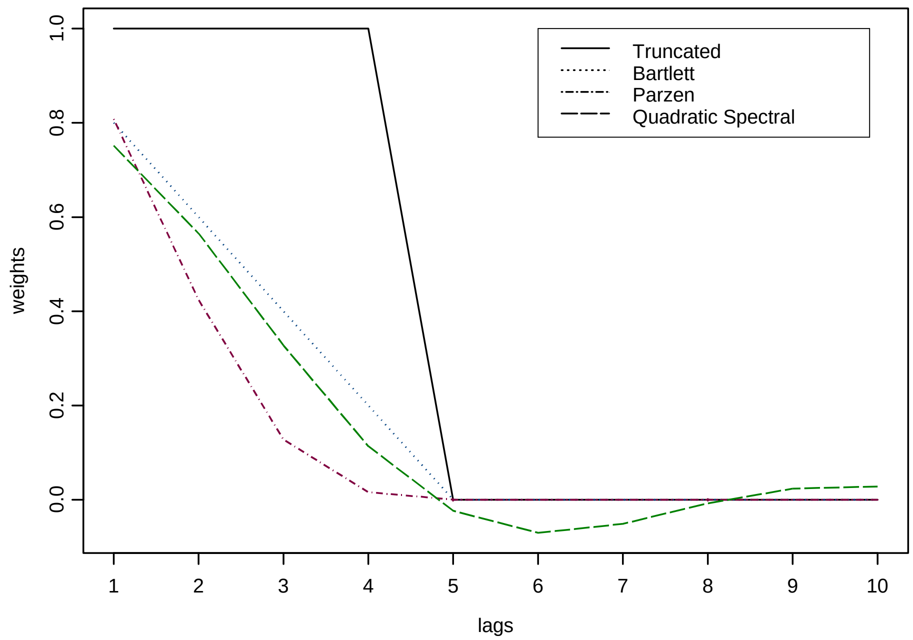

Figure 1: Common kernel functions.

Multivariate Time Series

Consider n time series variables $\{y_{1t}\}, \ldots, \{y_{nt}\}$. A *multivariate time series* is the $(n \times 1)$ vector time series $\{\mathbf{Y}_t\}$ where the i^{th} row of $\{\mathbf{Y}_t\}$ is $\{y_{it}\}$. That is, for any time t, $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$.

Multivariate time series analysis is used when one wants to model and explain the interactions and co-movements among a group of time series variables:

- Consumption and income, Stock prices and dividends, forward and spot exchange rates
- interest rates, money growth, income, inflation
- GMM moment conditions ($\mathbf{g}_t = \mathbf{x}_t \varepsilon_t$)

Stationary and Ergodic Multivariate Time Series

A multivariate time series $\mathbf{Y}_t$ is covariance stationary and ergodic if all of its component time series are stationary and ergodic.

$$
\begin{aligned}
E[\mathbf{Y}_t] &= \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \\
\mathrm{var}(\mathbf{Y}_t) &= \Gamma_0 = E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})'] \\
&= \begin{pmatrix} \mathrm{var}(y_{1t}) & \mathrm{cov}(y_{1t}, y_{2t}) & \cdots & \mathrm{cov}(y_{1t}, y_{nt}) \\ \mathrm{cov}(y_{2t}, y_{1t}) & \mathrm{var}(y_{2t}) & \cdots & \mathrm{cov}(y_{2t}, y_{nt}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(y_{nt}, y_{1t}) & \mathrm{cov}(y_{nt}, y_{2t}) & \cdots & \mathrm{var}(y_{nt}) \end{pmatrix}
\end{aligned}
$$

The correlation matrix of $\mathbf{Y}_t$ is the $(n \times n)$ matrix

$$
\mathrm{corr}(\mathbf{Y}_t) = \mathbf{R}_0 = \mathbf{D}^{-1}\boldsymbol{\Gamma}_0\mathbf{D}^{-1}
$$

where $\mathbf{D}$ is a diagonal matrix with j^{th} diagonal element $(\gamma_{jj}^0)^{1/2} = \mathrm{var}(y_{jt})^{1/2}$.

The parameters μ, Γ_0 and $\mathbf{R}_0$ are estimated from data $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ using the sample moments

$$\begin{aligned}
\bar{\mathbf{Y}} &= \frac{1}{T}\sum_{t=1}^{T} \mathbf{Y}_t \xrightarrow{p} E[\mathbf{Y}_t] = \boldsymbol{\mu} \\
\hat{\boldsymbol{\Gamma}}_0 &= \frac{1}{T}\sum_{t=1}^{T} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})' \xrightarrow{p} \mathrm{var}(\mathbf{Y}_t) = \Gamma_0 \\
\hat{\mathbf{R}}_0 &= \hat{\mathbf{D}}^{-1}\hat{\boldsymbol{\Gamma}}_0\hat{\mathbf{D}}^{-1} \xrightarrow{p} \mathrm{cor}(\mathbf{Y}_t) = \mathbf{R}_0
\end{aligned}$$

where $\hat{\mathbf{D}}$ is the $(n \times n)$ diagonal matrix with the sample standard deviations of y_{jt} along the diagonal.

The Ergodic Theorem justifies convergence of the sample moments to their population counterparts.

All of the lag k cross covariances and correlations are summarized in the $(n \times n)$ lag k cross covariance and lag k cross correlation matrices

$$\mathbf{\Gamma}_k = E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-k} - \boldsymbol{\mu})'] =$$
$$\begin{pmatrix} \mathsf{cov}(y_{1t}, y_{1t-k}) & \mathsf{cov}(y_{1t}, y_{2t-k}) & \cdots & \mathsf{cov}(y_{1t}, y_{nt-k}) \\ \mathsf{cov}(y_{2t}, y_{1t-k}) & \mathsf{cov}(y_{2t}, y_{2t-k}) & \cdots & \mathsf{cov}(y_{2t}, y_{nt-k}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{cov}(y_{nt}, y_{1t-k}) & \mathsf{cov}(y_{nt}, y_{2t-k}) & \cdots & \mathsf{cov}(y_{nt}, y_{nt-k}) \end{pmatrix}$$
$$\mathbf{R}_k = \mathbf{D}^{-1}\mathbf{\Gamma}_k\mathbf{D}^{-1}$$

The matrices $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are not symmetric in k but it is easy to show that $\mathbf{\Gamma}_{-k} = \mathbf{\Gamma}_k'$ and $\mathbf{R}_{-k} = \mathbf{R}_k'$.

The matrices $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are estimated from data $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ using

$$\begin{aligned}
\hat{\mathbf{\Gamma}}_k &= \frac{1}{T} \sum_{t=k+1}^{T} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t-k} - \bar{\mathbf{Y}})' \\
\hat{\mathbf{R}}_k &= \hat{\mathbf{D}}^{-1} \hat{\mathbf{\Gamma}}_k \hat{\mathbf{D}}^{-1}
\end{aligned}$$

Multivariate Wold Representation

Any $(n \times 1)$ covariance stationary multivariate time series $\mathbf{Y}_t$ has a Wold or linear process representation of the form

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k}, \;\; \boldsymbol{\Psi}_0 = \mathbf{I}_n \\ \boldsymbol{\varepsilon}_t &\sim \mathsf{WN}(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}$$

$\boldsymbol{\Psi}_k$ is an $(n \times n)$ matrix with (i,j)th element ψ_{ij}^k.

In lag operator notation, the Wold form is

$$\begin{aligned}\mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t \\ \boldsymbol{\Psi}(L) &= \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k L^k\end{aligned}$$

The moments of $\mathbf{Y}_t$ are given by

$$\begin{aligned}E[\mathbf{Y}_t] &= \boldsymbol{\mu} \\ \text{var}(\mathbf{Y}_t) &= \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\Sigma} \boldsymbol{\Psi}_k'\end{aligned}$$

Digression on Vector Autoregressive Models (VARs)

The most popular multivariate time series model is the *vector autoregressive* (VAR) model. The VAR model is a multivariate extension of the univariate autoregressive model. For example, a bivariate VAR(1) model has the form

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

or

$$\begin{aligned} y_{1t} &= c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} + \varepsilon_{2t} \end{aligned}$$

where

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim iid \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right)$$

The general VAR(p) model for $\mathbf{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ has the form

$$\begin{aligned}
\mathbf{Y}_t &= \mathbf{c} + \mathbf{\Pi}_1 \mathbf{Y}_{t-1} + \mathbf{\Pi}_2 \mathbf{Y}_{t-2} + \cdots + \mathbf{\Pi}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \; t = 1, \ldots, T \\
& \boldsymbol{\varepsilon}_t \sim \text{iid } (\mathbf{0}, \mathbf{\Sigma})
\end{aligned}$$

In lag operator notation, we have

$$\begin{aligned}
\mathbf{\Pi}(L)\mathbf{Y}_t &= \mathbf{c} + \boldsymbol{\varepsilon}_t \\
\mathbf{\Pi}(L) &= \mathbf{I}_n - \mathbf{\Pi}_1 L - \cdots - \mathbf{\Pi}_p L^p
\end{aligned}$$

If

$$\det(\mathbf{\Pi}(z)) = 0$$

has all roots outside the complex unit circle then $\mathbf{Y}_t$ is covariance stationary and ergodic and the Wold representation for Y_t can be found by "inverting" the VAR polynomial

$$\mathbf{Y}_t = \mathbf{\Pi}(L)^{-1}(\mathbf{c} + \boldsymbol{\varepsilon}_t) = \boldsymbol{\mu} + \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t$$

Long Run Variance

Let $\mathbf{Y}_t$ be an $(n \times 1)$ stationary and ergodic multivariate time series with $E[\mathbf{Y}_t] = \boldsymbol{\mu}$. Phillips and Solo's CLT for linear processes states that

$$\sqrt{T}(\bar{\mathbf{Y}} - \boldsymbol{\mu}) \xrightarrow{d} N\left(\mathbf{0}, \mathsf{LRV}\right)$$

$$\underset{(n\times n)}{\mathsf{LRV}} = \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j = \mathbf{\Psi}(1)\mathbf{\Sigma}\mathbf{\Psi}(1)'$$

$$\mathbf{\Psi}(1) = \sum_{k=0}^{\infty} \mathbf{\Psi}_k$$

Since $\mathbf{\Gamma}_{-j} = \mathbf{\Gamma}_j'$, LRV may be alternatively expressed as

$$\mathsf{LRV} = \mathbf{\Gamma}_0 + \sum_{j=1}^{\infty} (\mathbf{\Gamma}_j + \mathbf{\Gamma}_j')$$

Parametric Estimate of Long Run Variance Using VAR(p) Model

Consider estimating the LRV matrix for $\mathbf{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ assuming $\mathbf{Y}_t$ follows a stationary VAR(p) model

$$\begin{aligned}
\mathbf{\Pi}(L)\mathbf{Y}_t &= \mathbf{c} + \boldsymbol{\varepsilon}_t \\
\mathbf{\Pi}(L) &= \mathbf{I}_n - \mathbf{\Pi}_1 L - \cdots - \mathbf{\Pi}_p L^p \\
& \quad \boldsymbol{\varepsilon}_t \sim \text{iid } (\mathbf{0}, \mathbf{\Sigma})
\end{aligned}$$

Using the Wold form

$$\mathbf{Y}_t = \mathbf{\Pi}(L)^{-1}(\mathbf{c} + \boldsymbol{\varepsilon}_t) = \boldsymbol{\mu} + \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t$$

we have that

$$\begin{aligned}
\underset{(n\times n)}{\text{LRV}_{VAR}} &= \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j = \mathbf{\Psi}(1)\mathbf{\Sigma}\mathbf{\Psi}(1)' = \mathbf{\Pi}(1)^{-1}\mathbf{\Sigma}\mathbf{\Pi}(1)^{-1\prime} \\
\mathbf{\Pi}(1) &= \mathbf{I}_n - \mathbf{\Pi}_1 - \cdots - \mathbf{\Pi}_p
\end{aligned}$$

Estimated LRV

The VAR model parmeters can be estimated by least squares equation by equation, and the residual covariance matrix can be estimated using

$$\begin{aligned} \mathbf{\Sigma} &= \frac{1}{T}\sum \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' \\ \hat{\boldsymbol{\varepsilon}}_t &= (\hat{\varepsilon}_{1t}, \ldots, \hat{\varepsilon}_{nt})' \\ \hat{\varepsilon}_{jt} &= \text{OLS residual from } jth \text{ equation} \end{aligned}$$

Then the estimated LRV has the form

$$\begin{aligned} \underset{(n\times n)}{\widehat{\mathsf{LRV}}_{VAR}} &= \hat{\mathbf{\Pi}}(1)^{-1}\hat{\mathbf{\Sigma}}\hat{\mathbf{\Pi}}(1)^{-1\prime} \\ \hat{\mathbf{\Pi}}(1) &= \mathbf{I}_n - \hat{\mathbf{\Pi}}_1 - \cdots - \hat{\mathbf{\Pi}}_p \end{aligned}$$

Non-parametric Estimate of the Long-Run Variance

A consistent estimate of LRV may be computed using non-parametric methods. A popular estimator is the Newey-West weighted autocovariance estimator based on Bartlett weights

$$
\begin{aligned}
\widehat{\mathsf{LRV}}_{NW} &= \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{q(T)-1}\left(1-\frac{j}{q(T)}\right)\cdot\left(\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j'\right) \\
\hat{\mathbf{\Gamma}}_j &= \frac{1}{T}\sum_{t=j+1}^{T}(\mathbf{Y}_t-\bar{\mathbf{Y}})(\mathbf{Y}_{t-j}-\bar{\mathbf{Y}})' \\
q(T) &= \text{bandwidth}
\end{aligned}
$$

Remark: $\widehat{\mathsf{LRV}}_{NW}$ is often denoted $\hat{\mathbf{S}}_{\mathsf{HAC}}$, where HAC denotes "heteroskedasticity and autocorrelation consistent".

Regression Model With Autocorrelated Errors

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\delta}_0 + \varepsilon_t, t = 1, \ldots, T \\ E[\mathbf{x}_t\varepsilon_t] &= 0 \\ & \varepsilon_t \text{ is autocorrelated} \end{aligned}$$

OLS gives

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 = \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t\varepsilon_t$$

Remarks:

1. Previously, we assumed that $\mathbf{g}_t = \mathbf{x}_t\varepsilon_t$ was a MDS so that g_t is an uncorrelated process

2. Now we allow $\mathbf{g}_t$ to be autocorrelated so it is not a MDS

Assume $\{\mathbf{g}_t\} = \{\mathbf{x}_t \varepsilon_t\}$ is a mean zero, 1-summable linear process

$$\mathbf{g}_t = \mathbf{\Psi}(L)\boldsymbol{\eta}_t,\ \boldsymbol{\eta}_t \sim \mathsf{MDS}(\mathbf{0}, \mathbf{\Sigma})$$

satisfying the Phillips-Solo CLT with long-run variance

$$\begin{aligned}
\mathbf{S} &= \mathsf{LRV} = \mathbf{\Gamma}_0 + \sum_{j=1}^{\infty}(\mathbf{\Gamma}_j + \mathbf{\Gamma}_j') = \mathbf{\Psi}(1)\mathbf{\Sigma}\mathbf{\Psi}(1)' \\
\mathbf{\Gamma}_0 &= E[\mathbf{g}_t\mathbf{g}_t'] = E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2],\ \mathbf{\Gamma}_j = E[\mathbf{g}_t\mathbf{g}_{t-j}'] = E[\mathbf{x}_t\mathbf{x}_{t-j}'\varepsilon_t\varepsilon_{t-j}]
\end{aligned}$$

Then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t \xrightarrow{d} N(\mathbf{0}, \mathsf{LRV})$$

Asymptotics for OLS under Autocorrelated Errors

Consistency:

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t \xrightarrow{p} \boldsymbol{\Sigma}_{xx}^{-1}\times 0 = 0$$

Asymptotic Distribution:

$$\sqrt{T}\left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0\right) = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t$$

$$\xrightarrow{d} \boldsymbol{\Sigma}_{xx}^{-1}\times N(\mathbf{0}, \mathsf{LRV}) \equiv N(0, \boldsymbol{\Sigma}_{xx}^{-1}\mathsf{LRV}\boldsymbol{\Sigma}_{xx}^{-1})$$

Here,

$$\mathsf{avar}(\hat{\boldsymbol{\delta}}) = \frac{1}{T}\boldsymbol{\Sigma}_{xx}^{-1}\mathsf{LRV}\boldsymbol{\Sigma}_{xx}^{-1}$$

A consistent estimate for $\mathsf{avar}(\hat{\boldsymbol{\delta}})$ is

$$\begin{aligned}
\widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}) &= \frac{1}{T}\mathbf{S}_{xx}^{-1}\widehat{\mathsf{LRV}}_{NW}\mathbf{S}_{xx}^{-1} \\
\widehat{\mathsf{LRV}}_{NW} &= \hat{\boldsymbol{\Gamma}}_0 + \sum_{j=1}^{q(T)-1}\left(1 - \frac{j}{q(T)}\right)\cdot\left(\hat{\boldsymbol{\Gamma}}_j + \hat{\boldsymbol{\Gamma}}_j'\right) \\
\hat{\boldsymbol{\Gamma}}_j &= \frac{1}{T}\sum_{t=j+1}^{T}\mathbf{g}_t\mathbf{g}_{t-j}' = \frac{1}{T}\sum_{t=j+1}^{T}\mathbf{x}_t\mathbf{x}_{t-j}'\hat{\varepsilon}_t\hat{\varepsilon}_{t-j} \\
\hat{\varepsilon}_t &= y_t - \mathbf{x}_t'\hat{\boldsymbol{\delta}} = \text{OLS residual}
\end{aligned}$$

Remark

$$\mathsf{SE}_{NW}(\hat{\delta}_i) = \left(\frac{1}{T}\mathbf{S}_{xx}^{-1}\widehat{\mathsf{LRV}}_{NW}\mathbf{S}_{xx}^{-1}\right)_{ii}^{1/2} = \text{Newey-West standard error}$$

Single Equation Linear GMM with Serial Correlation

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\delta}_0 + \varepsilon_t,\ t = 1, \ldots, T \\ \mathbf{z}_t &= L \times 1 \text{ vector of explanatory variables} \\ \boldsymbol{\delta}_0 &= L \times 1 \text{ vector of unknown coefficients} \\ \varepsilon_t &= \text{random error term} \\ \mathbf{x}_t &= K \times 1 \text{ vector of exogenous instruments} \end{aligned}$$

Moment Conditions and identification for General Model

$$\begin{aligned} \mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta}_0) &= \mathbf{x}_t\varepsilon_t = \mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\delta}_0) \\ E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta}_0)] &= E[\mathbf{x}_t\varepsilon_t] = E[\mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\delta}_0)] = \mathbf{0} \\ E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta})] &\neq \mathbf{0} \text{ for } \boldsymbol{\delta} \neq \boldsymbol{\delta}_0 \end{aligned}$$

Serially Correlated Moments

It is assumed that $\{\mathbf{g}_t\} = \{\mathbf{x}_t \varepsilon_t\}$ is a mean zero, 1-summable linear process

$$\mathbf{g}_t = \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t \sim \mathsf{MDS}(\mathbf{0}, \mathbf{\Sigma})$$

satisfying the Phillips-Solo CLT with long-run variance

$$\begin{aligned}
\mathbf{S} &= \mathsf{LRV} = \mathbf{\Gamma}_0 + \sum_{j=1}^{\infty}(\mathbf{\Gamma}_j + \mathbf{\Gamma}_j') = \mathbf{\Psi}(1)\mathbf{\Sigma}\mathbf{\Psi}(1)' \\
\mathbf{\Gamma}_0 &= E[\mathbf{g}_t\mathbf{g}_t'] = E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] \\
\mathbf{\Gamma}_j &= E[\mathbf{g}_t\mathbf{g}_{t-j}'] = E[\mathbf{x}_t\mathbf{x}_{t-j}'\varepsilon_t\varepsilon_{t-j}]
\end{aligned}$$

Estimation of LRV

LRV is typically estimated using the Newey-West estimator (kernel estimator with Bartlett kernel)

$$
\begin{aligned}
\widehat{\mathsf{LRV}}_{NW} &= \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{q(T)-1} \left(1 - \frac{j}{q(T)}\right) \cdot \left(\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j'\right) \\
\hat{\mathbf{\Gamma}}_0 &= \frac{1}{T} \sum_{t=j+1}^{T} \mathbf{g}_t \mathbf{g}_t' = \frac{1}{T} \sum_{t=j+1}^{T} \mathbf{x}_t \mathbf{x}_t' \hat{\varepsilon}_t^2 \\
\hat{\mathbf{\Gamma}}_j &= \frac{1}{T} \sum_{t=j+1}^{T} \mathbf{g}_t \mathbf{g}_{t-j}' = \frac{1}{T} \sum_{t=j+1}^{T} \mathbf{x}_t \mathbf{x}_{t-j}' \hat{\varepsilon}_t \hat{\varepsilon}_{t-j} \\
\hat{\varepsilon}_t &= y_t - \mathbf{z}_t' \hat{\boldsymbol{\delta}},\ \hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0
\end{aligned}
$$

where

$$\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}(\mathbf{W}), \hat{\mathbf{W}} = \text{arbitrary 1st step weight matrix}$$

Efficient GMM

$$\begin{aligned}
\hat{\boldsymbol{\delta}}(\widehat{\mathsf{LRV}}_{NW}^{-1}) &= \hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}_{\mathsf{HAC}}^{-1}) = \arg\min_{\boldsymbol{\delta}} \; J(\hat{\mathrm{S}}_{\mathsf{HAC}}^{-1}, \boldsymbol{\delta}) \\
&= \arg\min_{\boldsymbol{\delta}} \; n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathrm{S}}_{\mathsf{HAC}}^{-1}\mathbf{g}_n(\boldsymbol{\delta}) \\
&\Rightarrow \hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}_{\mathsf{HAC}}^{-1}) = (\mathbf{S}_{xz}'\hat{\mathrm{S}}_{\mathsf{HAC}}^{-1}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathrm{S}}_{\mathsf{HAC}}^{-1}\mathbf{S}_{xy}
\end{aligned}$$

As with GMM with heteroskedastic errors, the following efficient GMM estimators can be computed

1. 2-step efficient

2. Iterated efficient

3. Continuously-Updated (CU)

Note: the CU estimator solves

$$\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}^{-1}_{\mathsf{HAC,CU}}) = \arg\min_{\boldsymbol{\delta}} \; n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathrm{S}}^{-1}_{\mathsf{HAC}}(\boldsymbol{\delta})\mathbf{g}_n(\boldsymbol{\delta})$$

$$\hat{\mathbf{S}}_{\mathsf{HAC}}(\boldsymbol{\delta}) = \hat{\boldsymbol{\Gamma}}_0(\boldsymbol{\delta}) + \sum_{j=1}^{q(T)-1}\left(1 - \frac{j}{q(T)}\right)\cdot\left(\hat{\boldsymbol{\Gamma}}_j(\boldsymbol{\delta}) + \hat{\boldsymbol{\Gamma}}'_j(\boldsymbol{\delta})\right)$$

where

$$\begin{aligned}
\hat{\boldsymbol{\Gamma}}_j(\boldsymbol{\delta}) &= \frac{1}{T}\sum_{t=j+1}^{T}\mathbf{x}_t\mathbf{x}'_{t-j}\varepsilon_t(\boldsymbol{\delta})\varepsilon_{t-j}(\boldsymbol{\delta}) \\
\varepsilon_t(\boldsymbol{\delta}) &= y_t - \mathbf{z}'_t\boldsymbol{\delta}
\end{aligned}$$

This can be quite numerically unstable!