Economía Aplicada

Secciones Cruzadas Repetidas o Datos Fusionados

Departamento de Economía
Universidad Carlos III de Madrid

Basado en Wooldridge cap.13

Dos Ejemplos de Secciones Cruzadas Repetidas

Encuestas CIS - España

- cada mes de manera independiente, el Centro de Investigaciones Sociológicas toma una muestra de la población española.
- la encuesta pregunta sobre la actualidad política, económica y social.

Current Population Survey (CPS) - EEUU

- cada mes de manera independiente, la oficina de estadísticas laborales de EEUU (the US Bureau of Labor Statistics) entrevista a aproximadamente 60.000 hogares.
- el objetivo es conocer la situación laboral en EEUU. Se reporta una estimación del número de empleados y desempleados cada mes.

- Datos con estas estructuras se conocen como secciones cruzadas repetidas (o fusiones de secciones cruzadas independientes).

Estimación con Secciones Cruzadas Repetidas

- Una razón para fusionar secciones cruzadas es tener una muestra más grande: estimadores más precisos y mayor potencia en los contrastes.
- Pero necesitamos realizar supuestos sobre el valor de los parámetros en cada período. Será útil fusionar datos si la relación que nos interesa permanece constante en el periodo de análisis.
- Ejemplo: Queremos analizar los efectos de la educación en el salario y contamos con datos fusionados de 1978 y 1985 (base cps78_85.gdt de Wooldridge en `gretl`).

Fusionando datos

- Suponiendo que los parámetros permanecen constantes podemos estimar usando MCO la siguiente ecuación:

$$log(w_{it}) = \beta_0 + \beta_1 educ_{it} + u_{it}, \quad t = 1978, 1985$$

Obteniendo:

$$\widehat{log(w_{it})} = \underset{(0.071)}{0.984} + \underset{(0.006)}{0.069}\, educ_{it}$$

El retorno a la educación estimado es de casi 7% en ambos años.

- El modelo supone que el efecto es constante en el tiempo.
- Alternativamente podemos flexibilizar el modelo a estimar, por ejemplo suponiendo que solamente la ordenada al origen cambia.

Un poco más flexible: modelo

- Esto se puede hacer incluyendo una variable binaria para cada período, excepto uno para evitar multicolinealidad perfecta. Normalmente se excluye el primer período, el período base.
- En nuestro ejemplo incluiremos una variable binaria para el año 1985: en la base $y85$ vale 1 si $t = 1985$ y 0 si $t = 1978$:

$$log(w_{it}) = \beta_0 + \beta_1 educ_{it} + \delta_1 y85_t + u_{it}$$

- En este modelo se permite variación en el tiempo para la variable dependiente pero el efecto de la educación se sigue suponiendo constante en el tiempo.

Un poco más flexible: estimación

- Los coeficientes estimados para las variables binarias temporales muestran la evolución de la variable dependiente con respecto al período base.
- La estimación MCO es:

$$\widehat{log(w)} = \underset{(0.068)}{0.886} + \underset{(0.005)}{0.063}\, educ + \underset{(0.029)}{0.348}\, y85$$

- Con $\widehat{\delta_1}$ se captura el aumento salarial entre 1978 y 1985, si mantenemos fijo el nivel educativo. El salario creció en promedio 34,8% entre 1985 y 1978, manteniendo el nivel educativo constante.
- El retorno a la educación estimado es de 6,3%. Y es por supuesto el mismo para ambos años.

Más flexible: variaciones temporales en los efectos

- También podemos investigar si el efecto de interés ha cambiado en el tiempo.
- Esto se puede hacer incluyendo interacciones entre la variable de interés y las variables ficticias anuales:
- Primero debemos generar variables que interactúan *educ* con cada una de las variables artificiales anuales ($educ \times y78$ y $educ \times y85$).
- Luego dos alternativas equivalentes:

Alternativa 1

- Correr MCO de $log(w)$ en una constante y en todas las interacciones de las variables binarias temporales con *educ*: cada pendiente mide los retornos a la educación cada período:

$$log(w_{it}) = \gamma_0 + \gamma_1 educ_{it} \times y78 + \gamma_2 educ_{it} \times y85_t + \gamma_3 y85 + u_{it}$$

- γ_1 representa el efecto de la educación en el salario en 1978 y γ_2 representa el efecto de la educación en el salario en 1985.

$$\widehat{log(w)} = \underset{(0.087)}{1.030} + \underset{(0.007)}{0.052}\,\text{educ78} + \underset{(0.008)}{0.077}\,\text{educ85} + \underset{(0.137)}{0.029}\,\text{y85}$$

- El efecto estimado de la educación en 1978 es 5,2%; en 1985 es 7,7%. Para compararlos es necesario contrastar: $H0)\gamma_2 - \gamma_1 = 0$

Alternativa 2

- La otra alternativa es correr MCO en una constante, en *educ* y todas las interacciones menos una, por ejemplo para el primer año.
- Cada pendiente de las interacciones mide los retornos a la educación con respecto al período base (en este caso el año 1978).

$$log(w_{it}) = \alpha_0 + \alpha_1 educ_{it} + \alpha_2 educ_{it} \times y85_t + \alpha_3 y85 + u_{it}$$

- La estimación MCO:

$$\widehat{log(w)} = \underset{(0.087)}{1.030} + \underset{(0.007)}{0.052}\,\text{educ} + \underset{(0.011)}{0.025}\,educ \times y85 + \underset{(0.137)}{0.030}\,y85$$

- El efecto estimado de la educación en 1978 es 5,2%; en 1985 es 2,5 puntos porcentuales más alto.
- Contrastar si el retorno a la educación se ha mantenido constante en el tiempo: $H0)\alpha_2 = 0$.

El Contraste de Chow de Cambio Estructural

Todos los parámetros pueden variar

$$log(w_{it}) = \beta_{0t} + \beta_{1t} educ_{it} + u_{it}$$

- Podemos aplicar un contraste F para analizar si los parámetros son diferentes en dos periodos temporales distintos. Este contraste se conoce como contraste de Chow.
- Computar el estadístico F requiere usar la suma de cuadrados de los residuos (SCE) del modelo restringido y del modelo no restringido.
- El modelo restringido es el modelo con los datos fusionados.
- La SCE del modelo no restringido se obtiene sumando las SCE de las estimaciones para cada período por separado.

Chow (cont.)

- En nuestro ejemplo de salarios podemos estimar:
  $log(w) = \beta_0 + \beta_1 educ + \delta_1 y85 + \delta_2 educ * y85 + u$
- En esta ecuación $\delta_1 = 0$ implica que no hay diferencia en la constante, $\delta_2 = 0$ que no hay diferencia en el retorno a la educación.
- La hipótesis nula del contraste de Chow es: $H_0 : \delta_1 = \delta_2 = 0$ (que no hay cambio estructural).
- Para obtener un contraste robusto en `gretl`, se puede estimar la regresión conjunta con interacciones, usando errores estándar robustos y contrastar las restricciones usando `omit`.
- También se puede estimar la regresión conjunta sin interacciones y usar el comando $chow\ y85 - -dummy$.

Ejemplo

- Basados en nuestro ejemplo de salarios estimaremos un modelo para analizar la evolución de la diferencia salarial por sexo.
- Estimemos el siguiente modelo:
  $log(w_{it}) = \beta_0 + \beta_1 y85 + \beta_2 educ_{it} + \beta_3 educ_{it} \times y85_t + \beta_4 exper + \beta_5 exper^2 + \beta_6 union + \beta_7 female + \beta_8 female \times y85_t + u_{it}$
- En 1978, manteniendo el resto de variables constantes, una mujer ganaba un 31.7 por ciento menos que un hombre. La diferencia salarial disminuye alrededor de 8.5 pp en el período, pero se rechaza la hipótesis nula de que la diferencia es cero frente a la alternativa unilateral de que es positiva.

Ejemplo - posibles preguntas

- Interprete el coeficiente de la variable $y85$. ¿Tiene alguna interpretación interesante?
- Si suponemos que todos los demás factores permanecen constantes, ¿Cuál es el aumento estimado en el salario nominal para un hombre con doce años de educación?
- Si usamos salarios reales (con base 1978), ¿qué coeficientes cambian respecto a la estimación anterior? Para hacerlo en gretl usar el deflactor para 1985 que es 1,65.
- Modificar el modelo de modo de analizar si el diferencial de salarios de los sindicalistas ha cambiado en el tiempo.