Teoría Econométrica I - EAE- 250-A

Problemas de Especificación y Datos II

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Septiembre/Octubre 2021

Introducción

- En la sección pasada relajamos el supuesto de matriz de varianzas esférica, centrándonos en Heteroscedasticidad y vimos:

  – Método para estimar nuestros parámetros: MCG/MCGF (y su validez para hacer pruebas de hipótesis)

  – Varios test de Heteroscedasticidad

  – Estimar por MCO y estimar la matriz de varianza y covarianzas con formula de White
- En esta sección veremos otros problemas de especificación del modelo de regresión lineal
- En particular, qué sucede cuándo la matriz de varianzas es no diagonal?
- Veremos también Mínimos Cuadrados No lineales y Multicolinealidad

Autocorrelación

- La matriz de varianzas y covarianzas es no escalar y además no diagonal
- Ahora el índice indexará tiempo y los errores estarán serialmente correlacionados.

$$y_t = x_t\beta + \varepsilon_t \qquad t = 1, 2, ..., T$$

- Donde $covar(\varepsilon_t, \varepsilon_s) \neq 0 \Longrightarrow$ la matriz de varianza-covarianza tendrá elementos fuera de la diagonal.
- La estructura del residuo del modelo se asume estacionaria (es decir, que el 1^{er} y el 2^{do} momento son finitos, que no dependen de t y que la covarianza entre ϵ_t y ϵ_s depende de la diferencia $t - s$).
- También podemos pensar en un subíndice r y pensar en correlación espacial

AR(1)

- Asumiremos un modelo simple, con una estructura de autocorrelación de orden 1 ($AR(1)$) para los errores, es decir:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \mu_t$$

- Donde decimos que $\mu_t \sim iid(0, \sigma^2_\mu)$, o que es un ruido blanco (RB).
- Estacionariedad en este modelo implica que: $\rho \in (-1, 1)$, es decir, vive dentro del círculo unitario.
- Esto además implica que $\mathbb{E}(\varepsilon_t) = 0$.
- Lo podemos verificar fácilmente

$$\begin{aligned}\mathbb{E}(\epsilon_t) = \rho\mathbb{E}(\epsilon_{t-1}) + \mathbb{E}(\mu_t)\\ \mathbb{E}(\epsilon)(1-\rho) = 0\\ \mathbb{E}(\epsilon) = 0\end{aligned}$$

donde la segunda igualdad viene del supuesto de estacionariedad (primer momento no depende de t) y del supuesto de que μ_t es RB.

AR(1)

- Una manera alternativa es usar el operador de rezagos L donde $L\epsilon_y = \epsilon_{t-1}$. así,

$$\varepsilon_t(1-\rho L) = \mu_t \iff \varepsilon_t = \frac{\mu_t}{1-\rho L} \implies \mathbb{E}(\varepsilon_t) = \frac{1}{1-\rho L}\mathbb{E}(\mu_t) = 0$$

- Además

$$var(\varepsilon_t) = \mathbb{E}(\varepsilon_t - \mathbb{E}(\varepsilon_t))^2 = \mathbb{E}(\varepsilon_t^2).$$

$$\varepsilon_t^2 = \rho^2\varepsilon_{t-1}^2 + \mu_t^2 + \rho\varepsilon_{t-1}\mu_t$$

- Por lo tanto, $\mathbb{E}(\varepsilon_t^2) = \rho^2\mathbb{E}(\varepsilon_{t-1}^2) + \mathbb{E}(\mu_t^2) + \rho\mathbb{E}(\varepsilon_{t-1}\mu_t)$
- Pero $\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\varepsilon_{t-1}^2)$, y además, $\mathbb{E}(\varepsilon_{t-1}\mu_t) = 0$, por lo que:

$$(1-\rho^2)\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\mu_t^2)$$

$$(1-\rho^2)\mathbb{E}(\varepsilon_t^2) = \sigma_\mu^2$$

$$\mathbb{E}(\varepsilon_t^2) = \frac{\sigma_\mu^2}{(1-\rho^2)}$$

AR(1)

- De esta manera, tenemos que:

$$var(\varepsilon_t) = \frac{\sigma_\mu^2}{1-\rho^2}$$

- Por otro lado, necesitamos encontrar una expresión para $covar(\varepsilon_t, \varepsilon_s) = \mathbb{E}(\varepsilon_t \varepsilon_s)$.
- Note que $\varepsilon_t = \rho^{|t-s|}\varepsilon_s + \sum_{i=0}^{|t-s|-1} \rho^i \mu_{t-i}$, entonces:

$$\begin{aligned}\mathbb{E}(\varepsilon_t \varepsilon_s) &= E\left[\rho^{|t-s|}\varepsilon_s\varepsilon_s + \sum_{i=0}^{|t-s|-1} \rho^i \mu_{t-i}\varepsilon_s\right] \\ &= \rho^{|t-s|}\frac{\sigma_\mu^2}{1-\rho^2}\end{aligned}$$

AR(1)

- De esta manera, tenemos que:

$$var(\varepsilon) = \frac{\sigma_\mu^2}{1-\rho^2}\Omega = \frac{\sigma_\mu^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}$$

- Podemos descomponer Ω^{-1}, dado que es una matriz simétrica, en $H'H$, con:

$$H = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}$$

AR(1)

- Pudiendo así transformar el modelo original, siendo $Y^* = HY$, y $X^* = HX$, tendríamos que:

$$\hat{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \Longleftrightarrow \hat{\beta}_{MCG} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$$

- ¿Qué ocurre si hacemos MCO? Ocurre más error tipo I del que quisiéramos, y resulta ineficiente, dado que:

$$\begin{aligned} var(\hat{\beta}_{MCO}) &= (X'X)^{-1}X'var(Y)X(X'X)^{-1} \\ &= \frac{\sigma^2}{1-\rho^2}(X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

- y $var(\hat{\beta}_{MCO})$ =

AR(1)

- Entonces si el signo de la cov de los ε es el mismo que el de los $X \rightarrow$ $var(\hat{\beta}_{MCO})$ más pequeños $\rightarrow$ intervalos mas angostos $\rightarrow$ Rechazaremos mas veces la nula $\rightarrow$ más error de tipo I

  ¿Cómo estimamos ρ?
- MCO $\rightarrow$ Estimamos el residuo $\rightarrow$ Regresionamos el residuo con su rezago. Es decir:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon_{t-1}}}{\sum_{t=2}^{T} \hat{\varepsilon_{t-1}}^2}$$

  Donde $\hat{\varepsilon}_t = y_t - x_t' \hat{\beta}_{MCO}$
- ¿Porque partimos de $t-2$?
- De esta manera, tenemos que la distribución asintótica de $\hat{\rho}$ se puede escribir:

$$\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{d} N(0, 1-\rho^2)$$

  Donde queda propuesto demostrar que $var(\hat{\rho}) = 1 - \rho^2$

Tests de autocorrelación

- Siendo la hipótesis nula $\rho = 0$, tendríamos que bajo ella: $\sqrt{T}\hat{\rho} \xrightarrow{d} N(0,1)$.
- En el caso en que $\rho \approx 1$ esta aproximación asintótica no es válida y $\hat{\rho}$ no es asintóticamente normal sino que sigue una distribución Dickey-Fuller (próximo curso)
- El test natural para la hipótesis nula de ausencia de autocorrelación es un test-t, pero este es un test asintóticamente normal.
- A diferencia de este test, el test por excelencia para la hipótesis nula de autocorrelación serial de primer orden es el de Durbin-Watson, que es un test exacto (su distribución es exacta, no asintótica)
- Vamos a ver tests que tengan como nula la ausencia de autocorrelación

Durbin Watson

- La hipótesis nula en estos tests será siempre de ausencia de autocorrelación y la alternativa de presencia de autocorrelación de algún orden
- Durbin y Watson proponen:

$$DW = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon_{t-1}})^2}{\sum_{t=2}^{T} \hat{\varepsilon_t}^2} \simeq 2(1 - \hat{\rho})$$

- tarea: demostrar $DW \simeq 2(1 - \hat{\rho})$
- Y encuentran valores críticos para esta prueba que no dependen de la estructura de los datos.
- Pero son cotas: superior e inferior
- Y quedan regiones de indefinición en el medio

Durbin Watson

- Dado esto, tenemos que $DW \in (0,4)$, y su estructura sería la siguiente:

$$DW \to 0 \Longrightarrow \rho \to 1 \Longrightarrow \text{Autocorrelación}(+)$$
$$DW \to 4 \Longrightarrow \rho \to -1 \Longrightarrow \text{Autocorrelación}(-)$$

- Al definir cotas superior e inferior, el test de Durbin-Watson satisface lo siguiente

$$\begin{aligned} Pr(DW < d_L) &\leq \alpha \\ Pr(DW > d_U) &\leq 1 - \alpha \end{aligned}$$

- Un test exacto a una cola con significancia α rechazará en favor de autocorrelación positiva si

$$H_0{:}\rho = 0$$
$$H_1'{:}\rho > 0 \text{ Rechaza si } DW < d_L$$

y es inconclusivo cuando $d_L \leq DW \leq d_U$.

Durbin Watson

- Por otra parte y dada la simetría, el test rechaza en favor de autocorrelación negativa si

$$H_0{:}\rho = 0$$
$$H_1{:}\rho < 0 \text{ Rechaza si } DW > 4 - d_L$$

- y es inconclusivo cuando $4 - d_U \leq DW \leq 4 - d_L$.
- Si se está en una zona inconclusiva podemos usar el test-t (que es asintótico pero equivalente a DW cuando no es inconcluso) u otro test como Breuch-Godfrey

Breuch-Godfrey

- Alternativa para testear autocorrelaciones de ordenes superiores a 1 y se basa en el test LM introducido anteriormente
- La nula, al igual que en todos los test de autocorrelación, es que los residuos no se encuentran correlacionados
- Consideremos para distintos valores de k, el siguiente conjunto de estadísticos:

$$r_k = \frac{\sum_{t=1}^{n} \hat{\epsilon}_t \hat{\epsilon}_{t-k}}{\sum_{t=1}^{n} \hat{\epsilon}_t^2} \tag{1}$$

note que si k=1, entonces estamos en una caso parecido al estadístico DW.

Pasos Breuch-Godfrey

1. Estimar el modelo por MCO y obtener los residuos $\hat{u}$. El modelo puede incluir rezagos de la variable dependiente.
2. Estimar una regresión auxiliar de $\hat{u}_t$ sobre p rezagos: $\hat{\epsilon}_{t-1}, \ldots, \hat{\epsilon}_{t-p}$, incluyendo las variables exógenas (X) del modelo original. Note que deberá excluir p observaciones.
3. Calcular el R^2 de la regresión auxiliar
4. Construir el estadistico $nR^2 \sim \chi^2_p$

- Así, la prueba de hipótesis la escribimos:

Test de Box-Pierce-Ljung (Q-Stat)

- Este test se basa en el cuadrado de las primeras p autocorrelaciones de los residuos MCO. El estadístico se define como:

$$Q = n \sum_{j=1}^{p} r_j^2 \tag{2}$$

donde:

$$r_j = \frac{\sum_{t=j+1}^{n} \hat{\epsilon}_t \hat{\epsilon}_{t-j}}{\sum_{t=1} \hat{\epsilon}_t^2}$$

- La distribución del estadístico bajo la nula de no-autocorrelación es χ^2 con grados de libertad igual a p menos el número de rezagos del error incluidos en la especificación autorregresiva del error.

Estimación: MCGF/FGLS

- Como vimos anteriormente en presencia de autocorrelación $Var(\epsilon) = \Omega \frac{\sigma_\mu^2}{1-\rho^2}$ donde la matriz Ω es:

$$\Omega = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$$

Se puede demostrar que la matriz H en este caso es:

$$H = P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

Modelo transformado

- Para aplicar FGLS transformamos el modelo premultiplicando por H
- Nos queda entonces:

- Para la primera observación tenemos:

$$\sqrt{1-\rho^2}y_1 = (\sqrt{1-\rho^2})x_1'\beta + (\sqrt{1-\rho^2})\epsilon_1 \tag{3}$$

- Y para el resto de las $(T-1)$ observaciones:

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + \underbrace{\epsilon_t - \rho\epsilon_{t-1}}_{u_t} \tag{4}$$

Cochrane Orcutt

- Si es que nuestro modelo es el siguiente:

$$\begin{aligned} y_t &= x_t\beta + \epsilon_t \\ \epsilon_t &= \rho\epsilon_{t-1} + u_t \end{aligned}$$

El modelo transformado es de la siguiente forma:

$$\begin{aligned} \underbrace{y_t - \rho y_{t-1}}_{y_t^*} &= \underbrace{\left(x_t - \rho x_{t-1}\right)}_{x_t^*}\beta + \underbrace{\epsilon_t - \rho\epsilon_{t-1}}_{u_t} \\ \Rightarrow y_t^* &= x_t^*\beta + u_t \end{aligned}$$

menos la primera observación.

Cochrane Orcutt: método iterativo

Cochrane-Orcutt es un método iterativo para obtener la estimación de β y ρ:

1. Estimar por Mínimos Cuadrados Ordinarios la regresión de interés
2. Utilizar los residuos MCO para estimar el parámetro ρ. Esto puede hacerse mediante una regresión de $\hat{\epsilon}_t$ contra $\hat{\epsilon}_{t-1}$, o a partir del estadístico DW de la estimación anterior.
3. Utilizar este parámetro $\hat{\rho}$ para transformar las variables, y obtener y_t^* y x_t^* (pierde una observación).
4. Estimar por MCO un modelo con las variables transformadas, para obtener un *nuevo* vector de coeficientes β.
5. Utilizar esta *nueva* estimación para computar otro vector de residuos, y utilizar estos residuos para obtener una *nueva* estimación de ρ
6. Repetir este procedimiento hasta que los β converjan. Por ejemplo, $||\beta_j - \beta_{j-1}|| < 10^{-5}$.

Prais-Winsten y el método de Durbin

- El método de **MCGF de Prais-Winsten (1954)** es igual al método de Cochrane Orcutt pero evita perder la primera observación transformando: $\sqrt{1-\rho^2}y_1$ y $\sqrt{1-\rho^2}x_1$

Método de Durbin

1. Use la transformación

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + u_t \quad t = 2, .., T$$

2. Estime la regresión (libre de autocorrelación)

$$y_t = \rho y_{t-1} + x_t\beta - \rho x_{t-1}\gamma + u_t \quad t = 2, .., T$$

   donde $\gamma = -\rho\beta$, obtenga $\hat{\rho}$.

3. Obtenga el estimador de MCGF:

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

Estimación por Máxima Verosimilitud: repaso

- Likelihood function (o función de máxima verosimilitud)
  - Es una función de la muestra y un set de parámetros desconocidos
  - Si sabemos cual es la distribución conjunta de nuestra muestra, hasta un vector de parámetros, sabemos la likelihood function (función de verosimilitud) hasta ese parámetro
  - Para escribir la likelihood function: basta escribir la densidad conjunta: En el caso de una muestra aleatoria será el producto de todas las marginales (idénticas)
- MLE o estimador de máxima verosimilitud
  - Tenemos una muestra, dada la likelihood function, ¿cuál es el valor de los parámetros que hace máxima la probabilidad de haber obtenido esta muestra?
  - Supuesto detrás: está probabilidad es máxima en el verdadero valor de los parámetros

Ejemplo usando distribución normal

- Si Tenemos una muestra aleatoria, y asumimos $\epsilon \sim N(0, \sigma^2)$

$$L(\theta; X) = \prod_{t=1}^{t=T} \frac{1}{\sigma\sqrt{2\Pi}} exp(\frac{-\epsilon_t^2}{2\sigma^2}) = (\frac{1}{\sigma\sqrt{2\Pi}})^T exp(\frac{-\sum_{t=1}^{t=T}\epsilon_t^2}{2\sigma^2})$$

donde $\epsilon_t = y_t - \beta x_t$

- Muchas veces usamos la $\log(L)$

$$l(\theta; X) = \sum_{t=1}^{t=T} \frac{-1}{2}\log(2\pi) - log(\sigma) - \frac{-1}{2\sigma^2}\epsilon_t^2$$

- Aquí $\theta = [\beta, \sigma^2]$
- Las condiciones de primer orden nos darán $\hat{\theta}_{MLE}$
- $\hat{\beta}_{MLE}$
- Tarea: $\hat{\sigma^2}_{MLE}$

ML para estimación en presencia de autocorrelación

- Tenemos

$$\begin{aligned} y_t &= x_t\beta + \epsilon_t \\ \epsilon_t &= \rho\epsilon_{t-1} + u_t \end{aligned}$$

- Escribimos la likelihood del modelo transformado con H , es decir usando u_t

$$l(\theta; X) = \frac{-T}{2}\ln(2\pi) - T\ln\left(\frac{\sigma_\mu}{\sqrt{1-\rho^2}} - \frac{-(1-\rho^2)}{2\sigma_\mu^2}\sum_{t=1}^{t=T} u_t^2\right.$$

- Reemplazamos u por el error autocorrelacionado

$$l(\theta; X) = \frac{-T}{2}\ln(2\pi) - T\ln\left(\frac{\sigma_\mu}{\sqrt{1-\rho^2}}\right) - \frac{-(1-\rho^2)((1-\rho^2)\epsilon_1^2 - \sum_{t=2}^{T}(\epsilon_t - \rho\epsilon_{t-1}}{2\sigma_\mu^2}$$

- La ventaja de este método es que puedo estimar simultáneamente β y ρ

Estimación por OLS con estimación consistente de varianzas y Covarianzas

- Estimamos por MCO (que es consistente) y corregir su matriz de varianzas y covarianzas con **estimador de Newey-West (1987)**

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

si $V(u) = \Sigma = \sigma^2\Omega$

$$Var(\hat{\beta}_{MCO}|X) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

luego,

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, D^{-1}VD^{-1})$$

donde

$$\begin{aligned} D &= \operatorname{plim}\frac{1}{T}X'X \\ V &= \operatorname{plim}\frac{1}{T}X'\Sigma X \end{aligned}$$

Estimando σ

- El estimador consistente de V es:

$$\hat{V} = S_0 + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T} w(l)\hat{\epsilon}_t\hat{\epsilon}_{t-l}(x_t x'_{t-l} + x_{t-l}x'_t) \qquad (5)$$

- donde $w(l) = 1 - l/(L+1)$ es un ponderador que le da menos peso a las observaciones cerca de T y L corresponde al orden máximo de autocorrelación del término de error (que no siempre es fácil de determinar).

$$S_0 = \frac{1}{T}\sum_{t=1}^{T} e_t^2 x'_t x_t$$

- Note que sigue la misma idea de Eicker-Huber-White.

Clustered standard errors

- FGLS permite una forma específica de correlación (dada por la matriz Ω) pero su consistencia dependerá de que esta esté correctamente especificada
- Newey-West nos entrega un método general de corrección de Heterocedasticidad y Autocorrelación que implica estimar muchos parámetros fuera de la diagonal de Ω.
- **Clustering** es una solución entre medio de Newey-West y FGLS: no especifica la forma de la correlación pero le pone un límite, sólo puede existir dentro de un “cluster” o grupo
- las unidades dentro de un cluster deben compartir algo en común, ya sea geografía o incluso ser la misma en distintos momentos del tiempo en el caso de panel

Clustered standard errors

- Formalizando la idea, suponga tiene datos individuales pero cada individuo m pertenece a un **grupo** g

$$\{(y_{gm}, x_g, z_{gm}) : m = 1, ..., M_g\}$$

- El modelo,

$$y_{gm} = \alpha + x_g\beta + z_{gm}\gamma + v_{gm}$$

  con $m = 1, ..., M_g$ y $g = 1, ..., G$. Además tenemos k regresores grupales (x) y L individuales (z)

- Ejemplo: un total de G cursos y M_g alumnos para cada curso g.
- El modelo lineal homocedástico asume v_{gm} iid. La corrección de Eicker-Huber-White permite heterocedasticidad, pero las correlaciones son 0. ¿Podemos relajar dicho supuesto?
- Supongamos que dentro de cada grupo $E(v_{gm}v_{gl}) \neq 0$. Sabemos que OLS seguirá siendo consistente pero no eficiente.

Clustered standard errors

- Podemos generalizar Eicker-Huber-White permitiendo que dichas correalaciones sean distintas de cero. Básicamente procedemos de similar forma, pero tomando en cuenta que tenemos individuos y grupos.
- Supongamos que G es grande. Además supongamos que los regresores son exógenos:

$$E(v_{gm}|x_g, z_{gm}) = 0$$

- Podemos hacer un pooled OLS y el estimador OLS de $\lambda = (\alpha, \beta', \gamma')$ es consistente si $G \longrightarrow \infty$ y M_g fijo.

Clustered standard errors

- Veamos una matriz de varianza y covarianzas robusta. Sea W_g la matriz de $M_g \times (1 + K + L)$ que incluye todos los regresores y todas las observaciones para el grupo g. Sea y_g el vector de $M_g \times 1$ de la variable dependiente para el grupo g. Así $y_g = W_g\lambda + v_g$
- note que si apilamos W_g e y_g para todos los grupos ($y = W\lambda + v$), el estimador OLS es

$$\hat{\lambda} = (W'W)^{-1}W'y = \left(\sum_{g=1}^{G} W_g'W_g\right)^{-1}\left(\sum_{g=1}^{G} W_g'y_g\right)$$

- Luego, la matriz de varianzas y covarianzas de $\hat{\lambda}$ es

$$Var(\hat{\lambda}) = (W'W)^{-1}W'E(vv')W(W'W)^{-1}$$

- Uno podría intentar seguir una alternativa tipo Newey-West (generalización de Eicker-Huber-White)

$$Var(\hat{\lambda} = (W'W)^{-1}W'\hat{v}\hat{v}'W(W'W)^{-1}$$

pero hay un problema: $W'\hat{v} = 0$!!! No podemos rellenar todas las correlaciones $E(v_i v_j) = \hat{v}_i\hat{v}_j$

Clustered standard errors

- Una forma de bypassear ese problema es asumir que muchas covarianzas son cero, luego al imponer algo que en la muestra no ocurrirá podemos tratar de hacer la fórmula sandwich.
- Luego, si asumimos que $E(v_g v_h') = 0$ para $g \neq h$, podemos obtener

$$Var(\hat{\lambda}) = \left(\sum_{g=1}^{G} W_g' W_g \right)^{-1} \left(\sum_{g=1}^{G} W_g' \hat{v}_g \hat{v}_g' W_g \right) \left(\sum_{g=1}^{G} W_g' W_g \right)^{-1}$$

  donde $\hat{v}_g$ es el vector $M_g \times 1$ de residuos del pooled OLS para el grupo g.
- Esto es factible de estimar dado que supones que $E(v_g v_h') = 0$ para $g \neq h$ y el "relleno" del sandwich no es 0.
- Esto es lo que hace STATA cuando usamos el comando **cluster (group)** donde group es la variable de grupo.
- la consistencia de esta matriz ocurre cuando G es grande, estamos promediando sobre G grupos. Según Bertrand et al. (2004) con 50 funciona bien pero probablemente 30 sea un número todavía aceptable. Bajo la nula de homocedasticidad converge a lo mismo que la de Eicker-Huber-White