

## Taller de Econometría Aplicada I

### *Taller 12*

Para esta práctica utilizará el archivo **hrs.dta** disponible en el sitio Canvas del curso, que proviene del quinto levantamiento de la *Health and Retirement Study* en EUA con una muestra limitada a todos aquellos beneficiarios del seguro médico federal.

Utilizaremos estos datos para analizar la compra de seguros privados **ins** de cualquier fuente. Las variables explicativas incluyen el estado de salud, las características socioeconómicas y la información relacionada con el cónyuge del individuo. La información sobre el estado de salud autorreportada se utiliza para generar la variable **hstatusg** que mide si el estado de salud es bueno, muy bueno o excelente. Otras medidas del estado de salud del individuo son el número de limitaciones (hasta cinco) en las actividades de la vida diaria **adl** y el número total de condiciones crónicas **chronic**.

Las variables socioeconómicas utilizadas son edad, género, raza, etnia, estado civil, años de educación y estado de jubilación (respectivamente **age**, **female**, **white**, **hisp**, **married**, **educyear**, **retire**); así como el ingreso familiar **hhincome**. El estado de jubilación del cónyuge **sretire** es una variable dicotómica que toma el valor de 1 si el individuo tiene un cónyuge jubilado.

Deberá entregar un archivo **do** con las funciones que permitan obtener los siguientes resultados así como sus comentarios con la interpretación de los mismos.

- 1) **Lectura de los datos.** Lea el conjunto de datos y vea su descripción mediante la función **describe**. Se sugiere definir los conjuntos de variables que vaya a utilizar con el comando **global**, por ejemplo, la lista de regresores x sociodemográficos puede ser definida como:

```
global xlist age hstatusg hhincome educyear married hisp
```

y utilizada posteriormente sólo como `xlist`.

- 2) **Descripción de los datos.** Utilice las funciones `summarize` e `inspect` para realizar una breve descripción de los datos, incluyendo aquellas variables cualitativas.
- 3) **Modelo con un sólo regresor.** Utilice la variable dependiente `ins` y `educyear` como única regresora. Estime los parámetros de los modelos logit, probit y lineal de probabilidad mediante MCO, utilice los errores estándar predeterminados y los robustos. Compare los coeficientes obtenidos entre los modelos y sus errores estandar, así como la conclusión de su significancia. Para cada modelo, calcule el efecto marginal de un año más de educación para alguien con una muestra de años de educación promedio. ¿Qué modelo se ajusta mejor a los datos?
- 4) **Modelo con varios regresores.** Utilice la variable dependiente `ins` pero ahora con la variable `retire` y la lista de regresoras `xlist` definida como antes. Estime de nuevo los tres modelos tanto con errores estandar predeterminados como robustos y con el comando `estimates table` construya una tabla que permita comparar los resultados. Interprete sus coeficientes y significancias. ¿Qué factores influyen más? ¿Qué modelo se ajusta mejor a la teoría económica? Verifique que se cumplen las siguientes reglas aproximadas entre los coeficientes:

$$\begin{aligned}\hat{\beta}_{logit} &\approx 4\hat{\beta}_{MCO} \\ \hat{\beta}_{probit} &\approx 2.5\hat{\beta}_{MCO} \\ \hat{\beta}_{logit} &\approx 1.6\hat{\beta}_{probit}\end{aligned}$$

- 5) **Pruebas de hipótesis.** Genere otras variables como `age2`, `agefem`, `agechr`, and `agewhi` dadas por el cuadrado de la edad y la interacción de edad con `female`, `chronic`, `white`. Estime de nuevo uno de los modelos anteriores (logit o probit) con estas nuevas regresoras y realice la prueba de significancia para ver si es necesario incluirlas en el modelo mediante el comando `test`. ¿Considera que estos efectos de interacción son necesarios?
- 6) **Selección del modelo.** Utilice los siguientes criterios para determinar cuál modelo ajusta mejor a los datos: logit o probit:

- a) Significancia de los coeficientes estimados
- b) Criterios de información AIC y BIC
- c) Valor de log-likelihood (sólo por tener el mismo número de regresores)
- d) pseudo  $R^2$  de MacFadden
- e) Comparación de predicción de probabilidades con frecuencias (utilice el comando `estat gof`)
- f) Comparación de predicción de probabilidades con valores reales (utilice el comando `estat classification`)

¿Qué modelo es el que considera logra un mejor ajuste?

- 7) **Predicción de probabilidades.** Utilizando el modelo elegido, considere los siguientes valores condicionales `age=65`, `retire=1`, `hstatusg=1`, `hhincome=60`, `educyear=17`, `married=1`, `hisp=0` para predecir con qué probabilidad un individuo tendría un seguro privado (utilice un intervalo de confianza al 95 % para su estimación). Obtenga y compare el efecto marginal de `age` sobre  $Pr(ins = 1|x)$  utilizando los comandos `mf` y `prchange`. Los resultados deberían ser equivalentes.