Econometría II. 3^o LADE
Mínimos Cuadrados Generalizados

1 Introducción.

Hasta el momento hemos supuesto que las propiedades del vector de estimadores coinciden con las especificadas para el modeloe lineal general. Es decir, suponemos que este vector sigue una distribución centrada en 0 y con matriz de varianzas y covarianzas $Var(u) = \sigma^2 I$. Por tanto, esta matriz es diagonal, lo que supone que todas las varianzas coinciden ($Var(u_i) = \sigma^2$, $\forall\, i = 1, 2, ..., N$) y que no existe correlación temporal entre los elementos del vector de perturbaciones ($E(u_i\, u_j) = 0$, $\forall\, i \neq j$).

Sin embargo, en la práctica no es habitual que se cumplan estas hipótesis. Por ejemplo, en presencia de datos de corte trasversal es frecuente la aparición de problemas de heteroscedasticidad, por lo que es posible observar que $Var(u_i) = \sigma_i^2 \neq \sigma^2$, $\forall\, i = 1, 2, ..., N$. Por otro lado, en datos de series temporales la presencia de correlación temporal de las perturbaciones es algo frecuente. En estos casos tenemos que $E(u_i\, u_j) \neq 0$ para al menos un par de valores i, j.

A lo largo de los siguientes temas vamos a estudiar lo que sucede cuando se incumplen estas hipótesis. En el presente capítulo estudiaremos los efectos en la estimación mínimo cuadrático ordinaria y los métodos que podemos utilizar para resolver este problema. En los siguientes, abordaremos por separado los problemas de heteroscedasticidad y autocorrelación particularizando para cada caso los métodos de detección.

2 Consecuencias sobre la estimación mco

Supongamos que queremos estimar el siguiente modelo

$$y = X\beta + u$$

donde asumimos que $E[u] = 0$, por lo que no existen problemas de especificación. Además, mantenemos el supuesto de que la X es una matriz de elementos no estocásticos. Por contra, en lugar de realizar el supuesto habitual sobre la matriz de varianzas y covarianzas del vector de perturbaciones, suponemos que:

$$Var\,(u) = \Omega \neq \sigma^2 I.$$

donde Ω es una matriz no singular. Adicionalmente, también suponemos que $\lim_{T\to\infty} \dfrac{X'\Omega X}{T} = \Omega_X$ es una matriz finita y definida positiva. En estas circunstancias, queremos estudiar el comportamiento de los estimadores mínimo cuadráticos ordinarios.

a) Insesgadez

Para estudiar si el vector de estimadores de parámetros de posición contiúa siendo insegado, debemos expresarlo de la siguiente manera:

$$\hat{\beta} = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'u$$

Es directo comprobar ahora que:

$$\begin{aligned} E\left[\hat{\beta}\right] &= E\left[\beta + (X'X)^{-1} X'u\right] = E[\beta] + E\left[(X'X)^{-1} X'u\right] \\ &= \beta + (X'X)^{-1} X'E[u] = \beta \end{aligned}$$

POr tanto, si suponemos que la matriz X no es estocástica, entonces el vector de estimadores mínimo cuadrático ordinario es insesgado.

b) Consistencia

Para estudiar esta propiedad, basta con aplicar límites en probabilidad y comprobar que $\text{plim}\left(\hat{\beta}\right) = \beta$. Para ello, debemos tener en cuenta que:

$$\begin{aligned} \text{p}\lim\left(\hat{\beta}\right) &= \text{p}\lim\left[\beta + (X'X)^{-1} X'u\right] = \text{p}\lim(\beta) + \text{p}\lim\left[\beta + (X'X)^{-1} X'u\right] \\ &= \beta + \text{p}\lim\left[(X'X)^{-1}\right] \;\; \text{p}\lim[X'u] \end{aligned}$$

A partir de los supuestos básicos, deducimos que $\text{plim}\left[(X'X)^{-1}\right] = \sum_{XX}^{-1}$. Debemos ahora conocer el valor de $\text{plim}\,[X'u]$. Para ello, hay que observar que $\dfrac{X'u}{T}$ tiene media 0 y matriz de varianzas y covarianzas

$$\frac{X'\Omega X}{T^2} \to 0$$

Si, como hemos asumido, plim $\frac{X'\Omega X}{T}$ tiende hacia un valor finito, entonces resulta inmediato probar que:

$$p\lim \frac{X'u}{T} = 0$$

En consecuencia:

$$p\lim \left(\hat{\beta}\right) = \beta.$$

Por lo que el vector de estimadores de parámetros de posición es consistente.

c) La matriz de varianzas y covarianzas del vector de estimadores de parámetros de posición es igual a:

$$\begin{aligned} Var(\hat{\beta}) &= E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right] = E\left[\left(X'X\right)^{-1} X'uu'X \left(X'X\right)^{-1}\right] \\ &= \left(X'X\right)^{-1} X' \; E\left[uu'\right] \; X \left(X'X\right)^{-1} = \left(X'X\right)^{-1} X' \; \Omega X \left(X'X\right)^{-1} \end{aligned}$$

Por tanto, la expresión de la matriz y covarianzas es distinta a la obtenida bajo el supuesto de matriz de varianzas y covarianzas escalar, $\sigma^2 \left(X'X\right)^{-1}$. Por tanto, esta matriz no se puede seguir utilizando en, por ejemplo, el cálculo de los t-ratios. Además, la nueva matriz de varianzas y covarianzas puede ser inferior o superior a $\sigma^2 \left(X'X\right)^{-1}$. Por último, a partir de este resultado no queda claro que el estimador mínimo cuadrático ordinaria continúe siendo eficiente.

No obstante, resulta directo probar que, ahora, el vector de estimadores de parámetros de posisición es se distribuye asintóticamente según la siguiente formula:

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \rightarrow N\left(0,\; Q_{XX}^{-1}\Omega_X Q_{XX}^{-1}\right)$$

donde

$$\begin{aligned} \Sigma_{XX}^{-1} &= \text{p}\lim \left(\frac{X'X}{T}\right)^{-1} \\ \Omega_X &= \text{p}\lim \frac{X'\Omega X}{T}. \end{aligned}$$

d) La estimación del parámetro de dispersión queda afectada por la introducción del nuevo supuesto sobre la distribución del vector de perturbaciones. Cuando su matriz de varianzas y covarianzas era escalar, entonces $\hat{\sigma}^2 = RSS/(T-k) = \hat{u}'\hat{u}/(T-k)$ es (en general) un estimador sesgado e inconsistente del parámetro σ^2. Para comprobarlo, debemos tener en cuenta que:

$$\begin{aligned}
E[\hat{u}'\hat{u}] &= E[u'M\,u] \\
&= tr\ [E[M\,uu']] \\
&= \sigma^2\, tr\,[M\ \Omega] \\
&\neq \sigma^2\,(T-k) \text{ en general.}
\end{aligned}$$

Por lo que el estimador es sesgado. Su inconsistencia se demuestra de la siguiente manera:

$$p\lim s^2 = \sigma^2 p\lim \frac{1}{T} tr\,[M\Omega] \neq \sigma^2.$$

e) No se puede realizar inferencia en este tipo de modelo. Al margen de los problemas encontrados en la estimación mco de los diferentes parámetros de interés, debemos tener que en cuenta que los estadísticos que nos permiten realizar inferencia están basados en la siguiente relación:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{T-k} \tag{1}$$

No obstante, esta relación ya no es cierta ya que:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{u'Mu}{\sigma^2} = \left(\frac{u}{\sigma}\right)' M \left(\frac{u}{\sigma}\right)$$

Cuando el ratio $\frac{u}{\sigma}$ sigue una distribución $N(0,I)$, se obtiene el resultado expresado en (1). Sin embargo, si la matriz de varianzas y covarianzas no es escalar, entonces $\frac{u}{\sigma} \nsim N(0,I)$, por lo que cualquier tipo de inferencia basado en los contraste habituales es incorrecta.

3 Estimación mínimo cuadrática generalizada.

Hemos visto en la sección anterior los problemas que causa la violación de los supuestos sobre la matriz de varianzas y covarianzas del vector de perturbaciones sobre la estimación mínimo cuadrática ordinaria. Como los estimadores pierden parte de sus buenas propiedades, esto conlleva que sean necesarios métodos de estimación alternativos. Uno de ellos es la estimación mínimo cuadrática generalizada. Para entender el funcionamiento de este método de estimación, supongamos que queremos estimar el siguiente modelo:

$$y = X\beta + u$$

donde $E[u] = 0$ y

$$V(u) = \Omega.$$

donde asumimos que el resto de las hipótesis básicas se cumple y que Ω is simetrica, no singular y positiva definida. Dadas estas características, entonces existe una matriz no singular D, tal que:

$$D'D = \Omega^{-1}.$$

Si pre-multiplicamos el modelo original por la matriz D, tenemos que:

$$Dy = DX\beta + Du$$

o, de forma similar,

$$y^* = X^*\beta + u^* \tag{2}$$

donde

$$\begin{aligned} y^* &= Dy \\ X^* &= DX \\ u^* &= Du \end{aligned}$$

Si estudiamos las propiedades del vector de perturbaciones del modelo transformado, podemos comprobar que este vector de perturbaciones cumple con las hipótesis básicas. Para comprobar este extremos, basta con considerar que la esperanza del vector u^* es igual

$$E[u^*] = E[Du] = 0,$$

Por otro lado, la matriz de varianzas y covarianzas es igual a:

$$\begin{aligned} V(u^*) = V(Du) &= DV(u)D' \\ &= D\Omega D' \\ &= I. \end{aligned}$$

En el último paso de la igualdad anterior hay que considerar que si premultiplicamos $D\Omega D'$ por la matriz D', tenemos que:

$$D'D\Omega D' = \Omega^{-1}\Omega D' = D'$$

Por tanto, dado que D' es no singular, debe cumplirse que:$D\Omega D' = I$.

Si aplicamos ahora mínimos cuadrados ordinarios al modelo (2), el estimador resultante ha de tener necesariamente las mismas propiedades que las analizadas para el estimador OLS en presencia de una matriz de varianzas y covarianzas escalar. Este nuevo vector de estimadores se le conoce cómo vector de estimadores mínimo cuadrático generalizado y se define de la siguiente manera:

$$\begin{aligned} \hat{\beta}_{GLS} &= \left(X^{*\prime}X^{*}\right)^{-1} X^{*\prime}y^{*} \\ &= \left(X'D'DX\right)^{-1} X'D'DY \\ &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \end{aligned} \tag{3}$$

Es sencillo demostrar la insesgadez de este vector de estimadores. Debemos comenzar notando este estimador de la siguiente manera:

$$\begin{aligned} \hat{\beta}_{GLS} &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \\ &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\left(X\beta + u\right) \\ &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}X\beta + \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u \\ &= \beta + \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u \end{aligned}$$

Si tomamos ahora esperanzas matemáticas a ambos lados de la igualdad anterior, resulta que:

$$\begin{aligned} E\left(\hat{\beta}_{GLS}\right) &= E\left(\beta + \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u\right) \\ &= E\left(\beta\right) + E\left[\left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u\right] \\ &= \beta + \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}E(u) \\ &= \beta \end{aligned}$$

donde resulta fundamental el supuesto sobre la no estocasticidad de la matriz X. La matriz de varianzas y covarianzas del vector de estimadores generalizado se puede obtener de forma sencilla:

$$
\begin{aligned}
Var\left(\hat{\beta}_{GLS}\right) &= Var\left[\beta + \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u\right] \\
&= Var\left[\left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u\right] \\
&= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Var(u)\,\Omega^{-1}X\ \left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1} X'\underbrace{\Omega^{-1}\,\Omega}_{I_T}\Omega^{-1}X\ \left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1} \underbrace{X'\Omega^{-1}\,X\ \left(X'\Omega^{-1}X\right)^{-1}}_{I_T} \\
&= \left(X'\Omega^{-1}X\right)^{-1} = \left(X^{*\prime}X^{*}\right)^{-1}
\end{aligned}
$$

Debemos destacar el hecho de que $Var\left(\hat{\beta}_{GLS}\right) \leq Var\left(\hat{\beta}\right)$, por lo que se cumple que:

$$
\left(X'\Omega^{-1}X\right)^{-1} - \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1}
$$

es una matriz semidefinida negativa. Este resultado es lógico por cuanto ahora el vector de estimadores $\hat{\beta}_{GLS}$ es eficiente, mientras que $\hat{\beta}$ no lo es. Sin embargo, en ocasiones la varianza de ambos vectores de estimadores puede coincidir.

Por otro lado, la estimación del parámetro de dispersión se puede obtener de la siguiente forma:

$$
\hat{\sigma}^2_{GLS} = \frac{\hat{u}^{*\prime}\hat{u}^{*}}{T-k}
$$

donde $\hat{u}^{*\prime}\hat{u}^{*}$ es la suma residual procedente de la estimación OLS del modelo (2). En función de las variables del modelo original, este estimador se puede expresar como sigue:

$$
\hat{\sigma}^2_{GLS} = \frac{\hat{u}^{*\prime}\hat{u}^{*}}{T-k} = \frac{\hat{u}'D'D\,\hat{u}}{T-k} = \frac{\hat{u}'\Omega^{-1}\,\hat{u}}{T-k} = \frac{u'\,M\,\Omega^{-1}M\,u}{T-k}
$$

donde $\hat{u} = y - X\hat{\beta}_{GLS}$.

En consecuencia, si la matriz Ω es conocida a priori, podemos obtener un estimador eficiente sin más que aplicar OLS al modelo transformado. En algunos casos concretos, podemos asumir que la matriz Ω puede ser conocida a priori. Esto no es cierto con tintes de generalidad, por lo que el método de estimación presentado en esta sección no es muy útil desde el punto de vista aplicado. Por tanto, debemos seguir buscando métodos que nos ofrezcan la posibilidad de estimar los parámetros del modelo en presencia de una matriz de varianzas y covarianzas escalar.

4 Míninos cuadrados generalizados factibles.

El método presentado en la sección anterior carece de aplicabilidad por cuanto parte de la base de que la matriz Ω es conocida a priori. Sin embargo, esto no es así. Una posible solución a este problema es estimar los elementos de la matriz Ω. No obstante, esto no es una tarea sencilla, por cuanto dado que la matriz Ω es de dimensión T, entonces es necesario estimar $\frac{1}{2}T\left(T+1\right)$ elementos. Si tenemos en cuenta que disponemos de tan sólo T observaciones, no podemos realizar esta estimación sin imponer algún tipo de restricción que disminuya el número de parámetros a estimar.

Supongamos que disponemos de una matriz $\hat{\Omega}$ tal que no ofrece una estimación consistente de la matriz Ω. En este caso, $\operatorname{plim}\hat{\Omega}_{ij} = \Omega_{ij}\ \forall i,j$. A partir, de esta matriz, podemos definir un nuevo estimador:

$$\hat{\beta}_{FG} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}X'\hat{\Omega}^{-1}Y$$

Se observa que este nuevo vector de estimadores, conocido como mínimo cuadrático generalizado factible, se obtiene sin más que sustituir la matriz $\hat{\Omega}$ por Ω en (3). Sus propiedades difieren notablemente de las del estimador generalizado y, en general, dependen fundamentalmente de la calidad de la matriz $\hat{\Omega}$.

4.1 Propiedades

1) El vector de estimadores $\hat{\beta}_{FG}$ no es (en general) ni insesgado, ni ELIO. Sin embargo, si asumimos que la distribución del vector de perturbaciones es simétrica, entonces para una amplia gama de casos $\hat{\beta}_{FG}$ será insesgado, tal y como se demuestra que Andrews (1986), *Econometrica*.
2) El vector de estimadores $\hat{\beta}_{FG}$ es consistente. Su demostración es sencilla. Basta con son considerar que:

$$\hat{\beta}_{FG} = \beta + \left(X'\hat{\Omega}^{-1}X \right)^{-1} X'\hat{\Omega}^{-1}u$$

Entonces:

$$\text{p}\lim \hat{\beta}_{FG} = \beta + \text{p}\lim \left(\frac{X'\hat{\Omega}^{-1}X}{T} \right)^{-1} \text{p}\lim \left(\frac{X'\Omega^{-1}u}{T} \right).$$

Si tenemos en cuenta que:

$$\text{p}\lim \left(\frac{X'\hat{\Omega}^{-1}X}{T} \right)^{-1} = \Omega_X$$

y

$$\text{p}\lim \left(\frac{X'\Omega u}{T} \right) = 0.$$

Es directo obtener que:

$$\hat{\beta}_{FG} \rightarrow_p \beta.$$

En general, si $\hat{\Omega} \rightarrow_p \Omega$, entonces $\hat{\beta}_{FG}$ tiene la misma distribución asintótica que $\hat{\beta}_{GLS}$ (cuando Ω es conocido). Un contraejemplo de esta afirmación se encuentra en Schmidt, p. 69.

3) Una condición suficiente para que el estimador generalizado y el generalizado factible tengan la misma distribución asintótica es que se cumpla:

$$\text{p}\lim X' \left(\hat{\Omega}^{-1} - \Omega^{-1} \right) X/T = 0$$

y

$$\text{p}\lim X' \left(\hat{\Omega}^{-1} - \Omega^{-1} \right) u/\sqrt{T} = 0$$

o

$$\text{p}\lim X'\Omega^{-1}X/T = \text{p}\lim X'\hat{\Omega}^{-1}X/T$$

y

$$\text{p}\lim X'\Omega^{-1}u/\sqrt{T} = \text{p}\lim X'\hat{\Omega}^{-1}u/\sqrt{T}.$$

5 Using OLS with corrected standard errors.

As we will see, using another estimation procedure to account for the possibility of heteroskedasticity and/or serial correlation in the residuals (non-spherical errors) usually entails assuming a particular parametric structure for the data generating process of the residuals (e.g., assuming the errors to be AR(1)). It is often seen as difficult or unrealistic to assume that we have a good enough idea of the process generating the residuals. If our guess is right, then using a feasible generalized least-squares approach (to be discussed later) will lead to more efficient estimates. However, making a wrong assumption on this structure is likely to lead to outright inconsistent estimators. Thus, in many cases, a sensible strategy to adopt is to be agnostic about this structure, i.e. impose as few conditions as possible, and use the standard OLS estimates with a corrected estimate for the covariance matrix. If we have such a consistent estimate, we will have consistent estimates of the parameters and valid asymptotic inference, though these will be less efficient than if we had "guessed" the right structure for the residuals. In this sense, OLS with a corrected covariance matrix is robust to a wide class of possible specifications for the errors. Feasible GLS is, on the other hand, more efficient if the right structure is assumed but may be inconsistent altogether if not. This is a standard trade-off between robustness and efficiency.

To see how to obtain an asymptotically valid estimate of the covariance matrix of the parameter estimate $\hat{\beta}$ let us look at this limiting covariance matrix. It is given by:

$$V = p\lim \left(\frac{X'X}{T}\right)^{-1} \frac{X'\Omega X}{T} \left(\frac{X'X}{T}\right)^{-1}$$

Since the only unknown quantity is Ω, a natural strategy would be to replace Ω by an estimator, say $\hat{\Omega}$. However, no such estimator can be consistent for Ω since it is a T by T matrix and, hence, its dimension increases with the sample size (there are $T(T+1)/2$ elements to estimates which cannot be done with only T observations). However, the matrix

$$S = p\lim \frac{X'\Omega X}{T}$$

is only of dimension k by k and, hence, could potentially be consistently estimated. Let us suppose that we have such a consistent estimator, i.e. we have an estimate $\hat{S}_T$ such that

$$p\lim \hat{S}_T = S = p\lim \frac{X'\Omega X}{T}.$$

Then

$$\hat{V}_T = \left(\frac{X'X}{T}\right)^{-1} \hat{S}_T \left(\frac{X'X}{T}\right)^{-1}.$$

is a consistent estimate of V and we could obtain the appropriate variance-covariance matrix of $\hat{\beta}_{OLS}$ and carry proper tests of hypothesis (valid asymptotically).

5.1 What is $p\lim \frac{X'\Omega X}{T}$?

Before trying to find a consistent estimate of S it is natural to ask first what is $S = p\lim \frac{X'\Omega X}{T}$? To do so, let us take the simple case with a single regressor (the results extend easily to the general case). Hence, the model is

$$y = \beta x + u$$

where x is a single regressor. We use the notation

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \cdots & \cdots & \omega_{1T} \\ \omega_{21} & \omega_{22} & & & \omega_{2T} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \omega_{T1} & \omega_{T2} & \cdots & \cdots & \omega_{TT} \end{pmatrix}$$

where

$$\omega_{ij} = E(u_i u_j),$$

and, by the symmetry of Ω,

$$\omega_{ij} = \omega_{ji}.$$

Then,

$$
\begin{aligned}
x'\Omega x &= \sum_{i=1}^{T}\sum_{j=1}^{T}\omega_{ij}x_i x_j \\
&= \sum_{i=1}^{T}\sum_{j=1}^{T}E(u_i u_j)x_i x_j \\
&= \sum_{i=1}^{T}\sum_{j=1}^{T}E[(u_i x_i)(u_j x_j)]
\end{aligned}
$$

using the fact that we assume fixed x's. Now let $v_t = u_t x_t$ and note that $E(v_t) = 0$. Assume, for simplicity that the sequence of random variables v_t is stationary. For example, x_t could be a constant and u_t a stationary and invertible $ARMA$ process. This assumption is made only for the sake of clarity of exposition and all the procedures that we shall discuss are applicable to the general case where heteroskedasticity and some form of non-stationarity is present in the process v_t. The derivations in this more general case are just more cumbersome. Then

$$
x'\Omega x = \sum_{i=1}^{T}\sum_{j=1}^{T}R_v(i-j) \tag{4}
$$

where $R_v(\tau)$ is the covariance function of v_t at delay τ. Consider now the following matrix

$$
\begin{pmatrix}
R_v(0) & R_v(1) & \cdots & \cdots & R_v(T-1) \\
R_v(1) & R_v(0) & \cdots & \cdots & R_v(T-2) \\
\vdots & & & & \vdots \\
\vdots & & & & \vdots \\
R_v(T-1) & R_v(T-2) & \cdots & \cdots & R_v(0)
\end{pmatrix} \tag{5}
$$

where we have used the fact that the autocovariance function is an even function, i.e. $R_v(\tau) = R_v(-\tau)$. Note that the double summation given by (4) is equivalent to adding each element of the matrix (5). To simplify this expression, we can use the fact that many of the elements of the matrix (5) are identical, namely those on a given diagonal. Then we can write (adding the sums of the diagonals)

$$x'\Omega x = TR_v(0) + 2\sum_{j=1}^{T-1}(T-j)R_v(j).$$

Hence,

$$\frac{x'\Omega x}{T} = R_v(0) + 2\sum_{j=1}^{T-1}(\frac{T-j}{T})R_v(j)$$

and

$$\begin{aligned} p\lim\frac{x'\Omega x}{T} &= R_v(0) + 2p\lim\sum_{j=1}^{T-1}(\frac{T-j}{T})R_v(j) \\ &= R_v(0) + 2p\lim\sum_{j=1}^{T-1}R_v(j) \end{aligned}$$

assuming $p\lim\sum_{j=1}^{T-1}R_v(j)$ is finite. Hence:

$$\begin{aligned} p\lim\frac{x'\Omega x}{T} &= R_v(0) + 2\sum_{j=1}^{\infty}R_v(j) \\ &= \sum_{j=-\infty}^{\infty}R_v(j) \end{aligned}$$

Remark 1 *For those who are familiar with the theory of spectral analysis, $p\lim\frac{x'\Omega x}{T}$ is therefore equivalent to (2π) times the spectral density function of the v_t (the pair $x_t u_t$) at frequency zero.*

5.2 How to consistently estimate $p\lim\frac{x'\Omega x}{T}$?

A useful expression to start searching for a consistent estimate of $p\lim\frac{x'\Omega x}{T}$ is:

$$p\lim\frac{x'\Omega x}{T} = R_v(0) + 2p\lim\sum_{j=1}^{T-1}R_v(j). \tag{6}$$

The autocovariance function $R_v(j)$ is unknown but can consistently be estimated by

$$\hat{R}_v(j) = T^{-1}\sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}$$

where

$$\hat{v}_t = x_t \hat{u}_t$$

and

$$\hat{u}_t = y_t - \hat{\beta} x_t$$

are the residuals from the OLS regression. We have

$$p \lim \hat{R}_v(j) = R_v(j) \tag{7}$$

basically because the OLS estimate $\hat{\beta}$ are consistent for β and, hence, the OLS residuals $\hat{u}_t$ are good estimates of the true residuals u_t.

Using (6) and (7), we can think of consistently estimating $p \lim \frac{x'\Omega x}{T}$ by

$$\hat{S}_{0,T} = \hat{R}_v(0) + 2 \sum_{j=1}^{T-1} \hat{R}_v(j).$$

It is easy to see, however, that such an estimator is asymptotically unbiased but inconsistent. The basic reason is that it is constructed using the estimates of T quantities. Since we have only T observations, there is not enough information to get good estimates for all the lagged autocovariances. The idea is to get rid of those autocovariances for which we have the least information, i.e. the higher lagged ones. This leads to the estimate

$$\hat{S}_{1,T} = \hat{R}_v(0) + 2 \sum_{j=1}^{m} \hat{R}_v(j),$$

where m is some truncation lag parameter. $\hat{S}_{1,T}$ will be a consistent estimate provided

$$m \to \infty \text{ and } \frac{m^2}{T} \to 0,$$

as $T \to \infty$. The requirement that $m \to \infty$ is to ensure that the estimate is asymptotically unbiased since as T increases all lagged autocovariances are used. The requirement that $m^2/T \to 0$ is to ensure that as the sample size increases there is still an infinite amount of information for all quantities being estimated, and this at a fast enough rate.

The estimate $\hat{S}_{1,T}$ still does not take into account the fact that the autocovariances at lag close to m are not as precisely estimated as the autocovariances at lags close to 0. One way to take into account this feature is to put linearly declining weights on the lagged autocovariances. This leads to the estimate

$$\hat{S}_{2,T} = \hat{R}_v(0) + 2\sum_{j=1}^{m}(1 - j/(m+1))\hat{R}_v(j).$$

In the context of spectral analysis, this particular weighting scheme was proposed by Bartlett. A feature of importance is that, unlike $\hat{S}_{1,T}$, $\hat{S}_{2,T}$ ensures a non-negative estimate.

In general, the use of some truncation or the linearly declining weights are quite arbitrary (though sensible). More generally what is required is that we give less weight to higher lagged autocovariances. Hence, we can think of a general class of estimators given by

$$\hat{S}_{w,T} = \hat{R}_v(0) + 2\sum_{j=1}^{T-1} w(j,m)\hat{R}_v(j),$$

where $w(j,m)$ is some weighting function. Following the literature on spectral analysis, the parameter m is called a bandwidth. If the weight function is such that

$$w(j,m) = 0 \text{ if } j > m$$

then m acts as a truncation lag parameter. Otherwise it dictates the general shape of the weighting function. Using the notation $x = j/m$, examples of weighting functions are:

1. Bartlett's triangular window:

$$\begin{aligned} w_{BT}(x) &= 1 - |x| && \text{if } |x| \le 1 \\ &= 0 && \text{otherwise} \end{aligned}$$

2. Parzen's window:

$$\begin{aligned} w_{PR}(x) &= 1 - 6x^2 + 6|x|^3 && \text{if } 0 \le |x| \le 1/2 \\ &= 2(1-|x|)^3 && \text{if } 1/2 \le |x| \le 1 \\ &= 0 && \text{otherwise} \end{aligned}$$

3. Tuckey-Hanning's window:

$$\begin{aligned} w_{TH}(x) &= (1 + \cos(\pi x))/2 && \text{if } |x| \le 1 \\ &= 0 && \text{otherwise} \end{aligned}$$

4. Quadratic Spectral window:

$$w_{QS}(x) = (\sin(\delta)/\delta - \delta)/(3\delta^2)$$

where $\delta = 6\pi x/5$.

Note that, except for the Quadratic Spectral window, the autocovariances at lags greater than m receive zero weight. Hence, the bandwidth acts as a truncation lag parameter. For the Quadratic Spectral window all lags of the sample autocovariance function are considered and the bandwidth parameter m acts to change the shape of the weight function.

5.3 The estimate in the general case with k regressors.

In the general case with k regressors,

$$S_T = \frac{X'\Omega X}{T}$$

is a k by k matrix the class of estimators is given by:

$$\hat{S}_{w,T} = (\frac{T}{T-k}) \sum_{j=-T+1}^{T-1} w(j/m)\hat{\Gamma}_v(j)$$

where

$$\begin{aligned} \hat{\Gamma}_v(j) &= T^{-1} \textstyle\sum_{t=j+1}^{T} \hat{v}_t \hat{v}'_{t-j} \quad \text{for } j \geq 0 \\ &= T^{-1} \textstyle\sum_{t=-j+1}^{T} \hat{v}_{t+j} \hat{v}'_t \quad \text{for } j < 0 \end{aligned}$$

and

$$\hat{v}_t = X_t \hat{u}_t$$

with

$$\hat{u}_t = y_t - X_t \hat{\beta}.$$

Remark 2 *The scaling factor $T/(T-k)$ is an optional small sample correction.*

5.4 Data dependent method to select m.

The following material is based on Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858. In that paper, Andrews showed that the Quadratic Spectral kernel is the window that minimizes the asymptotic mean square error in the class of windows that yield estimators that are positive semi-definite. As in many studies, he also found that using the Bartlett weights yields an estimator with rather poor properties.

To be more precise, we now write the bandwidth as m_T to highlight the fact that it should be a function of the sample size T. The asymptotic mean-squared error is a rather complicated function of:

1. the weight function $w(\cdot)$;
2. the bandwidth m_T;
3. the true data generating process of $v_t = X_t u_t$ through some summary measure which we shall call α, say.

So we have

$$MSE = f(w(\cdot), m_T, \alpha). \tag{8}$$

For a given data generating process and window, a value m_T^* can be evaluated which minimizes the asymptotic mean-squared error. Then an estimate $\hat{m}_T$ can be obtained using the so-called "plug-in" method. This method works as follows. Take a data-generating process which you believe is a good approximation to the true data generating process of $v_t = X_t u_t$. Then one can construct an estimate of the summary measure α as $\hat{\alpha}$ (this is discussed below). One can then solve for the value $\hat{m}_T$ which minimizes the function (8) to get

$$\hat{m}_T = g(\hat{\alpha}, w(\cdot)).$$

The resulting functions are the following for some popular windows:

1. Parzen: $\hat{m}_T = 2.6614(\hat{\alpha}T)^{1/5}$;
2. Tuckey-Hanning: $\hat{m}_T = 1.7462(\hat{\alpha}T)^{1/5}$;
3. Quadratic Spectral: $\hat{m}_T = 1.3221(\hat{\alpha}T)^{1/5}$.

To describe the possible methods to construct $\hat{\alpha}$ note first that v_t is a k dimensional vector and we can therefore approximate the data generating process of v_t using k different univariate data-generating processes. One could also use a multivariate model, but the method is then more complex. Consider approximating each element of the vector v_t by an autoregression of order 1. That is, we consider the following regression for the ith element

$$\hat{v}_{i,t} = \rho_i \hat{v}_{i,t-1} + e_{i,t}$$

Note that no constant need be included since $v_{i,t}$ has mean zero. Denote the estimates by $\hat{\rho}_i$ and $\hat{\sigma}_i^2$. Then:

$$\hat{\alpha} = \sum_{i=1}^{p} \frac{4\hat{\rho}_i^2 \hat{\sigma}_i^4}{(1-\hat{\rho}_i)^8} / \sum_{i=1}^{p} \frac{\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^4}.$$

In general, $k = p$ but if one considers a linear model with a constant included as a regressor, Andrews suggests not to include the component of v_t associated with the constant. The argument is that the estimator is then scale invariant. So here $p = k - 1$.

In the case where the X_t matrix is only a constant, then

$$\hat{\alpha} = 4\hat{\rho}^2/(1-\hat{\rho})^4$$

where $\hat{\rho}$ is the OLS estimate in the regression

$$(y_t - \bar{y}) = \rho(y_{t-1} - \bar{y}) + e_t.$$

In this case, the estimator is scale invariant automatically.

In some cases, using the AR(1) approximation for the data-generating process of v_t can give poor results, for example when a strong negative moving-average component is present. Then, a recommendation is to fit an ARMA(1,1) process of the form

$$\hat{v}_{i,t} = \rho_i \hat{v}_{i,t-1} + e_{i,t} + \psi_i e_{i,t-1}$$

with estimates denoted by $\hat{\rho}_i$, $\hat{\psi}_i$ and $\hat{\sigma}_i^2$. The expression for $\hat{\alpha}$ is then:

$$\hat{\alpha} = \sum_{i=1}^{p} \frac{4(1+\hat{\rho}_i\hat{\psi}_i)^2(\hat{\rho}_i+\hat{\psi}_i)^2\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^8} / \sum_{i=1}^{p} \frac{(1+\hat{\psi}_i)^4\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^4}.$$

Remark 3 *Not much is known about the small sample properties of $\hat{\beta}_{FG}$ compared to $\hat{\beta}_{OLS}$. The presumption is usually that using $\hat{\beta}_{FG}$ permits efficiency gains compared to using OLS and correcting for the standard errors. But it is not clear that this is, in general, the case.*