Econometría de Datos de Panel

Maestrías en Economía y Econometría

Lecture 4

Agenda

1. Modelos de Respuesta Binaria
   - Modelo Probit de Efectos no Observables
   - Modelo Logit de Efectos no Observables
2. Modelos Dinámicos de Respuesta Binaria con Efectos no Observables

Modelos de Respuesta Binaria

- En el caso de paneles, utilizar un modelo de probabilidad lineal cuando la variable dependiente es binaria tiene los mismos problemas que en el caso de una muestra de corte transversal. Go
- Por lo tanto comenzaremos discutiendo directamente los modelos Probit y Logit.
- Para ilustrar estos modelos veremos el caso más sencillo: Pooled Probit y Logit.
- Supongamos que el modelo es: $P\left(y_{jt} = 1 \mid x_{jt}\right) = G\left(x_{jt}\beta\right)\ t = 1, 2, ..., T.$
- Donde $G(\cdot)$ es una función conocida que toma valores en el intervalo abierto cero-uno. x_{jt} son las variables explicativas del modelo que pueden incluir dummies temporales, interacciones de estas variables con variables constantes o no en el tiempo etc.

Modelos de Respuesta Binaria

- Dada esta especificación uno puede obtener estimaciones consistentes de β maximizando el logaritmo de la función de verosimilitud:

$$\sum_{j=1}^{N}\sum_{t=1}^{T}\left\{y_{jt} logG\left(x_{jt}\beta\right) + \left(1 - y_{jt}\right) log\left[1 - G\left(x_{jt}\beta\right)\right]\right\}$$

- Sin supuestos adicionales se debe utilizar un estimador robusto de la matriz de varianzas y covarianzas.

Modelos de Respuesta Binaria

- En particular un estimador simple de esta matriz viene dado por:

$$\left[\sum_{j=1}^{N}\sum_{t=1}^{T} A_{jt}\left(\hat{\beta}\right)\right]^{-1}\left[\sum_{j=1}^{N} S_j\left(\hat{\beta}\right) S_j\left(\hat{\beta}\right)'\right]\left[\sum_{j=1}^{N}\sum_{t=1}^{T} A_{jt}\left(\hat{\beta}\right)\right]^{-1} \tag{1}$$

- Donde,

$$A_{jt}\left(\hat{\beta}\right) = \frac{\left\{g\left(x_{jt}\hat{\beta}\right)\right\}^2 x_{jt}' x_{jt}}{G\left(x_{jt}\hat{\beta}\right)\left[1 - G\left(x_{jt}\hat{\beta}\right)\right]}$$

Modelos de Respuesta Binaria

- y por otro lado se tiene que:

$$S_j\left(\hat{\beta}\right)=\sum_{t=1}^{T}S_{jt}\left(\hat{\beta}\right)=\sum_{t=1}^{T}\frac{g\left(x_{jt}\hat{\beta}\right)x_{jt}^{'}\left[y_{jt}-G\left(x_{jt}\hat{\beta}\right)\right]}{G\left(x_{jt}\hat{\beta}\right)\left[1-G\left(x_{jt}\hat{\beta}\right)\right]}$$

con

$$g\left(x_{jt}\hat{\beta}\right)=\frac{\partial G\left(x_{jt}\hat{\beta}\right)}{\partial x_{jt}\hat{\beta}}$$

Modelos de Respuesta Binaria

- Note que $S_{jt}\left(\hat{\beta}\right)$ es la denominada score function (i.e. la derivada primera del logaritmo de la función de verosimilitud); mientras que $A_{jt}\left(\hat{\beta}\right)$ es igual a menos la matriz Hesiana promedio estimada (i.e. la matriz de información de Fisher).
- El estimador en (1) contiene los productos cruzados de la función score y por lo tanto es robusta ante la presencia de heterocedasticidad y correlación serial.
- Si en todas las expresiones anteriores reemplazamos a $G(\cdot)$ y $g(\cdot)$ por las funciones $\Phi(\cdot)$ y $\phi(\cdot)$ respectivamente, tenemos el modelo Probit. $\Phi(\cdot)$ y $\phi(\cdot)$ son las funciones de distribución acumulada y de densidad de la normal estándar, respectivamente.
- Si en todas las expresiones anteriores reemplazamos a $G(\cdot)$ y $g(\cdot)$ por las funciones $\Lambda(\cdot)$ y $\lambda(\cdot)$ respectivamente, tenemos el modelo Logit. $\Lambda(\cdot)$ y $\lambda(\cdot)$ son las funciones de distribución acumulada y de densidad de la logística, respectivamente.

Agenda

1. Modelos de Respuesta Binaria
   - Modelo Probit de Efectos no Observables
   - Modelo Logit de Efectos no Observables
2. Modelos Dinámicos de Respuesta Binaria con Efectos no Observables

Modelo Probit de Efectos no Observables

- Este modelo tiene como supuesto principal que:

$$P\left(y_{jt}=1 \mid X_j, c_j\right)=P\left(y_{jt}=1 \mid x_{jt}, c_j\right)=\Phi\left(x_{jt}\beta+c_j\right), t=1,2,...,T \quad (2)$$

- La primera igualdad dice que x_{jt} es estríctamente exógeno condicional en c_j. Esto deja afuera a los modelos dinámicos vistos en el curso.
- La segunda igualdad de la condición anterior es el supuesto estándar del Probit.
- Además del supuesto de exogeneidad estricta, necesitamos el supuesto estándar de que la variable dependiente es independiente condicional en (X_j, c_j). Esto es:

$$y_{j1}, y_{j2}, ..., y_{jT} \text{ son independientes condicionando en } (X_j, c_j). \quad (3)$$

Modelo Probit de Efectos no Observables

- Bajo estos dos supuestos podemos construir la función de verosimilitud como:

$$L\left(c_j, \beta\right) = \prod_{j=1}^{N} \prod_{t=1}^{T} \Phi\left(x_{jt}\beta + c_j\right)^{y_{jt}} \left[1 - \Phi\left(x_{jt}\beta + c_j\right)\right]^{1-y_{jt}}$$

- Idealmente, uno podría estimar los parámetros del modelo sin restringir la relación entre c_j y x_{jt}. Con este espíritu, un modelo Probit de efectos fijos trata a los efectos no observables c_j como parámetros a estimar.
- El logaritmo de la función de verosimilitud es simplemente el logaritmo de la expresión anterior.
- Desafortunadamente, además de ser una expresión complicada de estimar, en este modelo estimar juntos los β y los c_j lleva a obtener estimadores inconsistentes para β con T fijo y $N \to \infty$.
- Para poder obtener estimaciones consistentes de β necesitamos hacer algún supuesto acerca de la relación entre c_j y x_{jt}.

Modelo Probit de Efectos no Observables

- El modelo tradicional adiciona el siguiente supuesto:

$$c_j | X_j \sim \text{Normal}(0,\ \sigma_c^2) \tag{4}$$

- Este es un supuesto fuerte porque implica que c_j y X_j son independientes y que c_j tiene distribución normal.
- Bajo los supuestos (2), (3) y (4) existe un enfoque de máxima verosimilitud condicional para estimar β y σ_c^2.
- Como los c_j no se observan, no pueden aparecer en la función de verosimilitud. En este caso, construimos la función de verosimilitud de $(y_{j1}, y_{j2}, ..., y_{jT})$ condicional a X_j, lo que requiere eliminar los c_j vía integrarlos entre menos y más infinito (i.e. tienen distribución normal). Por lo tanto: $L(c_j, \beta) =$

$$\int_{-\infty}^{\infty} \left[\prod_{j=1}^{N} \prod_{t=1}^{T} \Phi \left(x_{jt}\beta + c_j \right)^{y_{jt}} \left[1 - \Phi \left(x_{jt}\beta + c_j \right)^{1-y_{jt}} \right] \right] (1/\sigma_c)\, \phi \left(c/\sigma_c \right) dc$$

Modelo Probit de Efectos no Observables

- Tomando el logaritmo de la expresión anterior obtenemos la función objetivo a maximizar para obtener estimadores de β y σ_c^2 consistentes y asintóticamente normales.
- La estimación por máxima verosimilitud condicional recibe el nombre en la literatura de modelo Probit de efectos aleatorios.
- Los supuestos (3) y (4) son bastante fuertes y es posible relajarlos. Go Consideremos relajar el supuesto (3).
- Bajo los supuestos (2) y (4) solamente, $P\left(y_{jt} = 1 \mid X_j\right) = P\left(y_{jt} = 1 \mid x_{jt}\right) = \Phi(x_{jt}\beta_c),\ t = 1, 2, ..., T.$
- Donde $\beta_c = \beta / \left(1 + \sigma_c^2\right)^{1/2}$. Go

Modelo Probit de Efectos no Observables

- Acabamos de mostrar que Pooled Probit de y sobre X estima consistentemente: $\beta_c = \beta / \left(1 + \sigma_c^2\right)^{1/2}$ en lugar de β, por lo tanto, el problema de los efectos no observables es más grave en el modelo Probit que en modelos lineales ya que aun cuando c_j y X_j sean independientes los coeficientes estimados son inconsistentes.
- Sin embargo, usualmente uno está interesado en el efecto parcial de las variables independientes sobre la variable dependiente y en estos casos estimar β_c es igual de bueno que estimar β.
- Para ilustrar este punto calculemos el efecto parcial para una variable x_{jt} contínua,

$$\partial P\left(y_{jt} = 1 \mid X_j, c_j\right) / \partial x_{jt} = \beta_j \phi\left(x_{jt}\beta + c_j\right)$$

Modelo Probit de Efectos no Observables

- Usualmente lo que uno hace en la práctica es calcular el efecto parcial promedio. Esto es, se reemplazan las variables x_{jt} por sus promedios y se promedia la ecuación anterior a lo largo de la distribución de c en la población. Es decir: $E_c\left[\beta_j \phi\left(\overline{x}\beta + c_j\right)\right] = \left[\beta_j / \left(1+\sigma_c^2\right)^{1/2}\right] \phi\left[\overline{x}\beta / \left(1+\sigma_c^2\right)^{1/2}\right]$
- En otras palabras, pooled Probit de y sobre X estima consistentemente los efectos parciales promedio, que es lo que uno usualmente quiere.
- Sin embargo, como sucedía en el caso lineal, muchas veces el punto de introducir los efectos no observables es permitir explícitamente que estén correlacionados con las variables explicativas.
- Chamberlain (1980) permite esta correlación asumiendo una distribución normal condicional con esperanza lineal y varianza constante.

Modelo Probit de Efectos no Observables

- La versión de Mundlak (1978) del supuesto de Chamberlain es:

$$c_j \mid X_j \sim Normal\left(\psi + \overline{X}_j \xi, \sigma_a^2\right) \tag{5}$$

- Donde $\overline{X}_J$ es el promedio de las x_{jt}, $t = 1, 2, ..., T$ y σ_a^2 es la varianza de a_j en la ecuación: $c_j = \psi + \overline{X}_j \xi + a_j$
- Chamberlain permite una mayor generalidad reemplazando $\overline{X}_j$ con X_j en el supuesto anterior. Esta mayor generalización tiene como contrapartida menos grados de libertad.
- Chamberlain llamó al modelo dado por (2) y (5) un modelo Probit de efectos aleatorios, y la literatura se refiere a este modelo como el modelo Probit de efectos aleatorios de Chamberlain.

Modelo Probit de Efectos no Observables

- Si los supuestos (2), (3) y (5) se cumplen, la estimación de β, ψ, ξ, y σ_a^2 es directa porque podemos escribir el modelo en forma de variable latente como:

$$y_{jt}^* = \psi + x_{jt}\beta + \overline{X}_j\xi + a_j + u_{jt}$$

- En otras palabras, adicionando $\overline{X}_j$ a la ecuación para cada período temporal llegamos al modelo tradicional Probit con efectos aleatorios.
- Note que el efecto de asumir que la esperanza de los efectos no observables es lineal provoca el mismo efecto que en la corrección de Wooldridge en el caso de los paneles no balanceados. Allí aparecía en la ecuación de interés el término $\overline{X}_j\pi$ y ahora aparece $\overline{X}_j\xi$ debido al supuesto de Mundlak.

Modelo Probit de Efectos no Observables

- Dadas las estimaciones de ψ y ξ podemos estimar $E\left(c_j\right) = \psi + E\left(\overline{X}_j\right)\xi$ con $\hat{\psi} + \overline{X}\xi$
- Donde $\overline{X}$ es la media muestral de $\overline{X}_j$.
- Por lo tanto para cualquier vector X_t, podemos estimar la probabilidad de respuesta en $E(c_j)$ como: $\Phi\left(\hat{\psi} + X_t\hat{\beta} + \overline{X}\hat{\xi}\right)$
- Tomando derivadas con respecto a los elementos de X_t podemos estimar los efectos parciales de las variables sobre las probabilidades de respuesta.

Modelo Probit de Efectos no Observables

- Si solo asumimos (2) y (5):
  $P\left(y_{jt}=1 \mid X_j\right)=\Phi\left[\left(\psi+x_{jt}\beta+\overline{X}_j\xi\right)\left(1+\sigma_a^2\right)^{-1/2}\right] \equiv \Phi\left(\psi_a+x_{jt}\beta_a+\overline{X}_j\xi\right)$
- Donde el subíndice a indica que el vector de parámetros ha sido multiplicado por $\left(1+\sigma_c^2\right)^{-1/2}$.
- Se sigue inmediatamente que β_a, ψ_a, y ξ_a pueden estimarse consistentemente usando pooled Probit en y_{jt} sobre 1, x_{jt}, y $\overline{X}_j$.

Modelo Probit de Efectos no Observables

- Como los y_{jt} son dependientes (no estamos asumiendo (3)) necesitamos varianzas y covarianzas robustas.
- Una vez que estimamos β_a, ψ_a, y ξ_a pueden estimarse los efectos parciales promedios usando el promedio a través de j de: $\widehat{\beta}_{aj}\phi\left(\hat{\psi}_a + x^0\hat{\beta}_a + \hat{X}_j\hat{\xi}_a\right)$
- Donde x^0 es un vector no aleatorio de números que se eligen como valores interesantes de las variables explicativas.

Agenda

Modelo Logit de Efectos no Observables

- Reemplazando la función de distribución acumulada de la normal estándar por la de la logística en (2) tenemos:

$$P\left(y_{jt}=1 \mid X_j, c_j\right)=P\left(y_{jt}=1 \mid x_{jt}, c_j\right)=\Lambda\left(x_{jt}\beta+c_j\right), t=1,2,...,T \quad (6)$$

- Manteniendo los supuestos (3) y (4) llegamos a lo que se conoce como el modelo Logit de efectos aleatorios.
- Este modelo no es tan atractivo como el modelo Probit de efectos aleatorios porque no existen estimadores fáciles de obtener.
- El modelo Probit de efectos aleatorios utiliza el hecho de que la combinación de dos funciones normales es otra normal. En el caso del Logit, $P(y_{jt}=1 \mid Xj)$ no tiene una forma simple porque hay que integrar la función $\Lambda(X_t\beta+c)$ con respecto a la densidad de la normal $(1/\sigma_c)\phi(c/\sigma_c)$ lo que no da una forma funcional simple.

Modelo Logit de Efectos no Observables

- La mayor ventaja del modelo Logit de efectos no observables por sobre el Probit es que bajo los supuestos (6) y (3) es posible obtener estimaciones consistentes de β sin hacer supuestos acerca de la relación de c_j con x_{jt}.
- En el caso lineal, utilizamos la transformación de FE ó FD para eliminar c_j de la ecuación a estimar.
- En el caso del Logit, se puede emplear una estrategia similar. Lo que necesitamos hacer es encontrar la distribución conjunta de $y_j \equiv (y_{j1}, y_{j2}, ..., y_{jT})'$ condicional a X_j, c_j y $n_j \equiv \sum_{t=1}^{T} y_{jt}$.
- Se puede verificar que esta distribución no depende de c_j tal que queda la distribución de y_j condicionada a X_j y n_j.

Modelo Logit de Efectos no Observables

- Ilustremos el caso con $T = 2$, donde n_j puede adoptar los valores $0, 1, 2$. Intuitivamente la distribución condicional de $(y_{j1}, y_{j2})'$ dado n_j no puede ser informativa acerca de β cuando $n_j = 0$ ó $n_j = 2$ porque esos valores determinan completamente el resultado de y_j.
- Sin embargo para $n_j = 1$ tenemos:

$$P\left(y_{j2} = 1 \mid X_j, c_j, n_j = 1\right) = \frac{P\left(y_{j2} = 1, n_j = 1 \mid X_j, c_j\right)}{P\left(n_j = 1 \mid X_j, c_j\right)} =$$

$$\frac{P\left(y_{j2} = 1 \mid X_j, c_j\right) P\left(y_{j1} = 0 \mid X_j, c_j\right)}{\left\{P\left(y_{j1} = 0, y_{j2} = 1 \mid X_j, c_j\right) + P\left(y_{j1} = 1, y_{j2} = 0 \mid X_j, c_j\right)\right\}}$$

$$= \frac{\Lambda\left(x_{j2}\beta + c_j\right)\left[1 - \Lambda\left(x_{j1}\beta + c_j\right)\right]}{\left\{\left[1 - \Lambda\left(x_{j1}\beta + c_j\right)\right]\Lambda\left(x_{j2}\beta + c_j\right) + \Lambda\left(x_{j1}\beta + c_j\right)\left[1 - \Lambda\left(x_{j2}\beta + c_j\right)\right]\right\}}$$

$$= \Lambda\left[\left(x_{j2} - x_{j1}\right)\beta\right]$$

Modelo Logit de Efectos no Observables

- Similarmente, $P\left(y_{j1}=1 \mid X_j, c_j, n_j=1\right)=\Lambda\left[-\left(x_{j2}-x_{j1}\right)\beta\right]$ $=1-\Lambda\left[\left(x_{j2}-x_{j1}\right)\beta\right]$
- El logaritmo de la función de verosimilitud condicional para la observación j se puede escribir como:

$$\ell_j\left(\beta\right)=1\left[n_j=1\right]\left(w_j log \Lambda\left[\left(x_{j2}-x_{j1}\right)\beta\right]\right)+\left(1-w_j\right) log\left\{1-\Lambda\left[\left(x_{j2}-x_{j1}\right)\beta\right]\right\}$$

- Donde $w_j=1$ si $\left(y_{j1}=0, y_{j2}=1\right)$ y $w_j=0$ si $\left(y_{j1}=1, y_{j2}=0\right)$
- El estimador de máxima verosimilitud condicional se obtiene maximizando la suma de $\ell_j\left(\beta\right)$ sobre j.
- La función indicador $1[n_j=1]$ selecciona las observaciones para las que $n_j=1$.
- Note que la función de verosimilitud anterior es la función estándar del Logit de corte transversal para una regresión de w_j sobre $x_{j2}-x_{j1}$ utilizando las observaciones para las que $n_j=1$.

Modelo Logit de Efectos no Observables

- El estimador de máxima verosimilitud condicional que se obtiene en este caso se denomina estimador Logit de efectos fijos.
- Lo que hace este tipo de estimación es simplemente encontrar la distribución condicional, que describe la subpoblación con $n_j = 1$, que depende solo de los datos observados y de β.
- Para el caso de T general, la función de verosimilitud es un poco más complicada pero tiene un tratamiento posible.
- Primero, se tiene que:
  $P\left(y_{j1} = y_1, ..., y_{jT} = y_T \mid x_J, c_j, n_j = n\right) = \frac{P\left(y_{j1}=y_1,...,y_{jT}=y_T | X_j, c_j\right)}{P\left(n_j = n | X_j, c_j\right)}$
- Y el numerador se puede expresar como:
  $P\left(y_{j1} = y_1 \mid X_j, c_j\right) P\left(y_{j2} = y_2 \mid X_j, c_j\right) ... P\left(y_{jT} = y_T \mid X_j, c_j\right)$
- Usando el supuesto (3). El denominador es la parte complicada pero es fácil de describir.

Modelo Logit de Efectos no Observables

- $P(n_j = n \mid X_j, c_j)$ es la suma de las probabilidades de todos los posibles resultados de y_j tal que $n_j = n$.
- Usando la forma específica de la función logística podemos escribir:

$$\ell_j(\beta) = log\left\{exp\left(\sum_{t=1}^{T} y_{it}x_{jt}\beta\right)\left[\sum_{a\in R_j} exp\left(a_t x_{jt}\beta\right)\right]^{-1}\right\}$$

- Donde R_j es un subconjunto de R^T definido como $\left\{a \in R^T : a_t \epsilon \{0,1\} \text{ y } \sum_{t=1}^{T} a_t = n_j\right\}$.
- El logaritmo de la función de verosimilitud anterior debe sumarse a través de j para luego maximizarse y obtener estimadores de β consistentes y asintóticamente normales.
- El estimador Logit de efectos fijos de β nos da el efecto de cada elemento de X_t en el logaritmo de la tasa de probabilidad.

Modelo Logit de Efectos no Observables

- Esto es,

$$log\left\{\frac{\Lambda\left(X_t\beta+c\right)}{1-\Lambda\left(X_t\beta+c\right)}\right\}=X_t\beta+c$$

- Desafortunadamente, no se puede estimar el efecto parcial sobre la probabilidad de respuesta a menos que insertemos el valor de c.
- Como la distribución de c_j no esta especificada es difícil saber que valor insertar.
- Además, tampoco se pueden calcular los efectos parciales promedio porque requerirían encontrar $E[\Lambda(Xt\beta+c)]$ que necesita una especificación de la distribución de c_j.

Modelos Dinámicos con Efectos no Observables

- Supongamos que la primera observación es con $t = 0$, de forma tal que y_{j0} es la primera observación de la variable y. Para $t = 1, 2, ..., T$ estamos interesados el el modelo dinámico:

$$P\left(y_{jt} = 1 \mid y_{j,t-1}, ..., y_{j0}, Z_j, c_j\right) = G\left(z_{jt}\delta + \rho y_{j,t-1} + c_j\right)$$

- Donde z_{jt} es un vector de variables explicativas contemporáneas, $Z_j = (z_{j1}, \ldots, z_{jT})$ y G puede ser la función Probit o Logit.
- Los puntos importantes del modelo anterior son:
- Primero, los z_{jt} se asume que satisfacen un supuesto de exogeneidad estricta (condicional en c_j) dado que Z_j aparece en el conjunto de condiciones del lado izquierdo de la ecuación pero solo aparece z_{jt} en el lado derecho.
- Segundo, la probabilidad de éxito en t depende del resultado en $t - 1$ y de los efectos no observables, c_j.

Modelos Dinámicos con Efectos no Observables

- Como podríamos estimar δ y ρ en la ecuación anterior?
- Primero podemos construir la función de probabilidad conjunta de las observaciones de la muestra: $f\left(y_1, y_2, ..., y_T \mid y_0, z, c; \beta\right)$

$$= \prod_{t=1}^{T} f\left(y_t \mid y_{t-1}, ..., y_1, y_0, z_t, c; \beta\right)$$

$$= \prod_{t=1}^{T} G\left(z_t\delta + \rho y_{t-1} + c\right)^{y_t}\left[1 - G\left(z_t\delta + \rho y_{t-1} + c\right)\right]^{1-y_t} (D.1)$$

- El problema es que con T fijo, debido a la presencia del efecto no observable c, no se puede construir la función de verosimilitud que nos permita estimar β consistentemente.
- Tal como pasaba en el caso con variables estríctamente exógenas, tratar a los c_j como parámetros a ser estimados no permite obtener estimadores consistentes de δ y ρ cuando $N \to \infty$

Modelos Dinámicos con Efectos no Observables

- Lo que uno debe hacer en este caso es eliminar c vía integración como hicimos antes. El problema ahora es que para hacer esto hay que realizar algún supuesto acerca del comportamiento de la primera observación de y, y_{j0}.
- Una forma de proceder es obtener la distribución conjunta de $(y_{j1}, y_{j2}, ..., y_{jT})$ condicional en (y_{j0}, Z_j).
- Para obtener $f(y_1, y_2, ..., y_T \mid y_0, z, \theta)$ necesitamos la densidad de c_j dado (y_{j0}, Z_j)
- Dada una densidad $h(c \mid y_0, z; \gamma)$ tenemos $f(y_1, y_2, ..., y_T \mid y_0, z, \theta)$

$$= \int_{-\infty}^{\infty} f(y_1, y_2, ..., y_T \mid y_0, z, c; \beta)\, h(c \mid y_0, z; \gamma)\, dc$$

- La integral puede ser reemplazada por una suma ponderada si la distribución de c es discreta.

- Cuando $G = \Phi$ en el modelo (D.1) una elección conveniente para $h\left(c \mid y_0, z; \gamma\right)$ es: $Normal\left(\psi + \xi_0 y_{j0} + Z_j \xi, \sigma_a^2\right)$, que sigue de escribir $c_j = \psi + \xi_0 y_{j0} + Z_j \xi + a_j$, donde $a_j \sim Normal(0, \sigma_a^2)$ e independiente de (y_{j0}, Z_j).
- Por lo tanto se puede escribir,

$$y_{jt} = 1\left[\psi + z_{jt}\delta + \rho y_{j,t-1} + \xi_0 y_{j0} + Z_j \xi + a_j + e_{jt} > 0\right]$$

- Tal que y_{jt} dado $(y_{j,t-1}, ..., y_{j0}, Z_j, a_j)$ sigue un modelo Probit.

Modelos Dinámicos con Efectos no Observables

- Por lo tanto, la densidad de $(y_{j1}, ..., y_{jT})$ dado (y_{j0}, Z_j) tiene la siguiente forma:
  $L(c_j, \beta) =$

$$\int_{-\infty}^{\infty} \left[\prod_{j=1}^{N} \prod_{t=1}^{T} \Phi\left(x_{jt}\beta + c_j\right)^{y_{jt}} \left[1 - \Phi\left(x_{jt}\beta + c_j\right)\right]^{1-y_{jt}} \right]$$

$$* \left(1/\sigma_c\right) \phi\left(c/\sigma_c\right) dc$$

- Donde $x_{jt} = (1, z_{jt}, y_{j,t-1}, y_{j0}, Z_j)$ y con a y σ_a reemplazando a c y σ_c.
- Este resultado implica que podemos utilizar el software del Probit de efectos aleatorios para estimar ψ, δ, ρ, ξ_0, ξ, y σ_c^2.

Supuestos

- Se tienen los siguientes supuestos:

$$P\left(y_{jt}=1 \mid X_j, c_j\right)=P(y_{jt}=1 \mid x_{jt}, c_j)$$

$$=\Phi\left(x_{jt}\beta+c_j\right), t=1,2,...,T. \qquad (2)$$

$$y_{j1}, y_{j2}, ..., y_{jT} \text{ son independientes condicionando en}(X_j, c_j). \qquad (3)$$

$$c_j \mid X_j \sim \text{Normal}(0, \sigma_c^2) \qquad (4)$$

Back

Supuestos

- Si escribimos el supuesto (2) en forma de variable latente tenemos:
- $y_{jt}^* = x_{jt}\beta + c_j + u_{jt}$, con $y_{jt} = 1\left[y_{jt}^* > 0\right]$ y $u_{jt} \mid X_j, c_j \sim N(0,1)$. Además se cumple (4).
- Con estos supuestos $c_j + u_{jt} \sim N\left(0, 1 + \sigma_c^2\right)$.
- Por lo tanto: $P\left(y_{jt} = 1 \mid X_j\right) = P\left(c_j + u_{jt} > -x_{jt}\beta \mid X_j\right)$
  $= \Phi\left[x_{jt}\beta / \left(1 + \sigma_c^2\right)^{1/2}\right]$, $t = 1, 2, ..., T$.
- Se sigue inmediatamente de la ecuación anterior que pooled Probit de y sobre X estima consistentemente:

$$\beta_c = \beta / \left(1 + \sigma_c^2\right)^{1/2}$$

Back

References

- Chamberlain, G (1980) "Analysis of covariance with qualitative data," Review of Economic Studies, 47, pp. 225-238.
- Mundlak, Y. (1978) "On the pooling of time series and cross section data," Econometrica, 46, 69-85.

Corte Transversal

- Una variable binaria, es una variable que solo adopta dos valores (que por convención se denotan 0 y 1).
- Por ejemplo, la empresa Sony quiere saber si la venta de sus televisiones esta relacionada con el ingreso de las personas. Para esto, toma una muestra de 100 personas que en el último semestre compró un televisor y registra los valores de las siguientes variables:
- $Y = 1$ si la persona compró un televisor Sony.
- $Y = 0$ si la persona compró otra marca de TV.
- Además, registra para cada persona en la muestra el valor del salario horario.
- La empresa especifica la siguiente relación funcional:

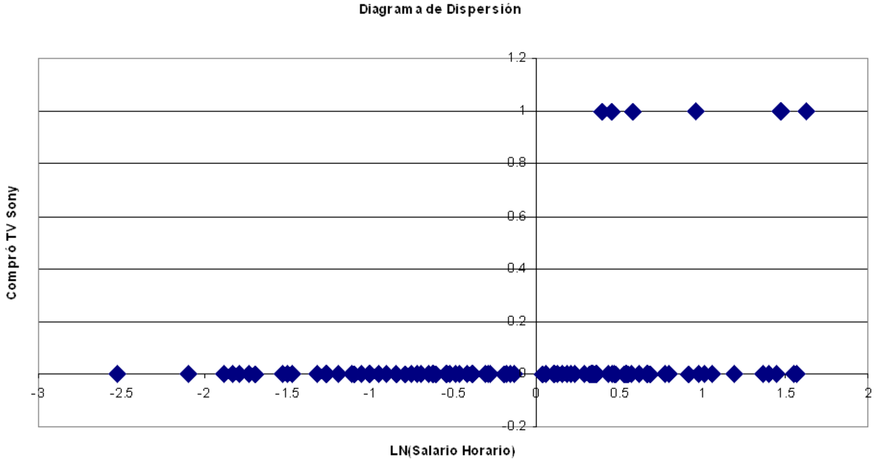
$$Y_i = \alpha + \beta X_i + \varepsilon_i; i = 1, 2, ..., 100.$$

Corte Transversal

- Donde X_i representa el logaritmo natural del salario horario.
- La idea de la firma es que si pudiera saber los valores de los parámetros del modelo entonces podría conocer si existe una asociación directa o inversa entre lo que gana la gente y la compra de sus televisores.
- Como primera aproximación al tema veamos un diagrama de dispersión de las variables.

Corte Transversal

- Se tiene el siguiente diagrama de dispersión:

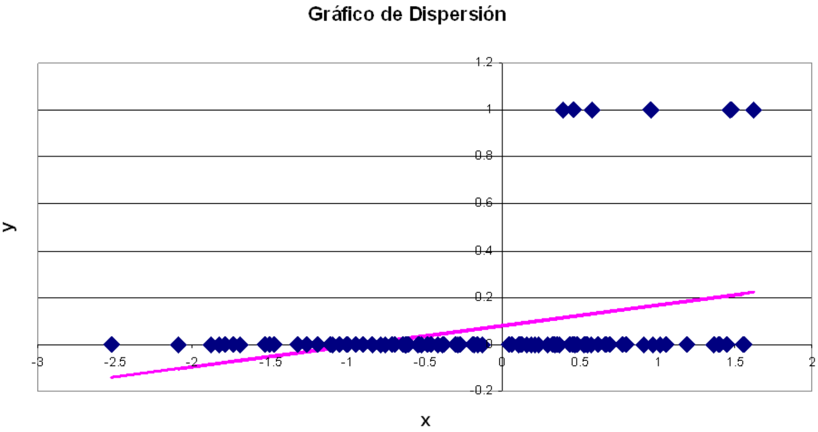


Corte Transversal

- Observando el gráfico de dispersión la empresa piensa que si bien hay evidencia de que la gente que compra sus televisores en general está ubicada en el sector de mayores ingresos necesita mayor información para poder segmentar el mercado.
- Esto es, si pudiera estimar los parámetros del modelo quizás pudiera distinguir a un nuevo cliente potencial.
- Utilizando el método de mínimos cuadrados clásicos, la empresa estima el modelo y obtiene los siguientes resultados: $\hat{Y}_i = 0.08 + 0.087X_i$, $i = 1, 2, ..., 100$, con valores t 3.2 y 3.4 respectivamente.

Modelos de Respuesta Binaria

- Esto es, existe una relación directa (positiva) entre el salario y la compra de televisores Sony como se apreciaba en el gráfico de dispersión.



Corte Transversal

- Como puede observarse claramente en el gráfico anterior, el problema con esta estimación es que hay muchos valores ajustados para la variable dependiente que son NEGATIVOS!
- Esta es una consecuencia del método de estimación elegido, ya que MCC no contiene ninguna restricción que nos diga que la estimación de la variable dependiente siempre tiene que ser cero o uno.
- El problema anterior puede re-plantearse de la siguiente manera. Cuál es la probabilidad de que una persona compre un televisor Sony, dado el valor de su salario horario?
- Evidentemente, en la muestra de 100 personas es poco probable que alguien pueda responder a esa pregunta. Sin embargo, podemos observar el valor de esta probabilidad ex—post.

Corte Transversal

- Es decir, una persona ex - post o compró el televisor Sony (probabilidad de realización igual a uno) o no lo compró (probabilidad de realización igual a cero).
- Esto significa que lo que observamos es la realización de una variable no observable (la probabilidad de compra del televisor Sony).
- Observe que si la empresa pudiera conocer la probabilidad ex – ante, entonces podría segmentar el mercado con esta probabilidad.
- Es decir que a nosotros nos interesa el valor de la probabilidad ex – ante. Esto es: $Pr\left(y_i = 1 \mid X_i\right)$
- Notemos que de acuerdo al modelo especificado anteriormente tenemos: $Y_i = \alpha + \beta X_i + \varepsilon_i$

Corte Transversal

- Por lo tanto tenemos que: (1) $E\left(Y_i \mid X_i\right) = \alpha + \beta X_i + E\left(\varepsilon_i \mid X_i\right) = \alpha + \beta X_i$
- Además, sabemos que por definición de esperanza matemática, la esperanza matemática condicional de una variable es la suma de cada uno de los valores que adopta la variable multiplicados por su probabilidad de ocurrencia.
- En este caso se tiene que: (2) $E\left(Y_i \mid X_i\right) = 1 * P\left(Y_i = 1 \mid X_i\right) + 0 * P\left(Y_i = 0 \mid X_i\right)$
- $= P(Y_i = 1 \mid X_i)$
- Igualando las ecuaciones (1) y (2) tenemos: (3) $P(Y_i = 1 \mid X_i) = \alpha + \beta X_i$

Corte Transversal

- Como puede observarse en la ecuación (3) la probabilidad condicional de que el evento Yi ocurra (en este caso la compra de la televisión Sony) dado que conocemos el valor de Xi esta expresada como una relación lineal.
- Debido a este hecho, los modelos de variable dependiente binaria reciben el nombre de Modelos de Probabilidad Lineal.
- Dado que Yi solo puede adoptar dos valores (cero ó uno) podemos obtener la distribución de probabilidad de la variable Y_i ilustrada con la siguiente tabla a continuación:

Yi	Pr(Yi | Xi)
1	$\alpha + \beta Xi$
0	$1 - (\alpha + \beta Xi)$

Corte Transversal

- Dada la distribución de la variable dependiente podemos utilizar el modelo para obtener la distribución de los errores.
- Esto es, como $Y_i = \alpha + \beta X_i + \varepsilon_i$, cuando:
- $y_i = 1 \rightarrow \varepsilon_i = 1 - \alpha - \beta X_i$ y cuando
- $y_i = 0 \rightarrow \varepsilon_i = -\alpha - \beta X_i$.
- Por lo tanto se tiene la siguiente tabla:

ε_i	Pr(Yi | Xi)
1 - α - β Xi	α + β Xi
- α - β Xi	1 - (α + β Xi)

- Cuáles son los momentos de esta distribución?
$E(\varepsilon_i \mid X_i) = [1 - (\alpha + \beta X_i)]\,(\alpha + \beta X_i) - (\alpha + \beta X_i)\,[1 - (\alpha + \beta X_i)] = 0$

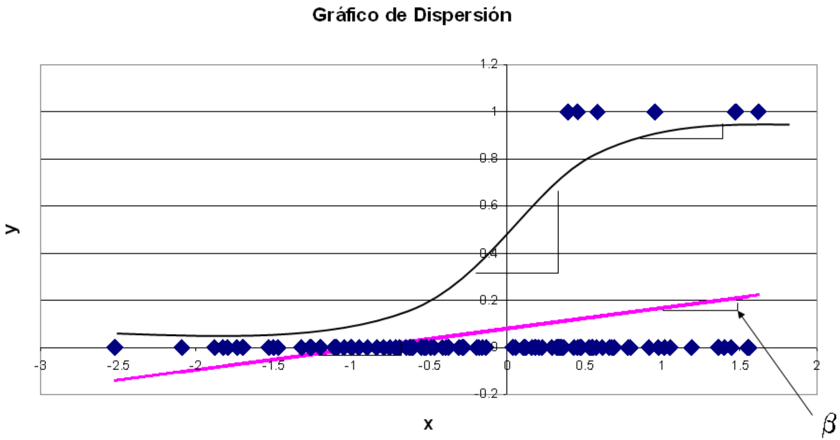
Corte Transversal

- Y por otro lado, la varianza: $Var\left(\varepsilon_i \mid X_i\right) = E\left(\varepsilon_i^2 \mid X_i\right) - E\left(\varepsilon_i \mid X_i\right)^2 =$ $E\left(\varepsilon_i^2 \mid X_i\right) = [1 - (\alpha + \beta X_i)] * (\alpha + \beta X_i)$
- Entonces, cuando la variable dependiente es binaria, el modelo tiene heterocedasticidad.
- Esto puede generalizarse a cualquier modelo que tenga por variable dependiente una variable categórica.
- Ahora podemos resumir las características de los modelos de probabilidad lineal (MPL):
  1. Estos modelos reciben este nombre porque la variable dependiente puede interpretarse como una probabilidad.
  2. Los valores estimados, por el método de MCC, de la variable dependiente pueden caer fuera del rango [0, 1] lo cual hace que la estimación sea pobre. .
  3. Los errores del modelo son heterocedásticos por lo tanto MCC nos dará estimadores ineficientes

Corte Transversal

- Otro de los problemas que sufren los modelos de probabilidad lineal es el de la interpretación.
- En nuestro caso particular el coeficiente β mide cuanto afecta a la probabilidad de comprar un televisor Sony un cambio en el salario de las personas.
- Económicamente uno pensaría que este efecto debiera ser pequeño para aquellos que ganan muy poco o ganan mucho y debiera ser más grande para el resto.
- Sin embargo, en el MPL el efecto es constante e igual a β lo que no coincide con lo que se desprende de la teoría económica.
- Cómo podríamos resolver este problema de interpretación y los problemas estadísticos?

- Lo que necesitamos es una función de probabilidad que no sea lineal. Esto es:



Corte Transversal

- Las características de la curva de la figura anterior resuelve nuestros problemas ya que:
    1. Empieza en cero y termina en uno. Esto es, solo adopta valores en el intervalo $[0, 1]$.
    2. Tiene diferentes pendientes en distintos puntos. Para valores muy pequeños y muy grandes de X_i la pendiente es chica y para valores intermedios de X_i la pendiente es más grande.
- Cualquier curva de probabilidad acumulada cumple con las características antes mencionadas.
- Por lo tanto, uno puede especificar el modelo utilizando la función de probabilidad acumulada de cualquier distribución.
- Las funciones más utilizadas son la distribución Normal y la distribución Logística. Back