## Capítulo 2.- Metodología

En la presente investigación, se contrastará el modelo propuesto en la investigación de Marisa Bucheli y Carlos Casacubierta, "Asistencia escolar y Participación en el mercado de trabajo de los adolescentes en Uruguay", mencionado en el capítulo anterior. En dicho trabajo se examina los determinantes de las decisiones de inversión en educación, estimándose un modelo probit para analizar el efecto de un conjunto de variables explicativas en la probabilidad de asistencia para Uruguay. Además, se considera un modelo de la decisión conjunta de asistencia escolar y participación en el mercado de trabajo, esto debido a que se encontró que en ese país la gran mayoría de los adolescentes o estudia o participa en el mercado de trabajo, por lo que se hace el supuesto de que la participación en el mercado de trabajo es opcional a la asistencia de enseñanza, de manera que ambos estados son excluyentes y resultado del mismo proceso de decisión.

### 2.1 Modelos probit.

En la econometría, el modelo de probabilidad lineal, es sencillo de estimar y de aplicar, pero posee ciertos inconvenientes. Las dos desventajas más importantes son que las probabilidades ajustadas pueden ser menores a cero o mayores que uno y que el efecto parcial de cualquier variable explicativa (que aparezca a su nivel) es constante (Wooldrige, 2001). Estas limitaciones del modelo de probabilidad lineal se superan mediante modelos de respuesta binaria más complejos.

En un modelo de respuesta binaria el interés descansa principalmente en la probabilidad de respuesta:

$$\mathbf{P\ (y=1|x)=P\ (y=1|x_1,\ x_2\ldots\ldots x_k)}$$

Donde x denota el conjunto de variables explicativas.

En el modelo de probabilidad lineal, suponemos que la probabilidad de respuesta es lineal en el conjunto de parámetros. Para evitar las limitaciones mencionadas del modelo de probabilidad lineal, se considera una clase de modelos de respuesta binaria de la siguiente forma:

$$\mathbf{P\ (y=1|x)=G(\beta_0+\beta_1 x_1+\ldots.\beta_k x_k)=G(\beta_0+x\beta)}$$

Donde G es una función que asume valores que se hallan estrictamente entre cero y uno $0<G(z)<1$, para todos los números reales (z). Lo anterior asegura que las probabilidades de respuesta se hallen estrictamente entre cero y uno.

En el modelo probit, G es la función de distribución acumulada de la normal estándar, que se expresa como una integral:

$$\mathbf{G\ (z)=\Phi\ (z) \equiv \int_{\infty}^{z}\ \varphi\ (v)\ dv}$$

Donde $\varphi(z)$ es la densidad de la normal estándar

$$\varphi (z)= (2\pi)^{-1/3}\exp (-z^2/2)$$

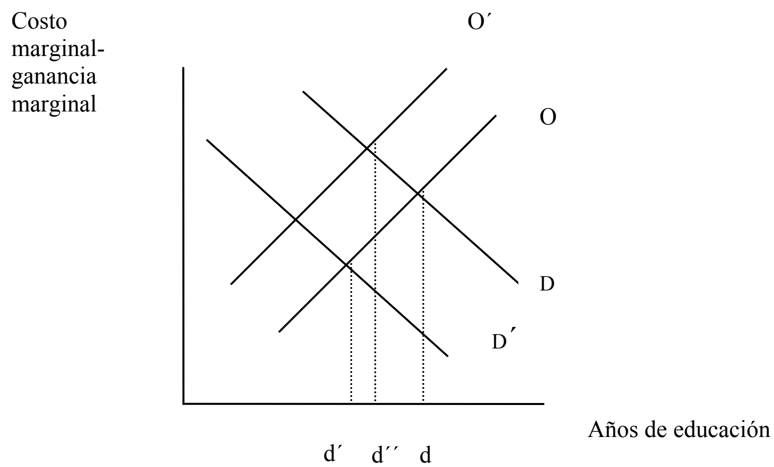
**2.2 Modelo probit en el artículo de Casacubierta y Bucheli.**

Como se mencionó en el capítulo anterior, en los modelos estándares de la teoría de la inversión en capital humano, la razón fundamental para estudiar, es aumentar los ingresos futuros y, por lo tanto, incrementar la riqueza total. En este contexto el número óptimo de años de educación de cada individuo será el resultado de comparar el beneficio y costo marginales asociados a dicha inversión. (Casacubierta, Bucheli, 2000).

El beneficio marginal está representado por la tasa de rendimiento asociada a cada monto de inversión. Se supone que por lo general las personas invierten considerando que las rentabilidades se incrementarán en el futuro.

La decisión individual del número óptimo de años de escolaridad es resultado de la interacción de la curva de demanda de educación y la oferta de fondos:

**Demanda por educación y oferta de fondos**




Este modelo puede ser utilizado para identificar las variables que explican las diferencias individuales en el capital educativo acumulado y por lo tanto la asistencia al sistema de enseñanza, por medio de su incidencia en la demanda por educación (ganancia marginal) y en la oferta de fondos (costo marginal). (Casacubierta, Bucheli, 2000).

Los individuos que esperan obtener mayores rentabilidades demandarán más capital educativo. La curva de demanda de los individuos con mayor rentabilidad está por encima, por lo que, para una misma curva de de oferta de fondos, su nivel óptimo de educación será mayor (en la gráfica, d versus d´). Por otro lado, la oferta de fondos, se puede ver, entre otras cosas con la disposición de becas, incluso las preferencias de los individuos y el ingreso del hogar, controlado por su tamaño, cuanto mayor es el número de personas que comparten gastos, menor es la cantidad disponible. Así las curvas de oferta de los individuos más pobres se sitúan más a la izquierda, dando lugar a un nivel óptimo de capital educativo inferior (en la gráfica, d´´ versus d).

En este contexto, en el trabajo mencionado, se dice que para cada individuo i puede estimarse $d_i$ suponiendo una relación lineal con las variables como el entorno familiar, la aptitud, la escolaridad de los adultos, el grado educativo alcanzado por el joven, etc. A partir de esta expresión lineal del grado óptimo de educación, puede deducirse otro modelo que distingue la decisión de asistir o no al sistema de enseñanza.

Para los desertores del sistema educativo, se cumple que ya han alcanzado el óptimo de educación. Sin embargo, esto no sucede con las personas que aún asisten. En estas condiciones, la variable $d_{i,}$ es importante, y que a partir de ésta se puede construirse otra variable, $y_i$ que toma valor de 1 si $d_i$ es mayor que el número de años de educación aprobados, en otras palabras, si el joven asiste al sistema de enseñanza, y el valor 0 en caso contrario. Suponiendo que el término aleatorio en la ecuación especificada para $d_i$ sigue una distribución normal. Esta especificación es un modelo probit de la probabilidad de asistir al sistema educativo condicional a un conjunto de variables explicativas. (Casacubierta, Bucheli, 2000).

En el modelo probit la probabilidad de que la variable binaria y tome el valor de 1, condicional a un vector de variables x, queda caracterizada por la ecuación siguiente:

$$\mathbf{P\ (y = 1/x) = \Phi\ (\beta'x)}$$

En la que φ(.) es la función de distribución normal estándar.

Este modelo se  estimó con un primer grupo de variables  relacionadas con el hogar. Como medida de los fondos para el financiamiento de la educación, se incluye el logaritmo del ingreso per capita excluidas las remuneraciones del adolescente. Por otro lado como indicadores del entorno familiar, se utiliza los años de escolaridad del jefe del hogar.. Un segundo grupo de variables refleja un conjunto de variables  individuales: el sexo, que toma valor de 1 para los varones y 0 para las mujeres, la edad y los años de educación aprobados.  Por último se incluye una variable dicotómica que recoge el lugar de residencia, la cual  toma valor  de 1 para quienes viven en la capital y 0 en caso contrario.

Aparte de la estimación de los coeficientes  de este modelo,  se procede a calcular los efectos marginales de cada variable, los cuales miden la repercusión de sus cambios en la probabilidad de asistir a la enseñanza.

El efecto marginal de una variable $x_k$ se obtiene a partir del cálculo  de la siguiente expresión, en la que el resto de las variables son medidas en sus valores promedio:

$$\frac{\partial P(y=1/x)}{\partial x_k} = \frac{\partial \phi(\beta' x)}{\partial x_k} = \varphi(\beta' x)\beta_k$$

en la que Φ y φ corresponden a las funciones de distribución y de densidad normal con media 0 y varianza 1. A su vez se calcula el efecto discreto de cambios en algunas de las variables. Para ello utilizando los parámetros estimados, se calcula la probabilidad de asistencia para distintos valores  de las variables explicativas.

### 2.2.1 Estimación del modelo de decisión conjunta de asistir al sistema de enseñanza y participar en el mercado de trabajo.

En esta investigación se incluye, además, un modelo probit bivariado, debido a que este explica mejor la situación de dos decisiones conjuntas, pero no independientes, mediante el sistema de dos ecuaciones. Cada una de ellas comprende una variable binaria ($y_a$, $y_p$) de manera que:

$Z_{Ai} = \beta_A X_{Ai} + \varepsilon_{Ai}$, con $y_{Ai}=1$ si $z_{Ai}>0$; $y_{Ai}=0$ en otro caso

$Z_{pi} = \beta_p X_{pi} + \varepsilon_{pi}$, con $y_{pi}=1$ si $z_{pi}>0$; $y_{pi}=0$ en otro caso

Las variables $Z_{Ai}$ y $Z_{pi}$ son no observables pero a partir de ellas puede construirse dos observables: $y_{Ai}$, que toma valor de 1 cuando el adolescente asiste al sistema de enseñanza y $y_{pi}$, con el valor 1 cuando participa en el mercado de trabajo. A su vez, $X_{Ai}$ y $X_{pi}$ son los vectores de las variables que afectan cada una de dichas decisiones. De acuerdo con esta especificación, los residuos de ambas ecuaciones están correlacionados, estipulándose que el vector ($\varepsilon_{Ai}$ ,$\varepsilon_{pi}$) tiene una distribución normal bivariada (0,0,1,1,p). En el modelo probit bivariado, las probabilidades conjuntas del tipo P ($y_A$=1,$y_p$=1|x) se definen como:

$$P\ (y_A=1, y_p=1|x)=\Phi(x'_A\beta_A,\ x'_p\beta p,\ \rho)$$

En la que $\Phi$ representa la función de distribución normal bivariada con parámetros (0,0,1,1,$\rho$).