

# Notas de Econometría

Pavel Solís  
2025

## 3 Análisis de Regresión Múltiple: Estimación

Sabemos cómo explicar una variable dependiente  $y$  en función de una variable independiente  $x$  con modelo RLS

- Pero es difícil hacer conclusiones ceteris paribus de cómo  $x$  afecta a  $y$
- Supuesto principal (RLS.4) es poco realista
  - RLS.4: Todos los otros factores que afectan a  $y$  no se correlacionan con  $x$

El modelo de regresión lineal múltiple (RLM) nos permite hacer análisis ceteris paribus porque podemos controlar explícitamente por otros factores que afectan simultáneamente a la variable dependiente

RLM permite usar datos no experimentales para:

- Verificar teorías económicas
- Evaluar políticas públicas

Modelo RLM permite incluir muchas variables explicativas que pueden estar correlacionadas

- Podemos aspirar a inferir causalidad en casos en que una RLS no es apropiada
- Permite explicar más de la variabilidad en  $y$ 
  - Mejores modelos para predecir  $y$
- Permite considerar formas funcionales más generales (flexibilidad)
- Vehículo más usado para el análisis empírico en ciencias sociales

¿Qué sigue?

- Estimación y propiedades estadísticas de MCO (falta de sesgo, eficiencia) para RLM

### 3.1 Motivación y definición

RLM se puede usar para resolver problemas que no se pueden resolver con RLS

- Factores diferentes a la variable de interés se pueden incluir en el modelo de regresión
- RLM permite generalizar las relaciones funcionales entre las variables

### 3.1.1 Modelo con 2 variables independientes

$y$  es determinada por 2 variables independientes y por otros factores no observados contenidos en  $u$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- $\beta_0$  es el parámetro del intercepto o la constante
- $\beta_1$  y  $\beta_2$  son los parámetros de las pendientes
  - Aún estamos interesados en  $\beta_1$
- Sacamos a  $x_2$  de  $u$  y la ponemos explícitamente en la ecuación
  - $\beta_1$  mide el efecto de  $x_1$  en  $y$ , manteniendo  $x_2$  fija (*ceteris paribus*)
  - $\beta_2$  mide el efecto *ceteris paribus* de  $x_2$  en  $y$
- Antes,  $x_2$  no debía estar correlacionada con  $x_1$  (o  $\beta_1$  estaría sesgado)

Ejemplos.

$$\text{salario} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + u$$

$$\text{consumo} = \beta_0 + \beta_1 \text{ingreso} + \beta_2 \text{ingreso}^2 + u$$

- Ambos ejemplos se pueden expresar en la forma general para 2 variables independientes
  - Pero la interpretación cambia
- El segundo modelo tiene una sola variable independiente pero aparece de dos formas distintas (3 parámetros)

$$\frac{\Delta \text{consumo}}{\Delta \text{ingreso}} \approx \beta_1 + 2\beta_2 \text{ingreso}$$

- Efecto marginal del ingreso sobre el consumo
- Depende de  $\beta_1$ ,  $\beta_2$  y el ingreso

Necesitamos supuestos de cómo  $u$  se relaciona con  $x_1$  y  $x_2$

- Supuesto principal en el modelo de 2 variables:  $\mathbb{E}(u \mid x_1, x_2) = 0$ 
  - Para cualesquier valores de  $x_1$  y  $x_2$  en la población, el promedio de los factores no observados es cero

Ejemplos.

$$\mathbb{E}(u \mid \text{educ}, \text{exper}) = 0$$

- Promedio de habilidad (en  $u$ ) es el mismo en todas las combinaciones de educación y experiencia

$$\mathbb{E}(u \mid \text{ingreso}, \text{ingreso}^2) = 0$$

- Es lo mismo que  $\mathbb{E}(u \mid \text{ingreso}) = 0$  porque  $\text{ingreso}^2$  se conoce al conocer

### 3.1.2 Modelo con $k$ variables independientes

RLM permite considerar varios factores que afectan a  $y$

- Ej. Para analizar el salario, podemos considerar además: capacitación, antigüedad, variables demográficas

El modelo de RLM o **modelo de regresión múltiple** se escribe

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

- $y$  es la variable dependiente, explicada, etc.
- $x_1, x_2, \dots, x_k$  son las variables explicativas, independientes, etc.
- Contiene  $k + 1$  parámetros poblacionales no conocidos
  - $\beta_0$  es el intercepto
  - $\beta_1, \beta_2, \dots, \beta_k$  son los parámetros de las pendientes asociadas con las  $x$
- $u$  es el término de error, contiene factores que afectan a  $y$  distintos a  $x_1, x_2, \dots, x_k$ 
  - Siempre habrá factores que no se puedan incluir

La interpretación de los parámetros es importante

Ejemplo.

$$\log(\text{salario}) = \beta_0 + \beta_1 \log(\text{ventas}) + \beta_2 \text{ceoantig} + \beta_3 \text{ceoantig}^2 + u$$

- $k = 3$
- $\beta_1$ : elasticidad del salario con respecto a las ventas ceteris paribus
- Si  $\beta_3 = 0$ ,  $(100 \times \beta_2)$  es aproximadamente el incremento porcentual ceteris paribus en el salario cuando *ceoantig* aumenta 1 año
- Si  $\beta_3 \neq 0$ , el efecto de *ceoantig* en el salario es

$$\frac{\Delta \log(\text{salario})}{\Delta \text{ceoantig}} \approx \beta_2 + 2\beta_3 \text{ceoantig}$$

Recordatorio: Lineal en RLM significa lineal en los parámetros  $\beta_j$

- Muchas aplicaciones de RLM involucran relaciones no lineales entre las variables

Supuesto clave para el modelo RLM:

$$\mathbb{E}(u \mid x_1, x_2, \dots, x_k) = 0$$

- Clave para probar falta de sesgo en MCO

- Todos los factores en  $u$  no están correlacionados con las variables explicativas
- Cualquier problema que cause correlación entre  $u$  y cualquier variable independiente hace que el supuesto no se cumpla
- La **función de regresión poblacional** (FRP) es:

$$\mathbb{E}(y \mid x_1, x_2, \dots, x_k) = \mathbb{E}(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u \mid \vec{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

## 3.2 Mecánica e interpretación de MCO

En RLM:

- Tenemos  $k$  variables independientes
- Buscamos  $k + 1$  estimadores  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  para los parámetros  $\beta_0, \beta_1, \dots, \beta_k$

### 3.2.1 Obtención de los estimadores de MCO

La ecuación de MCO estimada, la línea de regresión de MCO o la **función de regresión muestral** (FRM) es

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k$$

- $\hat{\beta}_0$  es el estimador de MCO para el intercepto
- $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$  son los estimadores de MCO para las pendientes

### Problema de minimización

El método de MCO escoge los estimadores ( $\hat{\beta}$ ) que minimizan la suma de los residuales al cuadrado dada una muestra de  $n$  observaciones  $\{(y_i, x_{i1}, x_{i2}, \dots, x_{ik}) \mid i = 1, 2, \dots, n\}$

$$\min \sum_{i=1}^n \hat{u}_i^2 = \min \sum_{i=1}^n (y_i - \hat{y})^2 = \min \sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik} \right)^2$$

- En  $x_{ij}$ ,  $i$  hace referencia a la observación y  $j$  a la variable independiente
  - Ej.  $educ_i, exper_i$
- La suma es sobre todas las observaciones

Este problema de minimización se resuelve con cálculo de varias variables para obtener las *condiciones de primer orden* de MCO

$$\sum_{i=1}^n \hat{u}_i = \sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik} \right) = 0$$

$$\sum_{i=1}^n x_{ij} \hat{u}_i = \sum_{i=1}^n x_{ij} \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik} \right) = 0 \quad \forall j = 1, 2, \dots, k$$

- Se aplica la derivada parcial con respecto a cada  $\hat{\beta}$ 
  - Se obtienen  $k + 1$  ecuaciones lineales con  $k + 1$  incógnitas  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$

- También se pueden obtener como momentos muestrales de  $\mathbb{E}(u) = 0$  y  $\mathbb{E}(x_j u) = 0$

### Solución

Formas de obtener las  $\hat{\beta}$ 's: a mano, con algebra lineal, con paquete estadístico

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

- Es más importante saber cómo interpretarlas
- Requiere suponer que las ecuaciones se pueden resolver de forma única
- Terminología: Corremos una regresión de MCO (o regresamos o proyectamos)  $y$  sobre  $x_1, x_2, \dots, x_k$ 
  - Indica que el método de MCO se utilizó para obtener la línea de regresión
- Generalmente siempre estimamos el intercepto junto con las pendientes

### 3.2.2 Interpretación de la ecuación de regresión de MCO

La línea de regresión de MCO expresada en cambios es

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \dots + \hat{\beta}_k \Delta x_k$$

- Indica la predicción del cambio en  $y$  dados cambios en las  $x$ 's
- $\hat{\beta}_0$  es la predicción de  $y$  cuando  $x_1 = x_2 = \dots = x_k = 0$ 
  - No siempre tiene sentido pero es necesaria

Si  $\Delta x_2 = \Delta x_3 = \dots = \Delta x_k = 0$ , entonces

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1$$

- $\hat{\beta}_1$  mide el cambio en  $y$  cuando  $x_1$  aumenta en 1 unidad, manteniendo todas las otras variables fijas
- Al incluir  $x_2, x_3, \dots, x_k$  en el modelo, obtenemos un coeficiente para  $x_1$ , con una interpretación *ceteris paribus*
- **Controlamos** por las variables  $x_2, x_3, \dots, x_k$  cuando estimamos el efecto de  $x_1$  en  $y$
- Misma interpretación para las otras  $\hat{\beta}_j$ 's

Todas las  $\hat{\beta}_j$ 's tienen una interpretación de efecto parcial o *ceteris paribus*

- Por eso RLM es tan útil

Ejemplo. Queremos entender las calificaciones universitarias en términos de las calificaciones de preparatoria (en escala de 4 puntos) y el examen de admisión

$$\widehat{caluni} = 1.29 + 0.453calprepa + 0.0094examadmi, \quad n = 141$$

- 1.29 es la predicción de  $caluni$  si  $calprepa = examadmi = 0$

- No es realista
- Relación positiva entre *caluni* y *calprepa*, manteniendo *examadmi* fijo
- Interpretación de 0.453:
  - Un punto adicional en *calprepa* está asociado con 0.453 de punto en *caluni*
  - Es la mejor predicción para el caso de dos alumnos con mismo *examadmi* pero  $calprepaA = calprepaB + 1$ ,  $caluniA = caluniB + 0.453$
- Interpretación de 0.0094:
  - Manteniendo *calprepa* fijo, si *examadmi* aumenta 10 puntos, *caluni* aumenta 0.09 puntos
  - \* No significativo porque para *examadmi* media = 36 y desvest = 24

Ejemplo. Queremos analizar el efecto de la educación en el salario

$$\log(\widehat{\text{salario}}) = 0.284 + 0.092educ + 0.0041exper + 0.022antig, \quad n = 526$$

Interpretación de coeficientes en porcentaje y como ceteris paribus

- Interpretación de 0.092:
  - Dejando *exper* y *antig* fijos, 1 año extra de *educ*, aumenta  $\log(\text{salario})$  en 0.092 y el *salario* en 9.2%
  - Diferencia proporcional en la predicción del salario con misma *exper* y *antig* pero con diferencia en *educ* de 1 año
  - Estimado mantiene 2 factores de productividad fijos
- Para saber si 0.092 es un buen estimado del rendimiento ceteris paribus de 1 año más de *educ*, necesitamos propiedades estadísticas de MCO

### 3.2.3 Significado de “mantener otros factores fijos” en RLM

$\hat{\beta}_j$ 's miden el cambio esperado en la variable dependiente  $\Delta\hat{y}$  manteniendo fijas las otras variables independientes

- Los datos vienen de una muestra aleatoria sin restricciones en los valores que pueden tomar las variables independientes
- Al dar a  $\hat{\beta}_j$  una interpretación de efecto parcial, es como si hubiéramos obtenido una muestra aleatoria con los mismos valores para las otras variables independientes y solo hubiera variación en la variable independiente de interés
  - En cuyo caso, hubiéramos corrido una RLS entre la variable dependiente y la variable independiente de interés

El poder de RLM es que nos permite:

- Hacer en ambientes no experimentales, lo que los científicos pueden hacer en un ambiente controlado de laboratorio (mantener otros factores fijos)
- Imitar esa situación sin restringir los valores de ninguna variable independiente
- Dar una interpretación ceteris paribus aunque los datos no se hayan recolectado de esa forma

### 3.2.4 Cambiar más de una variable independiente al mismo tiempo

Si cambia más de una variable a la vez, podemos utilizar la línea de regresión de MCO expresada en cambios

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \dots + \hat{\beta}_k \Delta x_k$$

Ejemplo. Para estimar el efecto en el salario de una persona que se queda en la empresa 1 año más (más experiencia y antigüedad, misma educación)

$$\Delta \log(\widehat{\text{salario}}) = 0.0041 \Delta \text{exper} + 0.022 \Delta \text{antig} = 0.0041 + 0.022 = 0.0261 \text{ o } 2.67\%$$

### 3.2.5 Valores ajustados y residuales de MCO

Valor ajustado o predicho para la observación  $i$ :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik}$$

- Usamos los valores de las variables independientes para cada observación  $i$  en la línea de regresión de MCO
- En general,  $y_i \neq \hat{y}_i$  porque MCO minimiza el promedio del error de predicción al cuadrado
  - No dice nada del error de predicción individual

Residual para la observación  $i$  (como en RLS):  $\hat{u}_i = y_i - \hat{y}_i \quad \forall i$

- Si  $\hat{u}_i > 0$ , la línea de regresión subestima  $y_i$
- Si  $\hat{u}_i < 0$ , la línea de regresión sobreestima  $y_i$

Propiedades de los valores ajustados y residuales (vienen de las condiciones de primer orden y son extensiones de RLS)

1. El promedio muestral de los residuales es cero

$$\bullet \frac{1}{n} \sum_{i=1}^n \hat{u}_i = 0 \implies \bar{y}_i = \bar{\hat{y}}_i$$

2. La covarianza muestral entre cada variable independiente y los residuales de MCO es cero

$$\bullet \text{Cov}(x_{ij}, \hat{u}_i) = 0 \implies \text{Cov}(\hat{y}_i, \hat{u}_i) = 0$$

3. El punto  $(\bar{y}, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$  siempre está sobre la línea de regresión de MCO

$$\bullet \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \dots + \hat{\beta}_k \bar{x}_k \implies \bar{y}_i = \bar{\hat{y}}_i$$

### 3.2.6 Interpretación de efecto parcial en RLM

Para algunas derivaciones, necesitamos fórmulas explícitas para las  $\hat{\beta}_j$ 's

- También ayudan a entender el funcionamiento de MCO

#### Modelo con 2 variables independientes

Cuando  $k = 2$ ,  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$  con interés en  $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1} y_i}{\sum_{i=1}^n \hat{r}_{i1}^2}$$

- $\hat{r}_{i1}$  son los residuales MCO de una RLS de  $x_1$  sobre  $x_2$  ( $y$  no se ocupa)
- Después, corremos una RLS de  $y$  sobre  $\hat{r}_{i1}$ 
  - $\hat{\beta}_1$  aquí es igual que en RLS porque  $\hat{r}_{i1} = 0$
  - \* Revisa regresión por el origen en RLS

Permite interpretar el efecto parcial de  $\hat{\beta}_1$  de otra forma

- Los residuales  $\hat{r}_{i1}$  son la parte de  $x_{i1}$  que no está correlacionada con  $x_{i2}$ 
  - $\hat{r}_{i1}$  es  $x_{i1}$  después de depurar los efectos de  $x_{i2}$
- Entonces,  $\hat{\beta}_1$  mide la relación en la muestra entre  $y$  y  $x_1$  después de depurar  $x_2$ 
  - $\hat{\beta}_1$  mide el cambio en  $y$  cuando  $x_1$  aumenta en 1 unidad manteniendo  $x_2$  fija
- Esto no existe en RLS porque no se incluyen otras variables independientes en el modelo

#### Modelo con $k$ variables independientes

Cuando hay  $k$  variables independientes,  $\hat{r}_{i1}$  viene de la regresión de  $x_1$  sobre  $x_2, x_3, \dots, x_k$  y  $\hat{\beta}_1$  se escribe igual que para el caso  $k = 2$

- $\hat{\beta}_1$  mide el efecto de  $x_1$  sobre  $y$  después de depurar los efectos de  $x_2, x_3, \dots, x_k$

### 3.2.7 Comparación de los estimados de regresión simple y múltiple

¿Cuándo se obtiene el mismo estimado para el efecto de  $x_1$  sobre  $y$  con RLS y RLM?

#### Modelo con 2 variables independientes

- RLS:  $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$
- RLM:  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

La relación entre  $\tilde{\beta}_1$  y  $\hat{\beta}_1$  está dada por:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$$

- $\tilde{\beta}_1$  no nos da el efecto parcial de  $x_1$  sobre  $\hat{y}$
- $\tilde{\delta}_1$  es el coeficiente de la pendiente en la RLS de  $x_{i2}$  sobre  $x_{i1}$ ,  $i = 1, 2, \dots, n$



- $x_2 = \delta_0 + \delta_1 x_1 + \varepsilon$ , de donde se obtienen  $\tilde{\delta}_0$  y  $\tilde{\delta}_1$
- $\tilde{\delta}_1 = \text{Cov}(x_1, x_2) / \text{Var}(x_1)$
- $\hat{\beta}_2 \tilde{\delta}_1$  se conoce como el término distorsionador
  - Efecto parcial de  $x_2$  sobre  $\hat{y}$  por la pendiente de la RLS de  $x_2$  sobre  $x_1$

Entonces,  $\tilde{\beta}_1 = \hat{\beta}_1$  en dos casos:

1. Cuando  $\hat{\beta}_2 = 0$ : Efecto parcial de  $x_2$  sobre  $\hat{y}$  es cero en la muestra
2. Cuando  $\tilde{\delta}_1 = 0$ :  $x_1$  y  $x_2$  no están correlacionadas en la muestra

En general,  $\tilde{\beta}_1 \neq \hat{\beta}_1$ , pero la relación nos sirve para caracterizar cuando son similares (o muy diferentes)

- Si  $\hat{\beta}_2 \approx 0$  o si  $\tilde{\delta}_1 \approx 0$ , esperaríamos que  $\tilde{\beta}_1 \approx \hat{\beta}_1$

### Modelo con $k$ variables independientes

Los estimados de  $\beta_1$  son idénticos en el caso de  $k$  variables independientes si [poco probable]

- Los coeficientes de MCO para  $x_2, \dots, x_k$  son todos cero

$$\hat{\beta}_2 = \dots = \hat{\beta}_k = 0$$

- $x_1$  no está correlacionada con ninguna de las variables  $x_2, \dots, x_k$

$$\tilde{\delta}_2 = \dots = \tilde{\delta}_k = 0$$

Pero  $\tilde{\beta}_1$  (de RLS) y  $\hat{\beta}_1$  (de RLM) pueden ser similares si

- Los coeficientes para  $x_2, \dots, x_k$  son pequeños
- Las correlaciones entre  $x_1$  y las otras variables independientes son pequeñas

### 3.2.8 Bondad de ajuste

Como en RLS, definimos la variación de cada una de las partes:

- Suma de cuadrados total:  $\text{SCT} = \sum_{i=1}^n (y_i - \bar{y})^2$ 
  - Variación de  $y_i$  (suponemos que no es constante,  $y_i \neq c \quad \forall i$ )
- Suma de cuadrados explicada:  $\text{SCE} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 
  - Variación de  $\hat{y}_i$
- Suma de cuadrados de los residuales:  $\text{SCR} = \sum_{i=1}^n \hat{u}_i^2$ 
  - Variación de  $\hat{u}_i$

La variación total en  $y$  siempre se puede expresar como:  $\text{SCT} = \text{SCE} + \text{SCR}$

Si  $\text{SCT} \neq 0$ ,

$$1 = \frac{\text{SCE}}{\text{SCT}} + \frac{\text{SCR}}{\text{SCT}} \implies R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}}$$

- Interpretación de  $R^2$ : Proporción de la variación muestral en  $y_i$  que es explicada por la línea de regresión de MCO
- $0 \leq R^2 \leq 1$  por definición
- $R^2$  es igual al coeficiente de correlación entre  $y_i$  y  $\hat{y}_i$  elevado al cuadrado

$$R^2 = [\text{Corr}(y_i, \hat{y}_i)]^2 = \frac{[\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{[\sum (y_i - \bar{y})^2][\sum (\hat{y}_i - \bar{\hat{y}})^2]}$$

- $R^2$  nunca decrece (generalmente aumenta) cuando se agrega cualquier variable independiente a la regresión (SCR nunca aumenta)
  - Mala herramienta para decidir si agregar variables al modelo
  - Una variable independiente va en el modelo si tiene un efecto parcial diferente de cero sobre  $y$  en la población
- $R^2$  además de medida de bondad de ajuste sirve para probar si un grupo de variables es importante para explicar  $y$

Ejemplo.

$$\widehat{caluni} = 1.29 + 0.453calprepa + 0.0094examadmi, \quad n = 141, \quad R^2 = 0.176$$

- Significa que *calprepa* y *examadmi* explican alrededor de 17.6% de la variación en *caluni* en esta muestra
- Hay muchos otros factores (personalidad, calidad de educación en prepa, etc.) que contribuyen a *caluni*

Una  $R^2$  baja:

- No significa que la ecuación no sirva
  - $\hat{\beta}_j$ 's pueden ser estimados confiables de los efectos ceteris paribus de cada variable independiente sobre  $y$ , eso no depende del valor de  $R^2$
- Significa que es difícil predecir valores individuales de  $y$  con exactitud
  - En ciencias sociales, es difícil predecir comportamientos individuales

### 3.2.9 Regresión por el origen

Teoría económica o sentido común pueden sugerir que  $\beta_0 = 0$ ,

$$\tilde{y} = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \dots + \tilde{\beta}_k x_k$$

- Si  $x_2 = x_3 = \dots = x_k = 0$ , la predicción para  $\tilde{y}$  es 0
- $\tilde{\beta}_j$ 's son los estimados de MCO de la regresión de  $y$  sobre  $x_1, x_2, \dots, x_k$  por el origen
- $\tilde{\beta}_j$ 's minimizan SCR pero fijando  $\beta_0 = 0$

Las propiedades de MCO anteriores no se cumplen en regresión por el origen

1. Los residuales de MCO ya no generan un promedio muestral cero

- $\frac{1}{n} \sum_{i=1}^n \hat{u}_i \neq 0, \bar{y}_i \neq \tilde{y}_i$

2.  $R^2 < 0$  es posible

- Significado:  $\bar{y}$  explica más de la variación en  $y$  que las variables explicativas
- Alternativa 1: Incluir un intercepto
- Alternativa 2: Concluir que las  $x$ 's explican mal a  $y$

3. Si se utiliza  $R^2$  como  $[\text{Corr}(y_i, \tilde{y}_i)]^2$ ,  $R^2 \geq 0$

- No hay regla sobre cómo reportar  $R^2$  en regresión por el origen

Si  $\beta_0 \neq 0$  en la población, las  $\tilde{\beta}_j$ 's estarán sesgadas

- Costo de estimar  $\beta_0$  cuando  $\beta_0 = 0$ : Varianzas de las  $\tilde{\beta}_j$ 's serán más grandes

### 3.3 Valor esperado de los estimadores de MCO

Ahora estudiamos las propiedades estadísticas de los estimadores de MCO

- No se refiere al caso de una muestra particular
- Para parámetros del modelo poblacional
- Supuesto: Podemos tomar muestras aleatorias de forma repetida

Las propiedades de interés son:

- Valor esperado
  - 4 supuestos (extensiones de RLS) permiten mostrar falta de sesgo
- Varianza
  - 1 supuesto adicional permite determinar eficiencia

Partimos de supuestos ideales para obtener propiedades deseables

#### 3.3.1 Falta de sesgo de MCO

Supuesto RLM.1. El modelo es lineal en parámetros (define el modelo)

El modelo poblacional se puede escribir como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

- $\beta_0, \beta_1, \dots, \beta_k$  son parámetros (constantes) de interés no conocidos
- $u$  es un error aleatorio no observado

Observaciones sobre el supuesto RLM.1

- Se conoce como el modelo poblacional o modelo verdadero
  - Incertidumbre: Podemos estimar un modelo diferente
- Clave: Modelo es lineal en los parámetros
  - Flexible porque permite capturar relaciones no lineales
  - $x$  y  $y$  pueden ser funciones de variables de interés

Supuesto RLM.2. La muestra es aleatoria

Tenemos una muestra aleatoria de  $n$  observaciones  $\{(y_i, x_{i1}, x_{i2}, \dots, x_{ik}) \mid i = 1, 2, \dots, n\}$  del modelo poblacional en RLM.1

Observaciones sobre el supuesto RLM.2

- Para una observación particular sacada de la población:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i \quad \forall i$$

- Primer subíndice indica el número de la observación y el segundo subíndice denota el número de la variable

Ejemplo.

$$\log(\text{salar}_i) = \beta_0 + \beta_1 \log(\text{ventas}_i) + \beta_2 \text{ceoantig}_i + \beta_3 \text{ceoantig}_i^2 + u_i$$

- $u_i$  contiene factores no observados para el CEO  $i$  que afectan su salario

- En aplicaciones, el modelo se escribe generalmente en forma poblacional
  - Enfatiza que queremos estimar una relación poblacional
- Aquí  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  son estimadores de  $\beta_0, \beta_1, \dots, \beta_k$ 
  - Antes, para una muestra particular  $\hat{\beta}_j$ 's eran escogidos tal que  $\sum_{i=1}^n \hat{u}_i = 0$  y  $\sum_{i=1}^n x_{ij} \hat{u}_i = 0 \quad \forall j \quad [\text{Cov}(x_{ij}, \hat{u}_i) = 0 \quad \forall j]$
  - Pero no incluimos condiciones para que los estimadores de MCO estuvieran bien definidos para una muestra

Supuesto RLM.3. No colinealidad perfecta (para poder calcular MCO)

En la muestra (y por tanto en la población),

- Ninguna de las variables independientes es constante, y
- No hay relaciones lineales exactas entre las variables independientes

Observaciones sobre el supuesto RLM.3

- Es más complicado que en RLS porque debemos evaluar la relación entre todas las variables independientes
- Si alguna variable independiente es una combinación lineal exacta de las otras variables independientes, el modelo sufre de **colinealidad perfecta**
  - En ese caso, no se puede estimar por MCO porque  $(X'X)^{-1}$  no existe
- RLM.3 permite que las variables independientes estén correlacionadas, solo no pueden estar perfectamente correlacionadas
  - Se puede verificar revisando que  $\text{Corr}(x_s, x_j) \neq 1 \quad \forall s, j$
  - Si no se permitiera ninguna correlación, RLM sería poco útil
  - De hecho, la razón de incluir una variable independiente es que puede estar correlacionada con la de interés y queremos fijar su efecto
- RLM.3 solo descarta correlación perfecta
  - Caso poco probable con una muestra aleatoria

Ejemplo. RLM3 no se cumple si una variable es múltiplo de otra

$$y = \text{consumo}, x_1 = \text{ingreso en pesos}, x_2 = \text{ingreso en miles de pesos}$$

- En ese caso, no se puede realizar un análisis ceteris paribus
- Funciones no lineales de la misma variable se permiten

$$x_1 = \text{ingreso}, x_2 = \text{ingreso}^2$$

Preguntas. ¿Cuál es el problema de estimar la siguiente ecuación?

$$\log(\text{consumo}) = \beta_0 + \beta_1 \log(\text{ingreso}) + \beta_2 \log(\text{ingreso}^2) + u$$

- ¿Qué solución se podría implementar?

Preguntas. En una campaña electoral con 2 candidatos (A y B), ¿cuál sería el problema de estimar la siguiente ecuación?

$$\text{votoA} = \beta_0 + \beta_1 \text{gastoA} + \beta_2 \text{gastoB} + \beta_3 \text{gastototal} + u$$

- ¿Tiene sentido un análisis ceteris paribus?
- ¿Qué solución se podría implementar?

Supuesto RLM.3 puede no cumplirse si

- No somos cuidadosos al especificar el modelo
- Tamaño de muestra es pequeño respecto al número de parámetros a estimar ( $n < k + 1$ )

- Necesitamos al menos  $k + 1$  observaciones para estimar  $k + 1$  parámetros
- Entre más observaciones, mejor

Si el modelo se especifica correctamente y  $n \geq k + 1$ , RLM.3 puede fallar por coincidencia al recolectar los datos

- Caso poco probable, a menos que la muestra sea muy pequeña

Supuesto RLM.4. Media condicional cero (el más importante para probar falta de sesgo)

$$\mathbb{E}(u|x_1, x_2, \dots, x_k) = 0$$

El error  $u$  tiene un valor esperado de cero dado cualquier valor de las variables independientes

Observaciones sobre el supuesto RLM.4

- RLM.4 no se cumple y los estimadores  $\hat{\beta}_j$  estarán sesgados cuando
  - Hay una mala especificación de la forma funcional
    - \* Ej. No incluir  $\text{ingreso}^2$  en  $\text{consumo} = \beta_0 + \beta_1 \text{ingreso} + \beta_2 \text{ingreso}^2 + u$
    - \* Ej. Usar  $\text{salarario}$  cuando el modelo verdadero es  $\log(\text{salarario})$
  - Se omite un factor importante que está correlacionado con  $x_1, x_2, \dots, x_k$ 
    - \* Variables omitidas son menos problema en RLM que en RLS
    - \* Pero puede haber limitación en datos o desconocimiento
  - Hay otras formas en que  $\text{Corr}(u, x_j) \neq 0$ 
    - \* Error de medición en las  $x$ 's
    - \*  $y$  y  $x$ 's determinadas conjuntamente (ej. precios y cantidades)
- Si se cumple RLM.4, decimos que tenemos variables independientes **exógenas**
- Si  $x_j$  se correlaciona con  $u$ , se dice que  $x_j$  es una variable independiente **endógena**
- Diferencia entre RLM.3 y RLM.4:
  - RLM.3 descarta ciertas relaciones entre las  $x$ 's y no tiene nada que ver con  $u$ 
    - \* Se puede probar
  - RLM.4 restringe relación entre las  $x$ 's y  $u$ 
    - \* Nunca sabremos, pero es un supuesto crítico

Teorema. Falta de Sesgo en MCO  
Bajo supuestos RLM.1 a RLM.4,

$$\mathbb{E}(\hat{\beta}_j) = \beta_j, \quad j = 0, 1, 2, \dots, k,$$

para cualquier valor de  $\beta_j$

MCO genera estimadores insesgados para los parámetros poblacionales

- Propiedad de los estimadores bajo muestras aleatorias repetidas

### 3.3.2 Inclusión de variables irrelevantes

**Sobre-especificación** del modelo:

- Incluir una o más variables independientes en el modelo que no tienen efecto parcial en la población [ $\beta_s = 0$ ]

Supongamos que el siguiente modelo satisface RLM.1 a RLM.4 pero  $\beta_3 = 0$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- $x_3$  puede estar correlacionada con  $x_1$  o  $x_2$ 
  - Pero una vez que se controla por  $x_1$  y  $x_2$ ,  $x_3$  no afecta a  $y$
  - Usando esperanza condicional:

$$\mathbb{E}(y \mid x_1, x_2, x_3) = \mathbb{E}(y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- No sabemos que  $\beta_3 = 0$  por lo que estimamos

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

- ¿Qué efecto tiene incluir una variable (irrelevante)  $x_3$  cuando  $\beta_3 = 0$ ?
  - No tiene efecto en términos de falta de sesgo para  $\hat{\beta}_1$  y  $\hat{\beta}_2$ 
    - \* Teorema aplica para cualquier valor de  $\beta_j$
    - \* Entonces  $\mathbb{E}(\hat{\beta}_0) = \beta_0$ ,  $\mathbb{E}(\hat{\beta}_1) = \beta_1$ ,  $\mathbb{E}(\hat{\beta}_2) = \beta_2$ ,  $\mathbb{E}(\hat{\beta}_3) = 0$
  - El efecto está en las varianzas de  $\hat{\beta}_1$  y  $\hat{\beta}_2$

### 3.3.3 Sesgo por variable omitida

**Sub-especificación** del modelo:

- Excluir una variable independiente relevante
  - La variable pertenece al modelo poblacional
- Este problema generalmente causa sesgo en los estimadores de MCO
  - Podemos derivar la dirección y el tamaño del sesgo
- Se conoce como **análisis de mala especificación**

### Modelo con 2 variables independientes

Empezamos con el modelo verdadero y suponemos que satisface los supuestos RLM.1 a RLM.4

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- Ej.  $y = \text{salario}$  (por hora),  $x_1 = \text{educ}$ ,  $x_2 = \text{habilidad}$  (innata)
- Estamos interesados en el efecto parcial de  $x_1$  sobre  $y$

- Para tener  $\mathbb{E}(\hat{\beta}_j) = \beta_j$ , debemos correr una regresión de  $y$  sobre  $x_1$  y  $x_2$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

- Pero por desconocimiento o falta de datos, estimamos el modelo sin incluir  $x_2$

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

- La tilde enfatiza que  $\tilde{\beta}_1$  viene de un modelo sub-especificado

Ejemplo. Modelo verdadero:

$$salario = \beta_0 + \beta_1 educ + \beta_2 habilidad + u$$

- Estimamos

$$salario = \beta_0 + \beta_1 educ + v$$

- $v = \beta_2 habilidad + u$  porque habilidad no se observa
- Obtenemos  $\tilde{\beta}_1$

Sabemos que

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$$

- $\hat{\beta}_1$  y  $\hat{\beta}_2$  son estimadores (imaginarios) de la regresión de  $y_i$  sobre  $x_{i1}$  y  $x_{i2} \forall i$
- $\tilde{\delta}_1$  es la pendiente en la RLS de  $x_{i2}$  sobre  $x_{i1} \forall i$ 
  - Solo depende de las variables independientes

Como el modelo satisface RLM.1 a RLM.4,  $\mathbb{E}(\hat{\beta}_1) = \beta_1$  y  $\mathbb{E}(\hat{\beta}_2) = \beta_2$ , entonces

$$\mathbb{E}(\tilde{\beta}_1) = \mathbb{E}(\hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1) = \mathbb{E}(\hat{\beta}_1) + \mathbb{E}(\hat{\beta}_2) \tilde{\delta}_1 = \beta_1 + \beta_2 \tilde{\delta}_1$$

$$\text{Sesgo}(\tilde{\beta}_1) = \mathbb{E}(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$$

- Se conoce como el **sesgo por variable omitida** (por excluir  $x_2$ )

Dos casos en que  $\tilde{\beta}_1$  es insesgada ( $\mathbb{E}(\tilde{\beta}_1) = \beta_1$ )

- $\beta_2 = 0$ :  $x_2$  no aparece el modelo verdadero
- $\tilde{\delta}_1 = 0$ :  $x_1$  y  $x_2$  no están correlacionadas (porque  $\tilde{\delta}_1 = \text{Cov}(x_1, x_2) / \text{Var}(x_1)$ )

Si  $x_1$  y  $x_2$  están correlacionadas,  $\tilde{\delta}_1$  adopta el signo de  $\text{Corr}(x_1, x_2)$

- $\tilde{\delta}_1 > 0$  si  $\text{Corr}(x_1, x_2) > 0$
- $\tilde{\delta}_1 < 0$  si  $\text{Corr}(x_1, x_2) < 0$

El *signo* o *dirección* del sesgo depende de los signos de  $\beta_2$  y  $\tilde{\delta}_1$



	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Sesgo+	Sesgo-
$\beta_2 < 0$	Sesgo-	Sesgo+

El *tamaño* del sesgo está determinado por el tamaño de  $\beta_2$  y  $\tilde{\delta}_1$

- Ej. 0.14% vs 3% relativo a 8.6%

En la práctica,  $\beta_2$  es un parámetro poblacional no conocido pero generalmente

- Tenemos idea de la dirección del efecto parcial de  $x_2$
- Podemos conjeturar el signo de la correlación entre  $x_1$  y  $x_2$

Ejemplo. A mayor habilidad, mayor productividad y mayor salario, entonces  $\beta_2 > 0$ . Si  $\text{Corr}(\text{educ}, \text{habilidad}) > 0$ ,  $\tilde{\beta}_1$  de la siguiente regresión en promedio será grande

$$\text{salarario} = \beta_0 + \beta_1 \text{educ} + v$$

Si  $\log(\text{salarario}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{habilidad} + u$  satisface RLM.1 a RLM.4 pero estimamos

$$\log(\widetilde{\text{salarario}}) = 0.584 + 0.083 \text{educ}, \quad n = 526, \quad R^2 = 0.186$$

- No sabemos si  $0.083 > \beta_1$  (porque  $\beta_1$  puede ser mayor o menor a 0.083)
- Pero en promedio  $\mathbb{E}(\tilde{\beta}_1) > \beta_1$

Preguntas. Supongamos que en escuelas primarias

$$\text{califprom} = \beta_0 + \beta_1 \text{gasto} + \beta_2 \text{tpobreza} + u$$

pero estimamos  $\beta_1$  con una regresión de *califprom* sobre *gasto*

- ¿Cuál podría ser el signo para  $\beta_2$  y para  $\text{Corr}(x_1, x_2)$ ?
- ¿Cuál es el posible sesgo en  $\tilde{\beta}_1$ ?
- ¿Cuál sería la implicación si  $\beta_1 = 0$  y  $\tilde{\beta}_1 > 0$ ?

Terminología al omitir una variable independientes relevante

- Sin importar el signo de  $\beta_1$ ,  $\tilde{\beta}_1$  tiene:
  - **Sesgo hacia arriba** si  $\mathbb{E}(\tilde{\beta}_1) > \beta_1$
  - **Sesgo hacia abajo** si  $\mathbb{E}(\tilde{\beta}_1) < \beta_1$
- $\tilde{\beta}_1$  está **sesgada hacia cero** si  $\mathbb{E}(\tilde{\beta}_1)$  está más cercano a cero que  $\beta_1$ 
  - Si  $\beta_1 > 0$ ,  $\tilde{\beta}_1$  está sesgada hacia cero si tiene sesgo hacia abajo

- Si  $\beta_1 < 0$ ,  $\tilde{\beta}_1$  está sesgada hacia cero si tiene sesgo hacia arriba
- En general, si  $|\mathbb{E}(\tilde{\beta}_1)| < |\beta_1|$

### Modelo con 3 variables independientes

Derivar el signo del sesgo por variable omitida es más complicado cuando hay varios regresores porque las  $x_j$ 's están correlacionadas en pares

- Si  $\text{Corr}(x_s, u) \neq 0$ , generalmente todos los estimadores de MCO  $\hat{\beta}_j$ 's estarán sesgados

Supongamos que el siguiente modelo poblacional satisface RLM.1 a RLM.4

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

pero omitimos  $x_3$  y estimamos

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2$$

Si  $\text{Corr}(x_1, x_3) \neq 0$  y  $\text{Corr}(x_2, x_3) = 0$

- No es correcto decir que  $\tilde{\beta}_1$  está sesgada y  $\tilde{\beta}_2$  no está sesgada
- Generalmente,  $\tilde{\beta}_1$  y  $\tilde{\beta}_2$  estarán sesgadas
  - Excepto si  $\text{Corr}(x_1, x_2) = 0$

Para aproximar la dirección del sesgo en  $\tilde{\beta}_1$ , suponemos que  $x_2$  no está en la población ni estimación ( $\text{Corr}(x_1, x_2) = 0$ ), entonces

$$\mathbb{E}(\tilde{\beta}_1) = \beta_1 + \beta_3 \tilde{\delta}_{31}$$

- Y Sesgo  $(\tilde{\beta}_1) > 0$  si  $\beta_3 > 0$  y  $\text{Corr}(x_1, x_3) > 0$ , etc.

Ejemplo.

$$\text{salario} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{habilidad} + u$$

- Si omitimos *habilidad*, los estimadores de  $\beta_1$  y  $\beta_2$  estarán sesgados
  - Aún si  $\text{Corr}(\text{exper}, \text{habilidad}) = 0$
- Si estamos interesados en el rendimiento de *educ* ( $\beta_1$ ), quisiéramos saber si  $\tilde{\beta}_1$  está sesgada hacia arriba o hacia abajo

Para aproximar, suponemos que  $\text{Corr}(\text{exper}, \text{habilidad}) = 0$  y  $\text{Corr}(\text{educ}, \text{exper}) = 0$

- Como  $\beta_3 > 0$  y  $\text{Corr}(\text{educ}, \text{habilidad}) > 0$ ,  $\tilde{\beta}_1$  tiene sesgo hacia arriba

En la práctica, seguimos un razonamiento similar para tener una idea del sesgo

- Hay que evaluar la relación entre la variable de interés y la variable omitida

### 3.4 Varianza de los estimadores de MCO

De la distribución muestral de las  $\hat{\beta}_j$ 's

- Sabemos las tendencias centrales (supuestos RLM.1 a RLM.4)
- Queremos saber la dispersión para analizar (bajo homocedasticidad)
  - Varianza
  - Eficiencia

Supuesto RLM.5. Homocedasticidad

El error  $u$  tiene la misma varianza dados cualesquier valores de las variables independientes

$$\text{Var}(u \mid x_1, x_2, \dots, x_k) = \sigma^2$$

La varianza es la misma para todas las posibles combinaciones de valores de las variables independientes

Ejemplo. Si el modelo poblacional es

$$\text{salarior} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{antig} + u,$$

entonces

$$\text{Var}(u \mid \text{educ}, \text{exper}, \text{antig}) = \sigma^2$$

no depende de  $\text{educ}$ ,  $\text{exper}$ ,  $\text{antig}$

Observaciones sobre el supuesto RLM.5

- Si RLM.5 no se cumple, hay **heterocedasticidad**
- Los supuestos RLM.1 a RLM.5 se conocen como los supuestos de Gauss-Markov (para corte transversal)
- Si escribimos  $\vec{x} = (x_1, x_2, \dots, x_k)$ , podemos expresar
  - RLM.1 a RLM.4 como

$$\mathbb{E}(y \mid \vec{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

\* Lineal en parámetros, depende de las variables independientes

- RLM.5 como

$$\text{Var}(y \mid \vec{x}) = \sigma^2$$

\* No depende de las variables independientes

Teorema. Varianzas Muestrales de los Estimadores de MCO (para las pendientes)  
Bajo los supuestos RLM.1 a RLM.5, y condicional en los valores muestrales de las

variables independientes,

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SCT}_j(1 - R_j^2)} \quad \forall j = 1, 2, \dots, k$$

- $\text{SCT}_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$  es la variación muestral total en  $x_j$
- $R_j^2$  es la  $R^2$  de la regresión de  $x_j$  sobre las otras variables independientes (con intercepto)
  - Proporción de la variación total en  $x_j$  explicada por las otras variables independientes

El tamaño de  $\text{Var}(\hat{\beta}_j)$  es importante

- Una varianza grande significa un estimador menos preciso e implica
  - Intervalos de confianza más amplios
  - Pruebas de hipótesis menos precisas

### 3.4.1 Componentes de las varianzas de MCO: Multicolinealidad

La varianza de cada  $\hat{\beta}_j$  depende de 3 factores:  $\sigma^2$ ,  $\text{SCT}_j$ ,  $R_j^2$

1. Varianza del error ( $\sigma^2$ ):

- Si sube  $\sigma^2$ , sube  $\text{Var}(\hat{\beta}_j)$  porque hay más ruido en la ecuación
  - Hace más difícil estimar el efecto parcial de cualquier variable independiente sobre  $y$
- $\sigma^2$  es una característica de la población
  - No tiene que ver con el tamaño de la muestra
  - Es desconocida por lo que queremos estimarla
- Dada  $y$ , la única forma en que  $\sigma^2$  puede bajar es agregando más variables independientes ('sacarlas' de  $u$ )
  - No siempre es posible

2. Variación muestral total en  $x_j$  ( $\text{SCT}_j$ ):

- Si sube  $\text{SCT}_j$ , baja  $\text{Var}(\hat{\beta}_j)$ 
  - Ceteris paribus queremos mayor variación en  $x_j$
  - Al aumentar el tamaño de la muestra, aumenta la variación en cada  $x_j$
- Componente de  $\text{Var}(\hat{\beta}_j)$  que depende sistemáticamente de la muestra
- Si  $\text{SCT}_j \approx 0$ , aumenta  $\text{Var}(\hat{\beta}_j)$  pero no viola RLM.3
  - $\text{SCT}_j = 0$  no está permitido por RLM.3
  - $n$  pequeño puede causar varianzas muestrales grandes

3. Relaciones lineales entre las variables independientes ( $R_j^2$ ):

- Si sube  $R_j^2$ , sube  $\text{Var}(\hat{\beta}_j)$
- $R_j^2$  no aparece en RLS porque solo hay una  $x_j$  y  $R_j^2 \neq R^2$ 
  - $R_j^2$  solo involucra variables independientes
- Si  $k = 2$ ,  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$  y  $\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\text{SCT}_1(1-R_1^2)}$  con  $R_1^2$  de la regresión de  $x_1$  sobre  $x_2$  (con intercepto)
  - Como  $R^2$  mide la bondad de ajuste,  $R_1^2 \approx 1$  implica que  $x_2$  explica mucha de la variación en  $x_1$  por lo que  $\text{Corr}(x_1, x_2)$  es alta
  - Conforme  $R_1^2 \rightarrow 1$ , sube  $\text{Var}(\hat{\beta}_1)$ , entonces una alta relación lineal entre  $x_1$  y  $x_2$  implica que  $\text{Var}(\hat{\beta}_1)$  y  $\text{Var}(\hat{\beta}_2)$  son altas
- Dados  $\sigma^2$  y  $\text{SCT}_j$ ,  $\text{Var}(\hat{\beta}_j)$  es más baja cuando  $R_j^2 = 0 \iff \text{Corr}(x_s, x_j) = 0 \quad \forall s$ 
  - Pero  $R_j^2 = 0$  es raro
- Caso  $R_j^2 = 1$  está descartado por RLM.3
  - $x_j$  sería una combinación lineal perfecta de las otras variables independientes
- Más interesante es el caso  $R_j^2 \approx 1$  porque cuando  $R_j^2 \rightarrow 1$ ,  $\text{Var}(\hat{\beta}_j) \rightarrow \infty$ 
  - Pero  $R_j^2 \approx 1$  no viola RLM.3
- **Multicolinealidad (MC)**: Se refiere a una correlación alta (pero no perfecta) entre 2 o más variables independientes
  - Problema de MC no está bien definido porque no viola supuestos
  - Efecto de  $R_j^2 \approx 1$  en  $\text{Var}(\hat{\beta}_j)$  depende de  $\sigma$  y  $\text{SCT}_j$
  - Lo que importa para inferencia estadística es  $\hat{\beta}_j$  relativo a  $\text{desvest}(\hat{\beta}_j)$
- Alta correlación entre variables independientes y  $n$  pequeño, implica que  $\text{Var}(\hat{\beta}_j)$  aumenta
- Ceteris paribus es mejor tener una correlación baja entre  $x_j$  y otras variables independientes
- Quitar variables independientes del modelo podría reducir MC, pero puede generar sesgo si las variables pertenecen al modelo poblacional
  - Ej. En lugar de estimar efectos separados de categorías de gasto altamente correlacionadas, sería mejor juntarlas
- La correlación entre variables independientes puede ser irrelevante para otra variable independiente (de interés)
  - Ej.  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$  con alta  $\text{Corr}(x_2, x_3)$  y altas  $\text{Var}(\hat{\beta}_2)$  y  $\text{Var}(\hat{\beta}_3)$  pero sin efecto en  $\text{Var}(\hat{\beta}_1)$

- \* Si  $\text{Corr}(x_1, x_2) = 0$  y  $\text{Corr}(x_1, x_3) = 0$ , entonces  $R_j^2 = 0$  y  $\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\text{SCT}_j}$  sin importar  $\text{Corr}(x_2, x_3)$
- Esto es importante porque generalmente se incluyen varios controles para aislar efecto causal de una variable
  - \* Alta correlación entre controles no dificulta determinar efectos
- Se pueden utilizar estadísticos para diagnosticar MC
  - Pero no distinguen si involucra controles que no importan
- Son más útiles estadísticos para coeficientes individuales, el más común es el **factor de inflación de la varianza**

$$\text{FIV}_j = \frac{1}{1 - R_j^2} \implies \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\text{SCT}_j} \text{FIV}_j$$

- $\text{FIV}_j$  es el factor por el que  $\text{Var}(\hat{\beta}_j)$  es más alta relativo a cuando  $x_j$  no se correlaciona con otras variables independientes
- Ceteris paribus quisiéramos un  $\text{FIV}_j$  bajo pero rara vez podemos escoger
- Un  $\text{FIV}_j$  “alto” para los controles, no debería afectar decisión de incluirlos
  - \* Ignoramos  $\text{FIV}_j$  de controles
- Fijar un valor límite para  $\text{FIV}_j$  y declarar que probablemente hay MC para  $\hat{\beta}_j$  es arbitrario
  - \* Pero a veces se usa 10 ( $R_j^2 > 0.9$ )
  - \*  $\text{FIV}_j > 10 \not\Rightarrow$  desvest  $(\hat{\beta}_j)$  sea alta porque desvest  $(\hat{\beta}_j)$  también depende de  $\sigma$  y  $\text{SCT}_j$  (puede crecer si  $n$  crece)
- En conclusión,  $\text{FIV}_j$  tiene un uso limitado

### 3.4.2 Varianzas en modelos mal especificados

Relación inversa entre sesgo y varianza:

- Puede ayudar a decidir sobre si incluir una variable en un modelo de regresión

Analicemos la varianza de los estimadores de MCO cuando se omite una variable relevante

- Supongamos que el siguiente modelo poblacional satisface los supuestos G-M

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

y consideremos 2 estimadores para  $\beta_1$ :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

- Si  $\beta_2 \neq 0$  y  $\text{Corr}(x_1, x_2) \neq 0$ ,  $\tilde{\beta}_1$  estará sesgado ( $\mathbb{E}(\tilde{\beta}_1) \neq \beta_1$ )
- Mientras que  $\mathbb{E}(\hat{\beta}_1) = \beta_1 \quad \forall \beta_2$ , incluido  $\beta_2 = 0$

- En este sentido,  $\widehat{\beta}_1$  parece mejor a  $\widetilde{\beta}_1$ , pero

$$\text{Var}(\widetilde{\beta}_1) = \frac{\sigma^2}{\text{SCT}_1} < \frac{\sigma^2}{\text{SCT}_1(1 - R_1^2)} = \text{Var}(\widehat{\beta}_1)$$

Conclusiones si  $\text{Corr}(x_1, x_2) \neq 0$

1. Si  $\beta_2 \neq 0$ ,

$$\mathbb{E}(\widetilde{\beta}_1) \neq \beta_1, \mathbb{E}(\widehat{\beta}_1) = \beta_1 \text{ y } \text{Var}(\widetilde{\beta}_1) < \text{Var}(\widehat{\beta}_1)$$

- Se prefiere  $\widehat{\beta}_1$  cuando  $n$  es grande
  - MC inducida por incluir  $x_2$  se vuelve menos importante
  - Ambas varianzas disminuyen cuando  $n \rightarrow \infty$
- Si  $x_2$  se excluye,  $x_2$  estaría en  $u$ ,  $\sigma^2$  subiría y  $\text{Var}(\widetilde{\beta}_1)$  subiría

2. Si  $\beta_2 = 0$ ,

$$\mathbb{E}(\widetilde{\beta}_1) = \beta_1, \mathbb{E}(\widehat{\beta}_1) = \beta_1 \text{ y } \text{Var}(\widetilde{\beta}_1) < \text{Var}(\widehat{\beta}_1)$$

- Se prefiere  $\widetilde{\beta}_1$
- Incluir  $x_2$  cuando no tiene efecto en  $y$ , agrava el problema de MC
  - El estimador de  $\beta_1$  es menos eficiente
- Una varianza más grande para el estimador de  $\beta_1$  es el costo de incluir una variable irrelevante en el modelo

### 3.4.3 Estimación de $\sigma^2$

Conocemos  $\text{Var}(\widehat{\beta}_1)$ , ¿por qué no podemos calcularla?

Al obtener un estimador insesgado de  $\sigma^2$ , tendríamos un estimador insesgado de  $\text{Var}(\widehat{\beta}_1)$

- Dado que  $\sigma^2 = \mathbb{E}(u^2)$ , un ‘estimador’ insesgado de  $\sigma^2$  sería  $\widetilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n u_i^2$ , pero no observamos los errores  $u_i$  porque no conocemos las  $\beta_j$ ’s

$$u_i = y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik}$$

- Si reemplazamos las  $\beta_j$ ’s con  $\widehat{\beta}_j$ ’s obtenemos los residuales de MCO

$$\widehat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \dots - \widehat{\beta}_k x_{ik}$$

- Sin embargo, no reemplazamos  $\widehat{u}_i$  por  $u_i$  para obtener  $\check{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \widehat{u}_i^2$  porque generaría un estimador sesgado de  $\sigma^2$ 
  - No toma en cuenta los grados de libertad

El estimador insesgado de  $\sigma^2$  en RLM es

$$\widehat{\sigma}^2 = \frac{1}{n - k - 1} \sum_{i=1}^n \widehat{u}_i^2 = \frac{\text{SCR}}{n - k - 1}$$

- $n - k - 1$  son los grados de libertad (g.l.)
  - Número de observaciones ( $n$ ) menos parametros estimados ( $k + 1$ )
  - En RLS,  $k = 1$
- Intuitivamente, el ajuste por g.l. se necesita por las condiciones de primer orden

$$\sum_{i=1}^n \hat{u}_i = 0 \quad y \quad \sum_{i=1}^n x_{ij} \hat{u}_i = 0 \quad \forall j = 1, 2, \dots, k$$

- Hay  $k + 1$  restricciones impuestas sobre los residuales
- Dados  $n - (k + 1)$  de los residuales, los  $k + 1$  residuales restantes se conocen
- Solo hay  $n - (k + 1)$  g.l. en los residuales
- $n > n - k - 1 \implies \frac{1}{n-k-1} > \frac{1}{n} \implies \frac{1}{n-k-1} \sum_{i=1}^n \hat{u}_i^2 > \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2$

Teorema. Estimación Insegada de  $\sigma^2$   
 Bajo supuestos RLM.1 a RLM.5,  $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$

$\hat{\sigma} = +\sqrt{\hat{\sigma}^2}$  se llama el **error estándar de la regresión** (EER)

- Estimador de la desviación estándar de  $u$
- También se conoce como
  - Error estándar del estimado
  - Raíz del error cuadrático medio

Si agregamos una variable independientes a la regresión,  $\hat{\sigma}$  puede subir o bajar

- SCR podría bajar (numerador)
- $n - (k + 1)$  podría bajar en 1 (denominador)
- No podemos saber de antemano que efecto va a dominar

### 3.4.4 Errores estándar de los estimadores de MCO

Para construir IC y hacer PH, necesitamos estimar la desviación estándar de  $\hat{\beta}_j$

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SCT}_j(1 - R_j^2)} \implies \text{desvest}(\hat{\beta}_j) = \frac{\sigma}{\sqrt{\text{SCT}_j(1 - R_j^2)}}$$

Como  $\sigma$  no es conocida, usamos su estimador  $\hat{\sigma}$  y obtenemos el **error estándar de  $\hat{\beta}_j$**

$$\text{errest}(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{\text{SCT}_j(1 - R_j^2)}}$$

- Los estimados de MCO y el error estándar se pueden calcular para cualquier muestra aleatoria



- $\text{errest}(\hat{\beta}_j)$  depende de  $\hat{\sigma}$  por lo que tiene una distribución muestral
  - Se utiliza para hacer inferencia
- $\text{errest}(\hat{\beta}_j)$  depende de RLM.5
  - No es válido para estimar  $\text{desvest}(\hat{\beta}_j)$  cuando los errores exhiben heterocedasticidad
  - Heterocedasticidad no genera sesgo en  $\hat{\beta}_j$  pero sí en  $\text{Var}(\hat{\beta}_j)$ 
    - \* Inválida  $\text{errest}(\hat{\beta}_j)$
  - Importante: Software reporta  $\text{errest}(\hat{\beta}_j)$  bajo RLM.5 por default
  - Si sospechamos de heterocedasticidad, los errores estándar ‘usuales’ no son válidos y tenemos que tomar medidas para corregirlos
- Si definimos  $\text{desvest}(x_j) = \sqrt{\text{SCT}_j/n}$ , podemos re-expresar
 
$$\text{errest}(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{n} \text{desvest}(x_j) \sqrt{1 - R_j^2}}$$
  - Muestra cómo  $n$  afecta directamente al error estándar
  - Otros 3 términos ( $\hat{\sigma}$ ,  $\text{desvest}(x_j)$ ,  $R_j^2$ ) cambian con diferentes muestras aleatorias pero convergen a constantes conforme  $n \rightarrow \infty$
  - $\text{errest} \rightarrow 0$  a una tasa  $1/\sqrt{n}$
- Valor de tener más datos:
  - La precisión de  $\hat{\beta}_j$  aumenta conforme  $n \rightarrow \infty$
- En contraste, falta de sesgo se cumple para cualquier  $n$  en tanto se puedan calcular los estimadores

### 3.5 Eficiencia de MCO: Teorema Gauss-Markov

Sabemos que bajo RLM.1 a RLM.4, MCO es insesgado

- Pero hay muchos estimadores insesgados

¿Hay estimadores con varianza menor que MCO?

- Si limitamos apropiadamente el conjunto de estimadores potenciales, se puede probar que MCO es el mejor en ese conjunto

Teorema Gauss-Markov

Bajo supuestos RLM.1 a RLM.5,  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  son los mejores estimadores lineales insesgados (MELI) de  $\beta_0, \beta_1, \dots, \beta_k$ , respectivamente

Significado de MELI:

- Estimador: Regla que se puede aplicar a los datos para producir un estimado
- Estimador insesgado:  $\mathbb{E}(\tilde{\beta}_j) = \beta_j \quad \forall j = 0, 1, \dots, k$
- Estimador lineal: Si se puede expresar como función lineal de la variable dependiente

$$\tilde{\beta}_j = \sum_{i=1}^n w_{ij} y_i$$

- Cada  $w_{ij}$  es una función de los valores de la muestra de todas las  $x_j$ 's
- Los estimadores de MCO son lineales,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1} y_i}{\sum_{i=1}^n \hat{r}_{i1}^2}$$

- Mejor estimador: Tiene la menor varianza
  - Dados 2 estimadores, preferimos el que tenga la menor varianza

Anotaciones sobre el teorema G-M:

- El teorema justifica el uso del método de MCO para estimar modelos de regresión múltiple
  - Para cada estimador  $\tilde{\beta}_j$  que sea lineal e insesgado,  $\text{Var}(\hat{\beta}_j) \leq \text{Var}(\tilde{\beta}_j)$ 
    - \* Generalmente con desigualdad estricta
  - Cuando los supuestos G-M se cumplen, ningún estimador lineal e insesgado será mejor que MCO
  - Si nos dan un estimador lineal e insesgado, sabemos que su varianza es al menos igual que la varianza de MCO
- Si queremos estimar una función lineal de las  $\beta_j$ 's, la combinación correspondiente de estimadores de MCO tiene la menor varianza entre todos los estimadores lineales insesgados
- Los supuestos RLM.1 a RLM.5 se conocen como supuestos G-M (para corte transversal) por este teorema
- Si alguno de los supuestos no se cumple, el teorema ya no se cumple
  - Sin RLM.4, MCO es sesgado
  - Sin RLM.5, MCO ya no tiene varianza mínima entre los estimadores lineales e insesgados
    - \* En ese caso, MC ponderados son mejores que MCO

### 3.6 Sobre el lenguaje de RLM

¿Cuál de los siguientes enunciados es correcto?

- Estimar un modelo MCO
- Estimar un modelo lineal por MCO

Diferencia entre modelo y método de estimación

- MCO es un método de estimación, no un modelo
  - Un modelo describe la población y depende de parámetros desconocidos
- El modelo lineal visto se puede escribir en la población como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

- $\beta_j$  son parámetros
  - Podemos hablar del significado de  $\beta_j$  sin tener datos
  - La interpretación viene del modelo
- Con una muestra aleatoria podemos estimar los parámetros
  - Hasta ahora, con MCO bajo los supuestos G-M
  - Bajo diferentes supuestos, se prefieren otros métodos de estimación
    - \* WLS, LAD, IV