Asymptotic distribution of GMM Estimator

Eduardo Rossi

University of Pavia

Econometria finanziaria

2010

Outline

1. **Asymptotic Normality of the GMM Estimator**

2. Long Run Covariance Matrix Estimation

Outline

1. Asymptotic Normality of the GMM Estimator
2. Long Run Covariance Matrix Estimation

To develop the asymptotic distribution of the estimator, we require an asymptotically valid closed form representation for

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$$

This representation comes from an application of the Mean Value Theorem. Let

$$\mathbf{G}_T = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \mathbf{g}(\mathbf{v},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$

The Mean Value Theorem implies that

$$\overline{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_T) = \overline{\mathbf{g}}(\boldsymbol{\theta}_0) + \mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T, \boldsymbol{\theta}_0, \boldsymbol{\lambda}_T)(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$$

where $\mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T, \boldsymbol{\theta}_0, \boldsymbol{\lambda}_T)$, is the matrix whose i-th row is the corresponding row of $\mathbf{G}_T(\overline{\boldsymbol{\theta}}_T^{(i)})$ and $\boldsymbol{\lambda}_T$ is $(L \times 1)$ where

$$\overline{\boldsymbol{\theta}}_T^{(i)} = \lambda_{T,i}\boldsymbol{\theta}_0 + (1 - \lambda_{T,i})\widehat{\boldsymbol{\theta}}_T \quad 0 \leq \lambda_{T,i} \leq 1$$

Premultiplying by $\mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T$

$$\begin{aligned}\mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T\overline{\mathbf{g}}(\widehat{\theta}_T) &= \mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T\overline{\mathbf{g}}(\theta_0) + \\ &\quad \mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T\mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T)(\widehat{\theta}_T - \theta_0)\end{aligned}$$

The F.O.C. imply that

$$\mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T\overline{\mathbf{g}}(\widehat{\theta}_T) = \mathbf{0}$$

it follows that

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) = -\left[\mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T\mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T)\right]^{-1}\mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T\sqrt{T}\,\overline{\mathbf{g}}(\theta_0)$$

The asymptotic behaviour of $\sqrt{T}\overline{\mathbf{g}}(\theta_0)$ is given by a version of the Central Limit Theorem.
To apply the Central Limit Theorem, it is necessary to assume the second moment matrices of the sample moment satisfy certain restrictions.

Properties of the Variance of the Sample Moment

1. $E\left[\mathbf{g}(\mathbf{v}_t, \theta_0)\mathbf{g}(\mathbf{v}_t, \theta_0)'\right]$ exists and is finite;
2. $\lim_{T\to\infty} Var\left[\overline{\mathbf{g}}_T(\theta_0)\right] = \mathbf{S}$

Assumptions

- **Strict stationarity**: The $(R \times 1)$ random vectors $\{\mathbf{v}_t; -\infty < t < \infty\}$ form a strictly stationary process with sample space $\mathbf{V} \subseteq \mathfrak{R}^r$. This assumption implies all expectations of functions of vt are independent of time.
- **Population Moment Condition**: The random vector $\mathbf{v}_t$ and the parameter vector θ_0 satisfy the $(L \times 1)$ population moment condition: $E[\mathbf{g}(\mathbf{v}_t; \theta_0)] = 0$
- **Ergodicity**: The random process $\{\mathbf{v}_t; -\infty < t < \infty\}$ is ergodic.
- **Properties of the Variance of the Sample Moment**
  1. $E\left[\mathbf{g}(\mathbf{v}_t, \theta_0)\mathbf{g}(\mathbf{v}_t, \theta_0)'\right]$ exists and is finite;
  2. $\lim_{T\to\infty} \overline{\mathbf{g}}_T(\theta_0)$

Central Limit Theorem for $\sqrt{T}\,\overline{\mathbf{g}}_T(\theta_0)$

With the Assumptions

$$\sqrt{T}\,\overline{\mathbf{g}}_T(\theta_0) \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{S})$$

Since

$$\widehat{\boldsymbol{\theta}}_T \xrightarrow{\text{p}} \boldsymbol{\theta}_0$$

and $\overline{\boldsymbol{\theta}}_T^{(i)}$ lies on the line segment between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0$ then it follows that

$$\overline{\boldsymbol{\theta}}_T^{i} \xrightarrow{\text{p}} \boldsymbol{\theta}_0 \quad \text{for} \quad i = 1, \ldots, K$$

Intuition suggests that this should imply both $\mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T)$ and $\mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T, \boldsymbol{\theta}_0, \boldsymbol{\lambda}_T)$ converge in probability to $\mathbf{G}_0 = E[\partial \mathbf{g}(\mathbf{v}_t, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}']$.

The following assumptions are needed:

- **Continuity of** $E\left[\frac{\partial \mathbf{g}(\mathbf{v}_t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]$ on some neighbourhood N_ϵ of $\boldsymbol{\theta}_0$.
- **Uniform Convergence of** $\mathbf{G}_T(\boldsymbol{\theta})$:

$$\sup_{\boldsymbol{\theta}\in N_\epsilon}\left\|\mathbf{G}_T(\boldsymbol{\theta})-E\left[\frac{\partial \mathbf{g}(\mathbf{v}_t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]\right\| \xrightarrow{\text{p}} 0.$$

- $\mathbf{W}_T \xrightarrow{\text{p}} \mathbf{W}$

When

$$\begin{aligned}\mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T) &\xrightarrow{\text{p}} \mathbf{G}_0\\ \mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T,\boldsymbol{\theta}_0,\boldsymbol{\lambda}_T) &\xrightarrow{\text{p}} \mathbf{G}_0\end{aligned}$$

$$\overline{\mathbf{M}}_T \equiv -\left[\mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T)'\mathbf{W}_T\mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T,\boldsymbol{\theta}_0,\boldsymbol{\lambda}_T)\right]^{-1}\mathbf{G}_T(\widehat{\boldsymbol{\theta}}_T)'\mathbf{W}_T \xrightarrow{\text{p}} \mathbf{M}$$

where $\mathbf{M}=\left(\mathbf{G}_0'\mathbf{W}\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}$, $(K\times L)$.

Asymptotic normality of the GMM estimator

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{MSM}')$$

This implies that an approximate $100(1-\alpha)\%$ confidence interval for $\theta_{0,i}$ in large samples is given by

$$\widehat{\theta}_{T,i} \pm z_{\alpha/2}\sqrt{\frac{\widehat{V}_{T,ii}}{T}}$$

where $\widehat{V}_{T,ii}$ is a consistent estimator of $\mathbf{MSM}'$.
A natural candidate is based on consistent estimators of the component matrices **M** and **S**:

$$\widehat{\mathbf{M}}_T = -\left[\mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T\mathbf{G}_T(\widehat{\theta}_T)\right]^{-1}\mathbf{G}_T(\widehat{\theta}_T)'\mathbf{W}_T$$

The consistent estimation of **S** is more complicated!!

Asymptotic Normality of $\mathbf{W}_T^{1/2} T^{1/2} \mathbf{g}_t(\widehat{\theta}_T)$ in *correctly specified models*.

$$\mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\widehat{\theta}_T) = \mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\theta_0) + \mathbf{W}_T^{1/2} \mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T) T^{1/2}(\widehat{\theta}_T - \theta_0)$$

with

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) = -\left[\mathbf{G}_T(\widehat{\theta}_T)' \mathbf{W}_T \mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T)\right]^{-1} \mathbf{G}_T(\widehat{\theta}_T)' \mathbf{W}_T \sqrt{T}\, \overline{\mathbf{g}}(\theta_0)$$

$$\begin{aligned}
\mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\widehat{\theta}_T) \quad = \quad & \mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\theta_0) - \\
& \mathbf{W}_T^{1/2} \mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T) \left[\mathbf{G}_T(\widehat{\theta}_T)' \mathbf{W}_T \mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T)\right]^{-1} \\
& \mathbf{G}_T(\widehat{\theta}_T)' \mathbf{W}_T \sqrt{T}\, \overline{\mathbf{g}}(\theta_0)
\end{aligned}$$

$$\mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\widehat{\theta}_T) \quad = \quad \mathbf{N}_T(\widehat{\theta}_T) \mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\theta_0)$$

$$\mathbf{N}_T(\widehat{\theta}_T) = \mathbf{I}_L - \mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T) \left[\mathbf{G}_T(\widehat{\theta}_T)' \mathbf{W}_T \mathbf{G}_T(\widehat{\theta}_T, \theta_0, \boldsymbol{\lambda}_T) \right]^{-1} \mathbf{G}_T(\widehat{\theta}_T)' \mathbf{W}_T^{1/2}$$

A random matrix, $\mathbf{N}_T$ times the random vector $\sqrt{T}\, \overline{\mathbf{g}}(\theta_0)$.

$$\mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\widehat{\theta}_T) \xrightarrow{\mathrm{d}} \mathcal{N}\left(\mathbf{0}, \mathbf{N} \mathbf{W}^{1/2} \mathbf{S} \mathbf{W}^{1/2} \mathbf{N}' \right)$$

where $\mathbf{N} = \mathbf{I}_L - \mathbf{P}(\theta_0)$

$$\mathbf{P}(\theta_0) = \mathbf{F}(\theta_0) \left[\mathbf{F}(\theta_0)' \mathbf{F}(\theta_0) \right]^{-1} \mathbf{F}(\theta_0)'$$

$$\mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\widehat{\theta}_T) = [\mathbf{I}_L - \mathbf{P}(\theta_0)]\, \mathbf{W}^{1/2} \sqrt{T}\, \mathbf{g}_T(\theta_0) + o_p(1)$$

the asymptotic behavior of the estimated sample moment is governed by the function of the data which appears in the overidentifying restrictions.
The

$$\lim_{T\to\infty} E\left[\mathbf{W}_T^{1/2} T^{1/2} \overline{\mathbf{g}}(\widehat{\theta}_T)\right] = \mathbf{0}$$

because the OI restrictions are satisfied at θ_0.
The covariance matrix is singular because

$$\text{rank}\{\mathbf{NSN}'\} = \text{rank}\{\mathbf{I}_L - \mathbf{P}(\theta_0)\} = L - K$$

the rank is the number of OI restrictions.

Definition of **S**:

$$
\begin{aligned}
\mathbf{S} &= \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_t(\theta_0)\right] \\
&= \lim_{T\to\infty} E\left[\left(T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_t(\theta_0) - E\left[T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_t(\theta_0)\right]\right)\times\right. \\
&\quad \left.\left(T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_t(\theta_0) - E\left[T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_t(\theta_0)\right]\right)'\right]
\end{aligned}
$$

since

$$
T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_t(\theta_0) - E\left[T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_t(\theta_0)\right] = T^{-1/2}\sum_{t=1}^{T}\left(\mathbf{g}_t(\boldsymbol{\theta}_0) - E[\mathbf{g}_t(\boldsymbol{\theta}_0)]\right)
$$

It follows that

$$\begin{aligned}\mathbf{S} &= \lim_{T\to\infty} E\Bigg[\Bigg\{T^{-1/2}\sum_{t=1}^{T}\Big(\mathbf{g}_t(\theta_0) - E[\mathbf{g}_t(\theta_0)]\Big)\Bigg\} \times \\ &\quad \Bigg\{T^{-1/2}\sum_{t=1}^{T}\Big(\mathbf{g}_t(\theta_0) - E[\mathbf{g}_t(\theta_0)]\Big)\Bigg\}'\Bigg] \\ &= \lim_{T\to\infty} E\Bigg[T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big(\mathbf{g}_t(\theta_0) - E[\mathbf{g}_t(\theta_0)]\Big)\Big(\mathbf{g}_s(\theta_0) - E[\mathbf{g}_s(\theta_0)]\Big)\Bigg]\end{aligned}$$

the stationarity assumption implies that

$$E\Big(\mathbf{g}_t(\theta_0) - E[\mathbf{g}_t(\theta_0)]\Big)\Big(\mathbf{g}_s(\theta_0) - E[\mathbf{g}_s(\theta_0)]\Big) = \mathbf{\Gamma}_j \quad \forall t$$

so

$$\mathbf{S} = \mathbf{\Gamma}_0 + \lim_{T\to\infty}\Bigg\{\sum_{j=1}^{T-1}\Bigg(\frac{T-j}{T}\Big(\mathbf{\Gamma}_j + \mathbf{\Gamma}_j'\Big)\Bigg)\Bigg\} = \mathbf{\Gamma}_0 + \sum_{i=1}^{\infty}(\mathbf{\Gamma}_i + \mathbf{\Gamma}_i')$$

- literature has focused on the ways to avoid any potential inconsistency caused by inappropriate assumptions about the dynamic specification of $\mathbf{g}_t$.
- all the contributions to this literature develop the properties of the estimator in question under the assumption that the model is correctly specified
$$E[\mathbf{g}_t(\theta_0)] = \mathbf{0}$$
- if this assumption is inappropriate then all the estimators discussed are inconsistent
- the consistency of a covariance matrix estimator depends on the validity of the assumptions about both the mean and dynamic structure of $\mathbf{g}_t(\theta_0)$.
- the use of an inconsistent covariance matrix estimator has a detrimental effect on the properties of certain tests for misspecification, and may in turn affect the properties of moment selection procedures based upon these tests

Estimators of $\mathbf{S}$ under three different sets of assumption about the dynamic structure of $\mathbf{g}_t(\theta_0)$:

1. $\{\mathbf{g}_t(\theta_0)\}$ serially uncorrelated sequence.
2. $\{\mathbf{g}_t(\theta_0)\}$ serially correlated sequence:
    1. $\{\mathbf{g}_t(\theta_0)\}$ generated by a VARMA. The potential disadvantage is that if this model for ft is incorrect then the resulting estimator of $\mathbf{S}$ may be inconsistent.
    2. The second approach uses a member of the class of *heteroscedasticity and autocorrelation covariance* (HAC) matrix estimators

If $\mathbf{g}_t(\theta_0)$ is a serially uncorrelated sequence then

$$\mathbf{\Gamma}_j = \mathbf{0} \quad \text{for } j \neq 0$$

so it follows that

$$\mathbf{S} = \mathbf{S}_{SU} = E[\mathbf{g}_t(\theta_0)\mathbf{g}_t(\theta_0)']$$

$$\widehat{\mathbf{S}}_{SU} = \frac{1}{T}\sum_t \mathbf{g}_t(\widehat{\theta}_T)\mathbf{g}_t(\widehat{\theta}_T)'$$

it can be shown that

$$\widehat{\mathbf{S}}_{SU} \xrightarrow{\text{p}} \mathbf{S}$$

This estimator is positive semi-definite by construction because

$$\widehat{\mathbf{S}}_{SU} = \frac{1}{T}\mathbf{H}'\mathbf{H}$$

where $\mathbf{H}$ is the $(T \times L)$ matrix with t-th rows $\mathbf{g}_t(\widehat{\theta}_T)'$.

This occurs because the underlying theory implies $\{\mathbf{g}_t\}$ is a m.d.s. with respect to the information set $\omega_{t-1} = \{\mathbf{g}_{t-1}, \mathbf{g}_{t-2}, \ldots, \mathbf{g}_1\}$ and satisfies

$$\begin{aligned} E[\mathbf{g}_t(\theta_0)] &= \mathbf{0} \\ E[\mathbf{g}_t(\theta_0)|\Omega_{t-1}] &= \mathbf{0} \\ E[\mathbf{g}_t(\theta_0)\mathbf{g}_s(\theta_0)|\Omega_{t-1}] &= \mathbf{0} \\ E[\mathbf{g}_t(\theta_0)\mathbf{g}_s(\theta_0)] &= \mathbf{0} \end{aligned}$$

If $\mathbf{g}_t(\theta_0)$ is generated by a stationary and invertible VARMA(m,n) model and $E[\mathbf{g}_t(\theta_0)] = \mathbf{0}$, then it has the representation

$$\mathbf{\Phi}(L)\mathbf{g}_t = \mathbf{\Theta}(L)\mathbf{e}_t$$

in which $\{\mathbf{e}_t\}$ is a sequence of independently and identically distributed random vectors

$$\begin{aligned} E[\mathbf{e}_t] &= \mathbf{0} \\ Var[\mathbf{e}_t] &= \mathbf{\Sigma} \end{aligned}$$

The VMA(∞) representation is

$$\mathbf{g}_t = \{\mathbf{\Phi}(L)\}^{-1}\mathbf{\Theta}(L)\mathbf{e}_t.$$

$$\mathbf{S}_{VARMA} = \{\mathbf{\Phi}(1)\}^{-1}\mathbf{\Theta}(1)\mathbf{\Sigma}\mathbf{\Theta}(1)\{\mathbf{\Phi}(1)\}^{-1}$$
$$\widehat{\mathbf{S}}_{VARMA} = \{\widehat{\mathbf{\Phi}}(1)\}^{-1}\,\widehat{\mathbf{\Theta}}(1)\,\widehat{\mathbf{\Sigma}}\,\widehat{\mathbf{\Theta}}(1)\,\{\widehat{\mathbf{\Phi}}(1)\}^{-1}$$

$$\begin{aligned}\widehat{\mathbf{\Phi}}(1) &= \mathbf{I}_L + \sum_{i=1}^{m} \widehat{\mathbf{\Phi}}_i \\ \widehat{\mathbf{\Theta}}(1) &= \mathbf{I}_L + \sum_{i=1}^{m} \widehat{\mathbf{\Theta}}_i\end{aligned}$$

$\{\widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Phi}}_i, \widehat{\mathbf{\Theta}}_i; i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\}$ are consistent estimators of

$$\{\mathbf{\Sigma}, \mathbf{\Phi}_i, \mathbf{\Theta}_i; i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\}$$

Since $\mathbf{g}_t$ is unobserved, these parameter estimates are obtained by estimating a VARMA model for $\widehat{\mathbf{g}}_t$.
The estimator of $\mathbf{\Sigma}$:

$$\widehat{\mathbf{\Sigma}} = \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{e}}_t \widehat{\mathbf{e}}_t'$$

is p.s.d. by construction.

VARMA are difficult to estimate due to the presence of the MA terms.
den Haan and Levin (1996) method.
Rewrite $\mathbf{S}_{VARMA}$ in terms of the coefficients in the VAR(∞) representation and avoid the computational problems associated with the MA terms.
Approximate the infinite VAR by a finite order VAR model, whose order increases with T.
To choose the order of the approximation is made via a data-based model selection criteria.

den Haan and Levin (1996) method.

1. Calculate

$$\widehat{\boldsymbol{\Sigma}}(0) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{g}}_t\widehat{\mathbf{g}}_t'$$

2. Estimate the model

$$\widehat{\mathbf{g}}_t = \mathbf{A}_1(k)\widehat{\mathbf{g}}_{t-1} + \ldots + \mathbf{A}_k(k)\widehat{\mathbf{g}}_{t-k} + \mathbf{e}_t(\widehat{k})$$

for $k = 1, 2, \ldots, K$ and $t = K+1, K+2, \ldots, T$ by OLS where $\widehat{\mathbf{g}}_t = \mathbf{g}(\mathbf{v}_t, \widehat{\theta}_T)$. The estimates are given by

$$\mathbf{A}(k) = \sum_{t=K+1}^{T}\widehat{\mathbf{g}}_t\mathbf{x}_t'\left\{\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right\}^{-1}$$

where

$$\mathbf{A}(k) = (\mathbf{A}_1(k), \mathbf{A}_2(k), \ldots, \mathbf{A}_k(k))$$
$$\mathbf{x}_t' = (\widehat{\mathbf{g}}_{t-1}', \widehat{\mathbf{g}}_{t-2}', \ldots, \widehat{\mathbf{g}}_{t-k}')$$

1. Construct the forecast error

$$\widehat{\mathbf{e}}_t = \widehat{\mathbf{g}}_t - \widehat{\mathbf{A}}(k)\mathbf{x}_t$$

$$\widehat{\boldsymbol{\Sigma}}(k) = \frac{1}{T}\sum_{t=K+1}^{T} \widehat{\mathbf{e}}_t\widehat{\mathbf{e}}_t'$$

2. Let $\widehat{k}$ the value that minimizes *Schwarz's Information Criterion*

$$\widehat{k} = \arg\min_{k\in\{0,1,\ldots,K\}} SIC(k) = \arg\min_{k\in\{0,1,\ldots,K\}} \log\left\{\left|\widehat{\boldsymbol{\Sigma}}(k)\right|\right\} + \frac{\log(T)kL^2}{T}$$

3. Estimate $\mathbf{S}_{VARMA}$ by

$$\widehat{\mathbf{S}}_{VARMA} = \left\{\mathbf{I}_L - \sum_{i=1}^{\widehat{k}} \widehat{\mathbf{A}}_i(k)\right\}^{-1} \widehat{\boldsymbol{\Sigma}}(k) \left\{\mathbf{I}_L - \sum_{i=1}^{\widehat{k}} \widehat{\mathbf{A}}_i(k)\right\}^{-1'}$$

To implement this method, it is necessary to choose K.
Den Haan and Levin (1996) show $\widehat{\mathbf{S}}_{VARMA} \xrightarrow{p} \mathbf{S}$ provided

$$K \to \infty \quad \text{as } T \to \infty$$

and

$$K = O(T^{1/3})$$

but an appropriate rule for picking K in finite samples remains an open question.

This choice has the advantage the lag selection procedure is consistent because

- if $n > 0$

$$\widehat{k} \rightarrow n \quad \text{as } T \rightarrow \infty$$

- if $n = 0$

$$\widehat{k} \rightarrow m$$

den Haan and Levin's method indicates that it is consistent provided the autocovariance structure of $\mathbf{g}_t$ is equivalent to that of some infinite order autoregression.
For this, it is only sufficient and not necessary that $\mathbf{g}_t$ be a VARMA process.

- VARMA processes may not be sufficiently general to capture the dependence structure of $\mathbf{g}_t$ in all cases of interest.
- Development of the class of *heteroscedasticity and autocorrelation covariance* (HAC) matrices which are consistent under relatively weak assumptions on the dependence structure of the process.

The definition of $\mathbf{S}$

$$\mathbf{S} = \mathbf{\Gamma}_0 + \sum_{i=1}^{\infty} (\mathbf{\Gamma}_i + \mathbf{\Gamma}_i')$$

Given this structure, it is natural to estimate $\mathbf{S}$ by truncating this infinite sum and using the sample autocovariances,

$$\widehat{\mathbf{\Gamma}}_j = \frac{1}{T} \sum_{t=j+1}^{t} \widehat{\mathbf{g}}_t \widehat{\mathbf{g}}_{t-j}$$

as estimates of their population analogs.

Truncated version (White and Domowitz (1984)):

$$\widehat{\mathbf{S}}_{TR} = \widehat{\mathbf{\Gamma}}_0 + \sum_{i=1}^{l_T} (\widehat{\mathbf{\Gamma}}_i + \widehat{\mathbf{\Gamma}}_i')$$

consistent when $l_T \to \infty$ as $T \to \infty$ and $l_T = o(T^{1/3})$.
$\widehat{\mathbf{S}}_{TR} \xrightarrow{p} \mathbf{A}$ where $\mathbf{A}$ is p.d.matrix.
It may be indefinite in finite samples.

The source of the trouble is not the truncation but the weights given to the sample autocovariances.
Simple case:

$$\mathbf{S} = \mathbf{\Gamma}_0 + \sum_{i=1}^{l} (\mathbf{\Gamma}_i + \mathbf{\Gamma}_i')$$

the correct order of the process is being used in the estimator but the estimator is still not positive semi-definite:

$$\widehat{\mathbf{S}}_{TR} = \widehat{\mathbf{\Gamma}}_0 + \sum_{i=1}^{l} (\widehat{\mathbf{\Gamma}}_i + \widehat{\mathbf{\Gamma}}_i')$$

it can be rewritten as:

$$\widehat{\mathbf{S}}_{TR} = \frac{1}{T} \mathbf{H}' \mathbf{D} \mathbf{H}$$

where

$$\mathbf{H}' = \begin{bmatrix} \widehat{\mathbf{g}}_1 & \widehat{\mathbf{g}}_2 & \dots & \widehat{\mathbf{g}}_T \end{bmatrix}$$

$$\mathbf{D}_{i,j} = 1 \quad \text{for } j = s_1(i), \ldots, s_2(i) \quad \text{for } i = 1, 2, \ldots, T$$

$$s_1(i) = \max(i - l, 1), \quad s_2(i) = \min(i + l, T)$$

Since D is not p.s.d., neither is $\widehat{\mathbf{S}}_{TR}$.

- The failure of p.s.d.-ness does not always imply negative sample variances.
- Rather it means that negative variances can occur for certain realizations of $\mathbf{H}$.
- In the limit, the problem disappears because all realizations from the process must satisfy

$$\widehat{\mathbf{S}}_{TR} \xrightarrow{p} \mathbf{\Gamma}_0 + \sum_{i=1}^{l} (\mathbf{\Gamma}_i + \mathbf{\Gamma}_i')$$

which is positive definite by definition.

If $l = 0$ then $\widehat{\mathbf{S}}_{TR} = \widehat{\mathbf{S}}_{SU}$ and this estimator is positive semi-definite by construction.
The problem stems from the inclusion of the sample autocovariance matrices: $\{\widehat{\mathbf{\Gamma}}_i, i = 1, \ldots, l\}$

Heteroscedasticity autocorrelation covariance (HAC) matrices

The solution is an estimator in which the contribution of the $\{\widehat{\mathbf{\Gamma}}_i, i = 1, \ldots, l\}$ matrices are weighted to downgrade their role sufficiently in finite samples to ensure p.s.d.-ness but have the weights tend to one as $T \to \infty$ to ensure consistency.

This class consists of estimators of the form:

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

where $\omega_{i,T}$ is known as the kernel (or weight).
The kernel must be chosen to ensure:

- consistency
- positive semi-definiteness.

Newey and West (1987), Bartlett kernel

$$\omega_{i,T} = 1 - \frac{i}{b_T + 1}$$

$b_T \geq 0$, integer, is the *bandwidth* parameter, it controls the number of $\widehat{\boldsymbol{\Gamma}}_i$ included in the HAC estimator:

$$\widehat{\mathbf{S}}_{NW} = \widehat{\boldsymbol{\Gamma}}_0 + \sum_{i=1}^{T-1} \left(1 - \frac{i}{b_T + 1}\right) (\widehat{\boldsymbol{\Gamma}}_i + \widehat{\boldsymbol{\Gamma}}_i')$$

Gallant (1987), Parzen kernel

$$\omega_{i,T} = \begin{cases} 1 - 6\left(\frac{i}{b_T+1}\right)^2 + 6\left(\frac{i}{b_T+1}\right)^3 & \text{for} \quad 0 \leq \frac{i}{b_T+1} \leq \frac{1}{2} \\ 2\left(1 - \frac{i}{b_T+1}\right)^3 & \text{for} \quad \frac{1}{2} \leq \frac{i}{b_T+1} \leq 1 \\ 0 & \text{for} \quad \frac{i}{b_T+1} > 1 \end{cases}$$

Andrews (1991), Quadratic spectral kernel

$$\omega_{i,T} = \frac{25}{12\pi^2 d_i^2}\left[\frac{\sin(m_i)}{m_i} - \cos(m_i)\right]$$

where

$$\begin{aligned} d_i &= \frac{i}{b_T} \\ m_i &= \frac{6\pi d_i}{5} \end{aligned}$$

- Newey and West (1994): the choice between the kernels is not particularly important.
- The bandwidth is a much more important determinant of the finite sample properties of $\widehat{\mathbf{S}}_{HAC}$.
- Newey and West (1994) propose a nonparametric method for selecting the bandwidth and show it minimizes the asymptotic mean square error criterion.

- For consistency, $b_T \to \infty$ with T.
- The optimal bandwidth for the Bartlett weights $b_T = cT^{1/3}$ for any choice of finite $c > 0$. `GRETL`: `set hac_lag nw1` $b_T = 0.75T^{1/3}$.
- Newey and West (1994) propose a nonparametric method for selecting the bandwidth and show it minimizes the asymptotic mean square error criterion.

1. Use the $(L \times 1)$ vector h to construct the scalar random variable $c_T = \mathbf{h}'\widehat{\mathbf{g}}_t$
2. Construct $\widehat{\sigma}_j = T^{-1} \sum_{t=j+1}^{T} c_t c_{t-j}, j = 0, 1, \ldots, n$
3. Calculate $\widehat{s}^{(\nu)} = 2 \sum_{j=1}^{n} j^{\nu} \widehat{\sigma}_j$ and $\widehat{s}^{(0)} = \widehat{\sigma}_0 + 2 \sum_{j=1}^{n} \widehat{\sigma}_j$
4. Calculate $\widehat{\gamma} = c_{\gamma}[(\widehat{s}^{(\nu)}/\widehat{s}^{(0)})^2]^{1/(2\nu+1)}$
5. b_T:
   - Bartlett and Parzen kernels: $b_T = \text{int}(\widehat{\gamma} T^{1/(2\nu+1)})$, int denotes the integer part of the number inside the brackets
   - Quadratic Spectral Kernel: $b_T = \widehat{\gamma} T^{1/(2\nu+1)}$

- The exact choice of n is not specified. Newey and West recommend that the calculations be repeated for different choices of n to ensure the resulting confidence intervals or hypothesis tests are not sensitive to the choice of n.
- The vector h must also be chosen. Newey and West focus on the case

$$\mathbf{g}_t = \mathbf{z}_t u_t$$

if $z_{t1} = 1$ then $\mathbf{h} = (0, 1, 1, \ldots, 1)$

The estimation error (for HAC estimators for which $\omega_{i,t} = 0$ for $i > b_T$)is

$$\begin{aligned}
\mathbf{S} - \widehat{\mathbf{S}}_{HAC} &= \mathbf{\Gamma}_0 + \sum_{i=1}^{\infty}(\mathbf{\Gamma}_i + \mathbf{\Gamma}_i') - \widehat{\mathbf{\Gamma}}_0 - \sum_{i=1}^{b_T} \omega_{i,T}(\widehat{\mathbf{\Gamma}}_i + \widehat{\mathbf{\Gamma}}_i') \\
&= \mathbf{\Gamma}_0 - \widehat{\mathbf{\Gamma}}_0 + \sum_{i=1}^{b_T} \omega_{i,T} \left\{ (\mathbf{\Gamma}_i - \widehat{\mathbf{\Gamma}}_i) + (\mathbf{\Gamma}_i' - \widehat{\mathbf{\Gamma}}_i') \right\} \\
&\quad + \sum_{i=1}^{b_T}(1 - \omega_{i,T})(\mathbf{\Gamma}_i + \mathbf{\Gamma}_i') + \sum_{i=b_T+1}^{\infty} (\mathbf{\Gamma}_i + \mathbf{\Gamma}_i')
\end{aligned}$$

$\widehat{\mathbf{S}}_{HAC}$ is unlikely to perform well in finite samples if the population autocovariance matrices of $\mathbf{g}_t$ die out *too slowly*. Such behaviour would be observed if $\mathbf{g}_t$ is generated by a process with a substantial autoregressive component.

Andrews and Monahan (1992) to propose a modification to the HAC estimator based on a technique called *prewhitening and recolouring*.

- *Prewhitening*: Filter $\mathbf{g}_t$ to reduce the size of its autoregressive component and hence to produce a series for which an HAC estimator works better.
  Andrews and Monahan (1992) recommend using a VAR(m) process to filter the data.
- *Recolouring*: The long run variance of $\mathbf{g}_t$ is estimated from the HAC and the properties of the filter.

Simulation evidence in Andrews and Monahan (1992) and Newey and West (1994) suggests the use of prewhitening and recolouring improves the finite sample performance of the asymptotic confidence intervals.

VAR(m):

$$\widehat{\mathbf{g}}_t = \mathbf{A}_1(m)\widehat{\mathbf{g}}_{t-1} + \mathbf{A}_2(m)\widehat{\mathbf{g}}_{t-2} + \ldots + \mathbf{A}_2(m)\widehat{\mathbf{g}}_{t-m} + \mathbf{e}_t(m)$$

must also satisfy certain properties:

$$\mathbf{A}(L) = \mathbf{I}_L - \sum_{i=1}^{m} \mathbf{A}_i(m)$$

must satisfy the conditions for stationarity. Andrews and Monahan propose modifying the filter to ensure it satisfies the required "stationarity" condition.
Newey and West (1994) recommend using $\widehat{\mathbf{S}}_{PWRC}$ with $m = 1$ and their method of bandwidth selection.

- Fit the VAR(1)

$$\widehat{\mathbf{g}}_t = \mathbf{A}\widehat{\mathbf{g}}_{t-1} + \mathbf{e}_t$$

OLS estimate: $\widehat{\mathbf{A}}$

- Decompose by the Singular Value Decomposition:

$$\widehat{\mathbf{A}} = \widehat{\mathbf{B}}\widehat{\boldsymbol{\Delta}}\widehat{\mathbf{C}}'$$

where $\widehat{\boldsymbol{\Delta}} = \text{diag}(\delta_{11}, \ldots, \delta_{LL})$ δ_{ii} denotes the i-th eigenvalue of $\widehat{\mathbf{A}}\widehat{\mathbf{A}}'$ and $\widehat{\mathbf{A}}'\widehat{\mathbf{A}}$. $\widehat{\mathbf{B}}$ contains the eigenvectors of $\widehat{\mathbf{A}}\widehat{\mathbf{A}}'$ and $\widehat{\mathbf{C}}$ those of $\widehat{\mathbf{A}}'\widehat{\mathbf{A}}$.
The condition for stationarity reduces to the requirement that the eigenvalues of

$$p \lim_{T \to \infty} \widehat{\mathbf{A}} = \mathbf{A}$$

be less than one in modulus.
The eigenvalues of $\widehat{\mathbf{A}}$ are guaranteed to satisfy the required constraint if all the elements of $\widehat{\boldsymbol{\Delta}}$ are less than or equal to 0.97.

- Andrews and Monahan (1992) propose modifying $\widehat{\mathbf{A}}$ to ensure its eigenvalues are less than 0.97 in absolute value:

$$\widetilde{\mathbf{A}} = \widehat{\mathbf{B}}\widetilde{\mathbf{\Delta}}\widehat{\mathbf{C}}'$$

where

$$\widetilde{\mathbf{\Delta}}_{ii} = \min\left(\widehat{\mathbf{\Delta}}_{ii}, 0.97\right)$$

- Construct

$$\widetilde{\mathbf{e}}_t = \widehat{\mathbf{g}}_t - \widetilde{\mathbf{A}}\widehat{\mathbf{g}}_{t-1}$$

- Use an HAC estimator in conjunction with Newey and West's method of bandwidth selection given above to construct the matrix

$$\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Gamma}}_0 + \sum_{i=1}^{T-1} \omega_{i,T}(\widehat{\mathbf{\Gamma}}_i + \widehat{\mathbf{\Gamma}}_i')$$

$$\widehat{\mathbf{\Gamma}}_i = \frac{1}{T} \sum_{t=i+1}^{T} \widetilde{\mathbf{e}}_t \widetilde{\mathbf{e}}_{t-i}'$$

- the estimator of **S**:

$$\mathbf{S}_{SE} = \left\{\mathbf{I}_L - \widehat{\mathbf{A}}\right\}^{-1} \widehat{\mathbf{\Sigma}} \left\{\mathbf{I}_L - \widehat{\mathbf{A}}\right\}^{-1}$$

1. Andrews D. W. K. and Monahan, J. C. (1992). "An improved heteroscedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 60: 953-66.
2. Newey W. K. and West, K. D. (1987). "A simple positive semidefinite heteroscedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55: 703-8.
3. Newey W. K. and West, K. D. (1994). "Automatic lag selection in covariance matrix estimation", *Review of Economic Studies*, 61: 631-53