

Microeconometría I
Problem Set 2
Extensiones de Modelos Logit y Probit

- **Ejercicio 1.** Considere la siguiente afirmación. “La estimación de un modelo de probabilidad lineal es más robusta que probit o logit porque el modelo de probabilidad lineal no asume homocedasticidad ni tiene supuestos acerca de la distribución de los errores.”

- **Ejercicio 2.** *Probit con una variable no observable.*

Considere el modelo Probit:

$$P(y = 1 \mid \mathbf{z}, q) = \Phi(\mathbf{z}_1 \delta_1 + \gamma_1 z_2 q),$$

donde q es independiente de \mathbf{z} y distribuido $\text{Normal}(0, 1)$; el vector \mathbf{z} es observado pero el escalar q no lo es.

1. Encuentre el efecto parcial de z_2 sobre la probabilidad de respuesta, a saber, $\frac{\partial P(y=1|\mathbf{z},q)}{\partial z_2}$.
2. Muestre que $P(y = 1 \mid \mathbf{z}) = \Phi\left[\mathbf{z}_1 \delta_1 / (1 + \gamma_1^2 z_2^2)^{1/2}\right]$.
3. Defina $\rho_1 \equiv \gamma_1^2$. ¿Cómo testaría la hipótesis $H_0 : \rho_1 = 0$?
4. Si tuviera motivos para creer que $\rho_1 > 0$, ¿cómo estimaría δ_1 junto con ρ_1 ?

- **Ejercicio 3.** *Probit con endogeneidad.*

Considere una gran muestra aleatoria de trabajadores en un momento dado. Sea $sick_i$ una variable que vale 1 si la persona i se reportó enferma durante los últimos 90 días, y vale 0 en caso contrario. Sea \mathbf{z}_i un vector de características del individuo y del empleador. Sea $cigs_i$ el número de cigarrillos que fuma el individuo i por día (en promedio).

1. Explique el experimento subyacente de interés cuando queremos examinar los efectos del tabaquismo en los días de trabajo perdidos.
2. ¿Por qué $cigs_i$ podría estar correlacionada con variables no observables que afectan a $sick_i$?
3. Una forma de escribir el modelo de interés es

$$P(sick_i = 1 \mid \mathbf{z}_i, cigs_i, q_1) = \Phi(\mathbf{z}_1 \delta_1 + \gamma_1 cigs_i + q_1)$$

donde \mathbf{z}_1 es un subconjunto de \mathbf{z} y q_1 es una variable no observable que posiblemente esté correlacionada con $cigs$. ¿Qué sucede si se ignora q_1 y se estima el probit de $sick$ sobre \mathbf{z}_1 y $cigs$?

4. ¿Puede $cigs$ tener una distribución normal condicional en la población? Explique.
5. Explique cómo probar si $cigs$ es exógeno. ¿Esta prueba se basa en $cigs$ que tienen una distribución normal condicional?
6. Suponga que algunos de los trabajadores viven en estados que recientemente implementaron leyes de no fumar en el lugar de trabajo. ¿La presencia de las nuevas leyes sugiere un buen candidato IV para $cigs$?

Microeconometría I
Problem Set 2 - Stata
Extensiones de Modelos Logit y Probit

■ **Ejercicio 1.** Utilice el conjunto de datos BWGHT.RAW para este problema.

1. Defina una variable binaria, *smokes*, si la mujer fuma durante el embarazo. Estime un modelo probit que relacione *smokes* con *motheduc*, *white* y $\log(faminc)$. En *white* = 0 y *faminc* evaluado en el promedio de la muestra, ¿cuál es la diferencia estimada en la probabilidad de fumar para una mujer con 16 años de educación y uno con 12 años de educación?
2. ¿Cree que *faminc* es exógena en la ecuación de *smokes*? ¿Qué pasa con *motheduc*?
3. Suponga que *motheduc* y *white* son exógenos en el probit de la parte 1. Suponga también que *fatheduc* es exógeno a esta ecuación. Estime la forma reducida de $\log(faminc)$ para ver si *fatheduc* está parcialmente correlacionada $\log(faminc)$.
4. Contraste la hipótesis nula de que $\log(faminc)$ es exógeno en el probit del inciso 1.

■ **Ejercicio 2.** Precios endógenos o exógenos

Una preocupación común cuando se utilizan precios autoinformados en la estimación de la prevalencia del tabaquismo con una base de datos de corte transversal (por ejemplo, *Global Adult Tobacco Survey* o GATS) es la potencial endogeneidad de esta variable. Para abordar este problema potencial, se construyen dos variables de precios diferentes. La primera variable de precio asigna a los fumadores el precio autoinformado pagado por la última compra y utiliza una imputación de regresión aleatoria (*random regression imputation*, a veces denominada imputación de regresión estocástica) para asignar un precio a los no fumadores de la muestra. La segunda variable de precio asigna a fumadores y no fumadores el promedio del precio autoinformado por unidad primaria de muestreo (UPM, o PSU por *primary sampling unit*). Siguiendo las recomendaciones en *Economics of Tobacco Toolkit: Economic analysis of demand using data from the Global Adult Tobacco Survey (GATS)* (John et al, 2019), se puede verificar la endogeneidad del precio autoinformado utilizando el test de Rivers-Vuong (1988).

1. ¿Por qué podrían ser endógenos los precios autoinformados?
2. Realice el test de Rivers-Vuong para los datos provistos en *pricedata.dta* utilizando las variables X en la primera etapa y Z en la segunda etapa.
3. En función de los resultados, estime la elasticidad de la prevalencia del tabaquismo con respecto a los precios.

■ **Ejercicio 3.** Probit heterocedástico y simulaciones

Buscamos simular el siguiente modelo:

$$\Pr(y = 1) = F\{(\beta_0 + \beta_1 x) / \exp(\gamma_1 x_{het})\}$$

Genere un dataset vacío con 1000 observaciones. Genere las siguientes variables:

$$\begin{aligned}x &\sim \mathbb{U}(-1, 1) \\x_{het} &\sim \mathbb{U}(0, 1) \\ \sigma &\sim e^{1.5 \cdot x_{het}} \\ p &\sim \mathcal{N}\left(\frac{\beta_0 + \beta_1 \cdot x}{\sigma}\right)\end{aligned}$$

con $\beta_0 = 0.3$ y $\beta_1 = 2$ y defina la variable dependiente y como una variable binaria que vale 1 si p es mayor o igual a una variable aleatoria uniforme en el intervalo $(0,1)$ y 0 en caso contrario. Estime el modelo probit heterocedástico y compare con las estimaciones del probit usual.