Aspectos teóricos avanzados del modelo de regresión

Gabriel V. Montes-Rojas

Regresores estocásticos

- Hasta ahora asumimos que teníamos **regresores no estocásticos**. Este supuesto simplifica las pruebas de insesgadez y de cálculo de varianza, dado que las Xs se pueden considerar números fijos.
- En la práctica si estamos dispuestos a asumir que tenemos una muestra aleatoria, esta muestra debera contener factores aleatorios en todas las variables, $\{y_i, X_i\}_{i=1}^{N}$.
- Como vamos a ver todo modelo de regresión tiene el objetivo de estimar una **esperanza condicional**, $E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K$, donde $\mathbf{x}$ es el conjunto de todas las $K+1$ variables explicativas, incluyendo una constante.
- "Condicionar" en una variable aleatoria es hacerla "fija". En teoría de la probabilidad se basa en la definición de probabilidad condicional.

Esperanzas condicionales

- Toda variable aleatoria y se puede **descomponer** en dos partes ortogonales entre sí:

  $$y = E(y|\boldsymbol{x}) + u,$$

  donde

  - (i) $E(u|\boldsymbol{x}) = 0$,
  - (ii) $E(h(\boldsymbol{x})u) = 0$ para cualquier función $h(.)$.

  Prueba: (i) Definamos $u \equiv y - E(y|\boldsymbol{x})$. Tomando esperanzas $E(u|\boldsymbol{x}) = E(y|\boldsymbol{x}) - E(y|\boldsymbol{x}) = 0$. (ii) Usando la ley de esperanzas iteradas $E(h(\boldsymbol{x})u) = E(E(h(\boldsymbol{x})u|\boldsymbol{x})) = E(h(\boldsymbol{x})E(u|\boldsymbol{x})) = 0$.

- Un resultado importante es que $E(u|\boldsymbol{x}) = 0$ implica que $E(u) = 0$. Esto es por la propiedad de esperanzas iteradas que dice que $E_u(u) = E_x[E_u(u|\boldsymbol{x})]$, donde la primera esperanza es con respecto a u y la segunda a $\boldsymbol{x}$.
- También, que $E(u|\boldsymbol{x})$ implica que $E(\boldsymbol{x}u) = 0$.
- Entonces, $cov(\boldsymbol{x}, u) \equiv E(\boldsymbol{x}u) - E(\boldsymbol{x})E(u) = 0$.
- Es decir, el supuesto $E(u|\boldsymbol{x}) = 0$ implica que los errores u de un modelo de regresión no están correlacionados con las $\boldsymbol{x}$.

Esperanzas condicionales

- La esperanza condicional es la solución al problema de minimización del valor esperado de las desviaciones al cuadrado, o sea $E(y|\mathbf{x}) = arg\min_{m(\mathbf{x})} E((y - m(\mathbf{x}))^2)$.
  Prueba: $(y - m(\mathbf{x}))^2 = ((y - E(y|\mathbf{x})) + (E(y|\mathbf{x}) - m(\mathbf{x})))^2 =$ $(y - E(y|\mathbf{x}))^2 + (E(y|\mathbf{x}) - m(\mathbf{x}))^2 + (y - E(y|\mathbf{x}))(E(y|\mathbf{x}) - m(\mathbf{x}))$. Notemos que el primer término no depende de $m(\mathbf{x})$, mientras que el tercero se puede escribir como $u(x)(E(y|\mathbf{x}) - m(\mathbf{x})) = u(x)h(x)$. Si tomamos la esperanza condicional del tercero tenemos 0 por (ii).
- Sin embargo, no sabemos la forma funcional de $E(y|\mathbf{x})$.

Esperanzas condicionales

- Para cualquier variable aleatoria y, tenemos la proyección poblacional sobre el espacio generado por las $\boldsymbol{x}$, $r(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}$ donde $\boldsymbol{\beta} = arg\min_{\boldsymbol{b}} E((y - \boldsymbol{x}\boldsymbol{b})^2)$.
- Si la esperanza condicional es lineal, entonces $E(y|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}$.
- Cada vez que corremos una regresión estamos estimando $E(y|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{b}$ asumiendo que es lineal en los parámetros. Conviene entonces decir que estamos estimando una esperanza condicional y levantar el supuesto de regresores no estocásticos.

Teorema de Gauss-Markov

- Supuesto 1: Lineal en parámetros La variable dependiente y se relaciona con X por una función lineal, i.e. $y = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K + u$.
- Supuesto 2: Muestreo aleatorio $\{(y_i, x_{1i}, x_{2i}, ..., x_{ni})\}$, $i = 1, 2, ...n$ es una muestra aleatoria del modelo del Supuesto 1.
- Supuesto 3: Ausencia de colinearidad perfecta en $\mathbf{X}$ Para esto necesitamos que $(\mathbf{X}'\mathbf{X})$ sea no singular o $rango(\mathbf{X}'\mathbf{X})^{-1} = K + 1$. Condición necesaria y suficiente para esto es que no haya una relación exacta entre los regresores.
- Supuesto 4: Media condicional cero $E[u|\mathbf{x}] = 0$

MCO es insesgado $E[\hat{\beta}_j|\mathbf{x}] = \beta_j$, $j = 0, 1, 2, ..., K$ *o* $E[\hat{\beta}|\mathbf{x}] = \beta$ *donde* β *es el vector de todos los parámetros.*
Prueba:...

Teorema de Gauss-Markov

- Supuesto 5: Homocedasticidad $Var[u|\mathbf{x}] = \sigma^2$

Teorema Gauss-Markov: Bajo los Supuestos 1-5, los estimadores MCO $(\hat{\beta}_0, ..., \hat{\beta}_k)$ *son los mejores estimadores lineales de* $\beta_0, ..., \beta_k$. *Note: MEJOR significa mínima varianza.*

Nociones básicas de teoría asintótica

- Consideremos una muestra aletoria de tamaño N de una variable aleatoria $\{x_1, x_2, ..., x_N\}$, $x \sim (\mu, \sigma^2)$. Esto es lo mismo que decir que es una muestra aleatoria de variables iid (independientes e idénticamente distribuídas).
- Ley de los grandes números: $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \xrightarrow{p} E[x] = \mu$ cuando $N \to \infty$. Decimos que $\bar{x}$ es consistente para μ. "$\xrightarrow{p}$" significa **convergencia en probabilidad**. También usamos la notación $\bar{x} = \mu + o_p(1)$, donde $o_p(1)$ significa the es un término que converge en probabilidad a 0.
- Teorema central del límite: $\frac{\sqrt{N}}{N}\sum_{i=1}^{N} x_i \xrightarrow{d} Normal(\mu, \sigma^2)$ cuando $N \to \infty$. "$\xrightarrow{d}$" significa **convergencia en distribución**.
- Decimos que $\bar{x}$ es asintóticamente normal: $\sqrt{N}(\bar{x} - \mu) \xrightarrow{d} Normal(0, \sigma^2)$ cuando $N \to \infty$.

Identificación

- Supuesto **OLS.0** (muestra aleatoria): El modelo poblacional es $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_K x_{Ki} + u_i$, $i = 1, 2, ...$ donde la $\{(y_i, \boldsymbol{x}_i, u_i) : i = 1, 2, ...\}$ es una secuencia de vectores aleatorios iid.
- Supuesto **OLS.1** (condición de ortogonalidad poblacional): $E(\boldsymbol{x}'u) = 0$
  - Dado que $\boldsymbol{x}$ tiene una constante, OLS.1 es equivalente a decir que $E(u) = 0$ y que $Cov(x_j, u) = 0, j = 1, 2, 3, ..., K$.
  - Una condición suficiente para OLS.1 es $E(u|\boldsymbol{x}) = 0$: $E(\boldsymbol{x}'u) = E(\boldsymbol{x}'E(u|\boldsymbol{x})) = 0$.
- Supuesto **OLS.2**: *rango* $E(\boldsymbol{x}'\boldsymbol{x}) = K + 1$
  - Dado que $E(\boldsymbol{x}'\boldsymbol{x})$ es una matrix simétrica $(K+1) \times (K+1)$, OLS.2 es equivalente a decir que $E(\boldsymbol{x}'\boldsymbol{x})$ es positiva definida y que $Cov(x_j, u) = 0, j = 1, 2, 3, ..., K$.

Bajo los supuestos OLS.1 y OLS.2, el vector de parámetros β está **identificado**. *En este contexto significa que se puede expresar como momentos poblacionales de variables observadas.*

$$\beta = [E(\boldsymbol{x}'\boldsymbol{x})]^{-1} E(\boldsymbol{x}'y)$$

Consistencia

Tomando la identificación, podriamos definir el estimador MCO como

$$\hat{\beta} = \left(N^{-1} \sum_{i}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(N^{-1} \sum_{i}^{N} \mathbf{x}_i' y_i \right) = \beta + \left(N^{-1} \sum_{i}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(N^{-1} \sum_{i}^{N} \mathbf{x}_i' u_i \right)$$

$$= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

Bajo los supuestos OLS.1 y OLS.2, el estimador MCO $\hat{\beta}$ de una muestra aleatoria es consistente para β.

Normalidad asintótica

La distribución asintótica depende de

$$\sqrt{N}(\hat{\beta}-\beta)=\left(N^{-1}\sum_{i}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(N^{-1/2}\sum_{i}^{N}\mathbf{x}_i'u_i\right)$$

- Probar que $\left(N^{-1}\sum_i^N \mathbf{x}_i'\mathbf{x}_i\right)^{-1}-\boldsymbol{A}^{-1}=o_p(1)$, donde $\boldsymbol{A}\equiv E(\mathbf{x}'\mathbf{x})$. Definamos $\widehat{\boldsymbol{A}}\equiv N^{-1}\sum_i^N \mathbf{x}_i'\mathbf{x}_i$.
- $\{(\mathbf{x}_i'u_i):i=1,2,\ldots\}$ es una secuendia iid con media cero y varianza finita. Entonces por una aplicación del teorema central del límite

$$N^{-1/2}\sum_{i}^{N}\mathbf{x}_i'u_i\xrightarrow{d}Normal(\mathbf{0},\boldsymbol{B}),$$

donde $\boldsymbol{B}\equiv E(u^2\mathbf{x}'\mathbf{x})$ es una matriz $(K+1)\times(K+1)$.
- Entonces,

$$\sqrt{N}(\hat{\beta}-\beta)=\boldsymbol{A}^{-1}\left(N^{-1/2}\sum_{i}^{N}\mathbf{x}_i'u_i\right)+o_p(1)$$

($o_p(1)$ es que se hace 0 asintóticamente)

Normalidad asintótica

Asumimos homoscedasticidad:
Supuesto **OLS.3**: $Var(\mathbf{x}'u) = E(u^2\mathbf{x}'\mathbf{x}) = \sigma^2\mathbf{A}$.

- Entonces tenemos
$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} Normal(\mathbf{0}, \sigma^2\mathbf{A}^{-1})$$
- Así, $AVar(\hat{\beta}) = \sigma^2\mathbf{A}^{-1}$. Sin embargo todavia tenemos que proponer un estimador de la varianza, $\widehat{AVar(\hat{\beta})} = \hat{\sigma}^2\hat{\mathbf{A}}^{-1}$.
- Lema: $\hat{\sigma}^2((\mathbf{X}'\mathbf{X})/N)^{-1} \xrightarrow{p} \sigma^2\mathbf{A}^{-1}$ donde $\hat{\sigma}^2 \equiv N^{-1}\sum_i^N \hat{u}_i^2$.
Prueba: Tenemos que probar que $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. Para ello consideremos
$\hat{\sigma}^2 \equiv N^{-1}\sum_i^N \hat{u}_i^2 = N^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = N^{-1}(\mathbf{M}_X\mathbf{y})'(\mathbf{M}_X\mathbf{y})$, donde
$\mathbf{M}_X = I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ que es simétrica e idempotente. Ahora, $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$ y $\mathbf{M}_X\mathbf{X} = \mathbf{0}$. Entonces,
$\hat{\sigma}^2 = N^{-1}\mathbf{u}'\mathbf{u} - N^{-1}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = N^{-1}\sum_i^N u_i^2 + o_p(1) = \sigma^2 + o_p(1)$.

Teorema de Frisch-Waugh-Lovell

Consideremos los siguientes modelos de regresión:

$$A.\ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K + u$$

$$B.\ y = \gamma_0 + \gamma_2 x_2 + \gamma_3 x_3 + ... + \gamma_K x_K + v$$

$$C.\ x_1 = \delta_0 + \delta_2 x_2 + \delta_3 x_3 + ... + \delta_K x_K + e$$

- Computar los residuos del modelo B ($\hat{v}$) y C ($\hat{e}$).
- Correr la siguiente regresión auxiliar: $\hat{v} = \alpha_0 + \alpha_1 \hat{e} + residuo$
- Chequear que $\hat{\alpha}_1 = \hat{\beta}_1$
- ¿Cómo se interpreta este resultado?

Teorema de Frisch-Waugh-Lovell

En notación matricial

$$\boldsymbol{y} = \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{u}$$

y consideremos $\hat{\beta} = [\hat{\beta}_1 \;\; \hat{\beta}_2]'$.
Ahora construyamos $\boldsymbol{M}_2\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\beta_1 + \boldsymbol{M}_2\boldsymbol{u}$ donde $\boldsymbol{M}_2$ es la proyección residual de $\boldsymbol{X}_2$. El teorema de FWL muestra que

$$\hat{\beta}_1 = (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y}$$

Prueba: Consideremos $\boldsymbol{y} = \boldsymbol{P}_X\boldsymbol{y} + \boldsymbol{M}_X\boldsymbol{y} = \boldsymbol{X}_1\hat{\beta}_1 + \boldsymbol{X}_2\hat{\beta}_2 + \boldsymbol{M}_X\boldsymbol{y}$. Multipliquemos ambos lados por $\boldsymbol{X}_1'\boldsymbol{M}_2$ y obtenemos $\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y} = \boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\hat{\beta}_1$. (Usamos el resultado $\boldsymbol{M}_2\boldsymbol{X}_2 = \boldsymbol{0}$ y $\boldsymbol{M}_X\boldsymbol{M}_2\boldsymbol{X}_1 = \boldsymbol{0}$). Resolver por $\hat{\beta}_1$,

$$\hat{\beta}_1 = (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y}$$

Algebra de MCO (nota para la slide anterior)

Proyección ortogonal: *Definamos la matriz* $\boldsymbol{P_X} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ *como la matriz que proyecta y en el espacio generado por* $\boldsymbol{X}$. *Tenemos que* $\hat{\boldsymbol{y}} = \boldsymbol{P_X y}$, *los valores predecidos.*

Proyección residual: *Definamos la matriz* $\boldsymbol{M_X} \equiv \boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ *como la proyección de y en el complemento del espacio de* $\boldsymbol{X}$. *Notemos que* $\hat{\boldsymbol{u}} = \boldsymbol{M_X y}$, *residuos de la regresión.*

Notar que $\boldsymbol{P_X M_X} = \boldsymbol{0}$, es decir son ortogonales. Además ambas matrices son idempotentes: $\boldsymbol{P_X P_X} = \boldsymbol{P_X}$, $\boldsymbol{M_X M_X} = \boldsymbol{M_X}$.

El problema con la multicolinearidad

- La multicolinearidad perfecta invalida el modelo de regresión, ej., $X'X$ no es invertible.
- La multicolinearidad no perfecta no es necesariamente un problema. La mayoría de las variables de control están correlacionadas entre sí (ej., educación y experiencia), y eso es justamente lo que hace al modelo de regresión múltiple.
- El problema es cuando la correlación es muy alta (ej., dos índices de precios del mismo país). El resultado es que la significatividad individual es baja (test t), pero la significatividad global del modelo (test F, R^2) es alta. Eso se puede ver con el Teorema de FWL.