Taller de Tesis – MAE UdeSA

Modelo matemático

Tomás Pacheco

En esta entrega para el seminario de tesis lo que hay que hacer es detallar el modelo matemático que contribuya a la tesis. Como he mencionado en las entregas anteriores, mi trabajo consiste en testear empíricamente el estimador propuesto por Romano y Wolf (2018). Por lo tanto, en esta entrega, detallaré formalmente el marco teórico en el que se va a desarrollar mi trabajo de graduación.

Ya he mencionado muchas veces que esta tesis es sobre un estimador de la varianza, en otras palabras, es un trabajo sobre inferencia. La inferencia, en el campo en el que nosotros nos movemos, acompaña a la estimación puntual de un parámetro. Lo que haremos es definir el estimador más utilizado, el de Mínimos Cuadrados Ordinarios:

$$\hat{\beta}_{OLS} \coloneqq (X'X)^{-1}X'y.$$

Bajo los supuestos clásicos, sabemos que este estimador es el insesgado y consistente. Adicionalmente, suponiendo homocedasticidad condicional, es el más eficiente por el Teorema de Gauss-Markov. El problema es que deja de serlo bajo heterocedasticidad condicional; esto quiere decir, cuando al función cedástica $v(\cdot)$ no es constante.

Una forma de obtener un estimador más eficiente es reponderando los datos (y_i, x_i') y luego estimar por MCO el modelo transformado de la siguiente manera:

$$\frac{y_i}{\sqrt{v(x_i)}} = \frac{x_i'}{\sqrt{v(x_i)}}\beta + \frac{\varepsilon_i}{\sqrt{v(x_i)}}.$$

Si definimos,

$$V := \begin{pmatrix} v(x_1) & & \\ & \ddots & \\ & & v(x_n) \end{pmatrix},$$

El estimador resultante puede escribirse como

$$\hat{\beta}_{BLUE} := (X'V^{-1}X)^{-1}X'V^{-1}y.$$

Este estimador es el mejor estimador lineal e insesgado (Teorema de Aitken) y es consistente. En particular, es más eficiente que el estimador de MCO. Lo que destacan los autores del principal paper para mi tesis, es que este estimador es "utópico" dado que asume que la función cedástica es conocida (Romano & Wolf, 2017).

Un enfoque factible sería estimar la función cedástica $v(\cdot)$ utilizando los datos y luego aplicar OLS al modelo:

$$\frac{y_i}{\sqrt{\hat{v}(x_i)}} = \frac{x_i'}{\sqrt{\hat{v}(x_i)}}\beta + \frac{\varepsilon_i}{\sqrt{\hat{v}(x_i)}}.$$

En donde $\hat{v}(\cdot)$ es un estimador de la función $v(\cdot)$. El estimador resultante es lo que se llama Mínimos Cuadrados Ponderados (WLS). Si definimos:

$$\hat{V} \coloneqq \begin{pmatrix} \hat{v}(x_1) & & \\ & \ddots & \\ & & \hat{v}(x_n) \end{pmatrix},$$

El estimador WLS pude escribirse como

$$\hat{\beta}_{WLS} \coloneqq \left(X'\hat{V}^{-1}X\right)^{-1}X'\hat{V}^{-1}y.$$

Este estimador no es necesariamente insesgado. Si $\hat{v}(\cdot)$ es un estimador consistente de la función cedástica, entonces el estimador WLS es asintóticamente más eficiente que OLS. Una de las contribuciones de este *paper*, que justifica la creación de *Adaptative Least Square* (ALS), es que por más que $\hat{v}(\cdot)$ sea un estimador **insconsistente** de $v(\cdot)$, WLS puede generar considerables ganancias de eficiencia si se compara con OLS en presencia de heterocedasticidad condicional.

Adaptative Least Squares

En esta sección del trabajo lo que haré es detallar la matemática del estimador propuesto por Romano y Wolf (2017), *Adaptative Least Squares*. Sabemos que bajo homocedasticidad condicional, OLS es más eficiente que WLS en muestras finitas. Pero, se puede mostrar que bajo ciertos supuestos sobre cómo estimar la función cedástica, OLS y WLS son asintóticamente equivalentes. Es también cierto que bajo heterocedasticidad condicional, WLS es más eficiente, tanto es muestras finitas en un "First-order asymtotic sense"[1].

[1] Profundizaré este concepto cuando me ponga a hacer la tesis.

Por el párrafo anterior es justamente la motivación para la derivación del estimador. Los autores consideran que es tentador que los datos sean los que digan cuál ruta tomar: OLS o WLS. Existen varios tests para detectar heterocedasticidad condicional, tales como el de White (1982), Breush-Pagan (1979), etc. Bajo la hipótesis nula de estos tests, el modelo es homocedástico condicional en las variables. Bajo la alternativa, es heterocedástico. Entonces, en caso de no rechazar, habría que utilizar OLS, mientras que si se rechaza, habría que usar WLS.

La motivación es que, bajo homocedasticidad condicional, el estimador ALS va a ser igual al estimador de WLS con una baja probabilidad (esto depende del test de homocedasticidad condicional a utilizar). En este caso, ALS sería más eficiente que WLS en muestras finitas, aunque menos eficiente que OLS. En cambio, bajo heterocedasticidad condicional, el estimador ALS va a ser igual a WLS con una probabilidad cerca a uno (siempre depende del test). Por lo tanto, para muestras lo suficientemente grandes, ALS debería ser tan eficiente como WLS, mientras que para muestras chicas, cuando la potencia del test de homocedasticidad no es cercana a uno, la eficiencia se esperaría que esté entre la de OLS y la de WLS. En consecuencia, ALS sacrifica ganancias de eficiencia por sobre WLS bajo heterocedasticidad condicional en favor de estar más cerca de la performance de OLS bajo homocedasticidad condicional.

Estimación de la función cedástica

Para poder utilizar WLS, como dijimos anteriormente, es necesario estimar la función cedástica. Los autores proponen utilizar un modelo paramétrico $v_\theta(\cdot)$, donde $\theta \in \mathbb{R}^d$ es un parámetro de dimensión finita. El modelo puede provenir de la teoría económica, del gráfico de los residuos, o por decisión del investigador. Lo que se asume es que para cada $\sigma^2 > 0$ existe un único $\theta := \theta(\sigma^2)$ tal que $v_\theta(x) \equiv \sigma^2$. Wolf y Romano sugieren un modelo flexible:

$$v_\theta(x_i) := \exp\left(\nu + \gamma_1 \log\left|x_{i,1}\right| + \gamma_2 \log\left|x_{i,2}\right| + \cdots + \gamma_K \log\left|x_{i,K}\right|\right),$$

con $\theta := (\nu, \gamma_1, \gamma_2, \ldots, \gamma_K)'$. Puede asumirse que $x_{i,1} \equiv 1$, es decir, que el modelo de regresión original tiene una constante (esto es lo que se supondrá de acá en adelante). Otro modelo que se puede plantear es:

$$v_\theta(x_i) := \nu + \gamma_2\left|x_{i,2}\right| + \cdots + \gamma_K\left|x_{i,K}\right|,$$

con $\theta := (\nu, \gamma_2, \ldots, \gamma_K)'$. La ventaja de la primera especificación en relación a la segunda es que la primera garantiza que las varianzas son no negativas.

Luego de elegir la especificación, hay que hacer el test de homocedasticidad condicional. Los autores recomiendan no utilizar tests que sean demasiado generales como del de White (1980), a no ser que la especificación de la $v_\theta(\cdot)$ sea la misma del test. En la primera etapa de los tests, hay que estimar θ a través de OLS, computar el coeficiente de determinación multiplicado por la cantidad de observaciones y compararlo con el cuantil de la distribución chi-cuadrado.

Referencias

Romano, J. P., & Wolf, M. (2017). Resurrecting weighted least squares. *Journal of Econometrics*, *197*(1), 1-19.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817-838.