Econometría 1 - Actividad 7

Modelos de Variable Dependiente Limitada

Ejercicios 5-6: LPM, Probit, Logit

División de Economía - CDE
Dr. Francisco Cabrera

1 Introducción: Modelos de Variable Dependiente Limitada

1.1 El Problema

Hasta ahora hemos trabajado con variables dependientes **continuas**:

- Salarios: log(wage)
- Años de educación
- Producción, consumo, etc.

Pero muchas veces la variable dependiente es **binaria** (0 o 1):

$$Y_i = \begin{cases} 1 & \text{si el evento ocurre} \\ 0 & \text{si el evento no ocurre} \end{cases} \tag{1}$$

Ejemplos:

- ¿Se gradúa el estudiante? (1 = sí, 0 = no)
- ¿Se rechaza el crédito? (1 = sí, 0 = no)
- ¿Participa en la fuerza laboral? (1 = sí, 0 = no)
- ¿Compra el producto? (1 = sí, 0 = no)

1.2 Por Qué MCO No Funciona Bien

Si usamos MCO directamente con Y binaria:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \tag{2}$$

Tenemos varios problemas:

Problemas de MCO con variable dependiente binaria:

1. **Predicciones fuera de [0,1]:**
   - MCO puede predecir $\hat{Y}_i < 0$ o $\hat{Y}_i > 1$
   - No tiene sentido como probabilidad
2. **Heterocedasticidad por construcción:**

$$\text{Var}(u_i|X_i) = P_i(1 - P_i) \tag{3}$$

   donde $P_i = P(Y_i = 1|X_i)$. La varianza depende de X_i.

3. **No linealidad:**
   - El efecto marginal de X no es constante
   - Debería depender del nivel de X

2 Tres Modelos Principales

2.1 Modelo de Probabilidad Lineal (LPM)

Modelo de Probabilidad Lineal:

$$P(Y_i = 1|\mathbf{X}_i) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} \tag{4}$$

Se estima usando MCO estándar.

Ventajas del LPM

- Simple de estimar (MCO estándar)
- Coeficientes son directamente interpretables como efectos marginales
- No requiere supuestos distribucionales

Desventajas del LPM

- Puede predecir probabilidades < 0 o > 1
- Heterocedasticidad por construcción
- Asume efectos marginales constantes (no realista)

Solución a la heterocedasticidad

Usar errores estándar **robustos** (heteroskedasticity-robust):

$$\widehat{\text{Var}}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Omega}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{5}$$

donde $\widehat{\mathbf{\Omega}} = \text{diag}(\hat{u}_i^2)$ (estimador HC1 o HC3).

2.2 Modelo Probit

Modelo Probit:

$$P(Y_i = 1|\mathbf{X}_i) = \Phi(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) \tag{6}$$

donde $\Phi(\cdot)$ es la función de distribución acumulada (CDF) de la distribución normal estándar $N(0,1)$:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{7}$$

Interpretación

El modelo Probit puede derivarse de un modelo de **utilidad latente**:

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i \tag{8}$$

donde $\varepsilon_i \sim N(0,1)$ y:

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* > 0 \\ 0 & \text{si } Y_i^* \leq 0 \end{cases} \tag{9}$$

Por lo tanto:

$$\begin{aligned} P(Y_i = 1|\mathbf{X}_i) &= P(Y_i^* > 0|\mathbf{X}_i) \\ &= P(\varepsilon_i > -\mathbf{X}_i\beta) \\ &= P(\varepsilon_i \leq \mathbf{X}_i\beta) \\ &= \Phi(\mathbf{X}_i\beta) \end{aligned} \tag{10}$$

2.3 Modelo Logit

Modelo Logit:

$$P(Y_i = 1|\mathbf{X}_i) = \Lambda(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) \tag{11}$$

donde $\Lambda(\cdot)$ es la función logística:

$$\Lambda(z) = \frac{\exp(z)}{1+\exp(z)} = \frac{1}{1+\exp(-z)} \tag{12}$$

Interpretación en términos de odds

El modelo Logit puede expresarse en términos de **log-odds**:

$$\log\left(\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}\right) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} \tag{13}$$

Esto hace que el modelo sea lineal en los log-odds.

2.4 Probit vs Logit

Diferencias entre Probit y Logit:

- **Probit:** Supone distribución normal de errores

$$\varepsilon_i \sim N(0, 1) \tag{14}$$

- **Logit:** Supone distribución logística de errores

$$\varepsilon_i \sim \text{Logística}(0, \pi^2/3) \tag{15}$$

La distribución logística tiene **colas más pesadas** que la normal.
En la práctica:

- Ambos modelos suelen dar resultados MUY SIMILARES
- Los coeficientes del Logit son $\approx$ 1.6-1.8 veces los del Probit
- Razón: $\sigma_{\text{logística}} = \pi/\sqrt{3} \approx 1.81$ vs $\sigma_{\text{normal}} = 1$
- Los AME (efectos marginales promedio) son casi idénticos

3 Efectos Marginales

3.1 Por Qué los Coeficientes No Son Efectos Marginales

En los modelos Probit y Logit, los coeficientes β_j **NO** son efectos marginales.

Probit

El efecto marginal de X_j sobre la probabilidad es:

Efecto marginal en Probit:

$$\frac{\partial P(Y_i = 1|\mathbf{X}_i)}{\partial X_{ji}} = \phi(\mathbf{X}_i\beta) \cdot \beta_j \tag{16}$$

donde $\phi(\cdot)$ es la función de densidad de probabilidad (PDF) de $N(0, 1)$:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{17}$$

El efecto marginal **depende** de $\mathbf{X}_i$ (no es constante).

Logit

Efecto marginal en Logit:

$$\frac{\partial P(Y_i = 1|\mathbf{X}_i)}{\partial X_{ji}} = \Lambda(\mathbf{X}_i\beta) \cdot [1 - \Lambda(\mathbf{X}_i\beta)] \cdot \beta_j \tag{18}$$

Equivalentemente:

$$\frac{\partial P}{\partial X_{ji}} = P_i(1 - P_i)\beta_j \tag{19}$$

donde $P_i = P(Y_i = 1|\mathbf{X}_i)$.

3.2 Efecto Marginal Promedio (AME)

Como el efecto marginal varía con $\mathbf{X}_i$, se reporta el **Average Marginal Effect**:

AME (Average Marginal Effect):

$$\text{AME}_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial P(Y_i = 1|\mathbf{X}_i)}{\partial X_{ji}} \tag{20}$$

Para Probit:

$$\text{AME}_j = \frac{1}{n} \sum_{i=1}^{n} \phi(\mathbf{X}_i\hat{\beta}) \cdot \hat{\beta}_j \tag{21}$$

Para Logit:

$$\text{AME}_j = \frac{1}{n} \sum_{i=1}^{n} \hat{P}_i(1 - \hat{P}_i)\hat{\beta}_j \tag{22}$$

donde $\hat{P}_i = \Lambda(\mathbf{X}_i\hat{\beta})$.

3.3 Efecto Marginal para Variable Binaria

Para una variable **binaria** (como single), el efecto marginal es:

$$\Delta P_i = P(Y_i = 1|X_{ji} = 1, \mathbf{X}_{-j}) - P(Y_i = 1|X_{ji} = 0, \mathbf{X}_{-j}) \tag{23}$$

El AME para variable binaria es:

$$\text{AME}_{\text{bin}} = \frac{1}{n}\sum_{i=1}^{n}[P(Y_i = 1|X_{ji} = 1, \mathbf{X}_{-j,i}) - P(Y_i = 1|X_{ji} = 0, \mathbf{X}_{-j,i})] \tag{24}$$

4 Ejercicio 5: Modelo Logit - Graduación de Estudiantes-Atletas

4.1 Contexto del Problema

Tenemos datos de 420 estudiantes-atletas. La variable dependiente es:

$$\text{grad}_i = \begin{cases} 1 & \text{si se gradúa en 5 años} \\ 0 & \text{si no se gradúa} \end{cases} \tag{25}$$

Variables explicativas:

- hsGPA: Promedio de calificaciones en high school
- SAT: Puntaje en examen SAT
- study: Horas por semana en sala de estudio

4.2 Modelo Estimado

El modelo Logit estimado es:

$$\hat{P}(\text{grad} = 1|\text{hsGPA}, \text{SAT}, \text{study}) = \Lambda(-1.17 + 0.24 \cdot \text{hsGPA} + 0.00058 \cdot \text{SAT} + 0.073 \cdot \text{study}) \tag{26}$$

donde:

$$\Lambda(z) = \frac{\exp(z)}{1 + \exp(z)} \tag{27}$$

4.3 Ejercicio 5a: Diferencia en Probabilidad

Objetivo: Calcular la diferencia en probabilidad de graduarse entre alguien que pasa 10 horas vs 5 horas en study hall, manteniendo hsGPA = 3.0 y SAT = 1200.

Cálculo para study = 10

$$\begin{aligned} z_1 &= -1.17 + 0.24(3.0) + 0.00058(1200) + 0.073(10) \\ &= -1.17 + 0.72 + 0.696 + 0.73 \\ &= 0.976 \end{aligned} \tag{28}$$

$$P_1 = \Lambda(0.976) = \frac{\exp(0.976)}{1 + \exp(0.976)} = \frac{2.654}{3.654} = 0.726 \tag{29}$$

Cálculo para study = 5

$$\begin{aligned} z_2 &= -1.17 + 0.24(3.0) + 0.00058(1200) + 0.073(5) \\ &= -1.17 + 0.72 + 0.696 + 0.365 \\ &= 0.611 \end{aligned} \tag{30}$$

$$P_2 = \Lambda(0.611) = \frac{\exp(0.611)}{1 + \exp(0.611)} = \frac{1.842}{2.842} = 0.648 \tag{31}$$

Diferencia

$$\Delta P = P_1 - P_2 = 0.726 - 0.648 = 0.078 \tag{32}$$

Interpretación:

Un estudiante que pasa 10 horas por semana en study hall tiene una probabilidad de graduarse que es **7.8 puntos porcentuales mayor** que un estudiante que solo pasa 5 horas, manteniendo constantes hsGPA = 3.0 y SAT = 1200.

Efecto Marginal Instantáneo (opcional)

Para calcular el efecto marginal en el punto medio (study = 7.5):

$$z_{\text{medio}} = -1.17 + 0.24(3.0) + 0.00058(1200) + 0.073(7.5) = 0.7935 \tag{33}$$

$$\Lambda(0.7935) = 0.689 \tag{34}$$

$$\frac{\partial P}{\partial \text{study}} = \Lambda(z)[1 - \Lambda(z)] \cdot \beta_{\text{study}} = 0.689 \cdot 0.311 \cdot 0.073 = 0.0156 \tag{35}$$

Una hora adicional aumenta la probabilidad en aproximadamente 1.56 puntos porcentuales.

5 Ejercicio 6: Datos HMDA

5.1 Contexto

Datos de solicitudes de hipoteca en Boston (1990).

Variable dependiente:

$$\text{deny}_i = \begin{cases} 1 & \text{si la solicitud fue rechazada} \\ 0 & \text{si fue aprobada} \end{cases} \tag{36}$$

Variable explicativa principal: pirat (payment-to-income ratio).

5.2 Ejercicio 6a: Modelo de Probabilidad Lineal

LPM:

$$P(\text{deny}_i = 1|\text{pirat}_i) = \beta_0 + \beta_1 \text{pirat}_i \tag{37}$$

Estimación: MCO con errores estándar robustos.

Interpretación de β_1:

Un aumento de 1 unidad en pirat aumenta la probabilidad de rechazo en $\beta_1 \times 100$ puntos porcentuales.

5.3 Ejercicio 6b: Modelo Probit

Probit:

$$P(\text{deny}_i = 1|\text{pirat}_i) = \Phi(\beta_0 + \beta_1 \text{pirat}_i) \tag{38}$$

Estimación: Máxima verosimilitud.

Nota: Los coeficientes no son directamente interpretables. Calcular AME.

5.4 Ejercicio 6c: Ecuación Estimada del Probit

Después de estimar, la ecuación toma la forma:

$$\hat{P}(\text{deny} = 1|\text{pirat}) = \Phi(\hat{\beta}_0 + \hat{\beta}_1 \text{pirat}) \tag{39}$$

5.5 Ejercicio 6d: Predicción para Cambio en pirat

Para pirat = 0.3:

$$z_1 = \hat{\beta}_0 + \hat{\beta}_1(0.3) \tag{40}$$

$$\hat{P}_1 = \Phi(z_1) \tag{41}$$

Para pirat = 0.4:

$$z_2 = \hat{\beta}_0 + \hat{\beta}_1(0.4) \tag{42}$$

$$\hat{P}_2 = \Phi(z_2) \tag{43}$$

Diferencia:

$$\Delta\hat{P} = \hat{P}_2 - \hat{P}_1 \tag{44}$$

5.6 Ejercicio 6e: Modelo con Variable 'single'

Probit extendido:

$$P(\text{deny}_i = 1|\text{pirat}_i, \text{single}_i) = \Phi(\beta_0 + \beta_1\text{pirat}_i + \beta_2\text{single}_i) \tag{45}$$

donde:

$$\text{single}_i = \begin{cases} 1 & \text{si es soltero} \\ 0 & \text{si no es soltero} \end{cases} \tag{46}$$

AME de 'single'

Para una variable binaria:

$$\text{AME}_{\text{single}} = \frac{1}{n}\sum_{i=1}^{n}[\Phi(\hat{\beta}_0 + \hat{\beta}_1\text{pirat}_i + \hat{\beta}_2) - \Phi(\hat{\beta}_0 + \hat{\beta}_1\text{pirat}_i)] \tag{47}$$

5.7 Ejercicio 6f: Ecuación Estimada con 'single'

$$\hat{P}(\text{deny} = 1|\text{pirat}, \text{single}) = \Phi(\hat{\beta}_0 + \hat{\beta}_1\text{pirat} + \hat{\beta}_2\text{single}) \tag{48}$$

Interpretación de signos:

- Si $\hat{\beta}_1 > 0$: Mayor pirat $\rightarrow$ Mayor probabilidad de rechazo
- Si $\hat{\beta}_2 > 0$: Ser soltero $\rightarrow$ Mayor probabilidad de rechazo

5.8 Ejercicio 6g: Modelo Logit

Logit:

$$P(\text{deny}_i = 1|\text{pirat}_i, \text{single}_i) = \Lambda(\beta_0 + \beta_1\text{pirat}_i + \beta_2\text{single}_i) \tag{49}$$

donde:

$$\Lambda(z) = \frac{\exp(z)}{1 + \exp(z)} \tag{50}$$

Relación con Probit:
Los coeficientes del Logit suelen ser $\approx$ 1.6-1.8 veces los del Probit:

$$\hat{\beta}_{\text{Logit}} \approx \frac{\pi}{\sqrt{3}}\hat{\beta}_{\text{Probit}} \approx 1.81\,\hat{\beta}_{\text{Probit}} \tag{51}$$

5.9 Ejercicio 6h: Pseudo R² y LR Test

Pseudo R² (McFadden)

Pseudo R² de McFadden:

$$\text{Pseudo } R^2 = 1 - \frac{\ln \mathcal{L}_1}{\ln \mathcal{L}_0} \tag{52}$$

donde:

- $\ln \mathcal{L}_1$ = log-verosimilitud del modelo completo
- $\ln \mathcal{L}_0$ = log-verosimilitud del modelo nulo (solo intercepto)

Interpretación:

- Similar al R^2 de MCO, pero NO es proporción de varianza explicada
- Valores típicos: 0.2–0.4 se consideran buenos en modelos binarios
- Valores cercanos a 1 son raros (e inusuales)

Prueba de Razón de Verosimilitud (LR Test)

LR Test:

Hipótesis:

$$\begin{aligned} &H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ &H_1 : \text{Al menos un } \beta_j \neq 0 \end{aligned} \tag{53}$$

Estadístico:

$$LR = 2[\ln \mathcal{L}_1 - \ln \mathcal{L}_0] \tag{54}$$

Distribución bajo H_0:

$$LR \sim \chi^2(k) \tag{55}$$

donde k = número de restricciones (número de coeficientes excluido el intercepto).

Regla de decisión:

- Si $LR > \chi^2_{k,\alpha} \rightarrow$ Rechazar H_0
- Equivalentemente: Si p-valor $< \alpha \rightarrow$ Rechazar H_0

Interpretación:

Si rechazamos H_0, concluimos que el modelo completo es significativamente mejor que el modelo nulo. Las variables explicativas tienen poder predictivo.

6 Comparación de Modelos

Aspecto	**LPM**	**Probit**	**Logit**
Método de estimación	MCO	MLE	MLE
Probabilidades en [0,1]	No	Sí	Sí
Coef. = Efectos marginales	Sí	No	No
Heterocedasticidad	Sí	No (condicional)	No (condicional)
Efectos marginales	Constantes	Variables	Variables
Interpretación	Fácil	Requiere AME	Requiere AME
Supuesto distribucional	Ninguno	Normal	Logística

Table 1: Comparación de modelos para variable dependiente binaria

6.1 Recomendaciones Prácticas

1. **Para análisis exploratorio:**
   - Comenzar con LPM (simple y rápido)
   - Usar errores estándar robustos
   - Verificar si hay predicciones fuera de [0,1]
2. **Para análisis final:**
   - Preferir Probit o Logit
   - Reportar AME además de coeficientes
   - Comparar resultados de ambos modelos
3. **Probit vs Logit:**
   - En práctica, da igual (resultados muy similares)
   - Probit: más común en economía por tradición
   - Logit: más común en otras ciencias sociales
   - Elegir según preferencia o convención del campo
4. **Reportar:**
   - Coeficientes con errores estándar
   - AME para variables clave
   - Pseudo R^2 o LR test
   - Interpretación sustantiva de efectos

7 Resumen de Fórmulas Clave

Fórmulas Esenciales:

1. **Función logística:**

$$\Lambda(z) = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + \exp(-z)}$$

2. **CDF normal estándar:**

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

3. **PDF normal estándar:**

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

4. **Efecto marginal (Probit):**

$$\frac{\partial P}{\partial X_j} = \phi(\mathbf{X}\beta) \cdot \beta_j$$

5. **Efecto marginal (Logit):**

$$\frac{\partial P}{\partial X_j} = \Lambda(\mathbf{X}\beta)[1 - \Lambda(\mathbf{X}\beta)] \cdot \beta_j$$

6. **AME:**

$$\text{AME}_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial P_i}{\partial X_{ji}}$$

7. **Pseudo R^2:**

$$\text{Pseudo } R^2 = 1 - \frac{\ln \mathcal{L}_1}{\ln \mathcal{L}_0}$$

8. **LR test:**

$$LR = 2[\ln \mathcal{L}_1 - \ln \mathcal{L}_0] \sim \chi^2(k)$$