

QLAB

Práctica Calificada - Módulo de IV: Replicando a David Card, Premio Nobel de Economía 2021

Docente: Dr. Juan Manuel del Pozo Segura

Para la realización de este ejercicio se requiere utilizar el software STATA, pero se permite el uso de otro programa estadístico siempre que pueda realizar lo solicitado. Se requiere la entrega de un informe en word o pdf (latex) con las respuestas a los ejercicios. Este informe deberá ser elaborado de forma grupal en grupos de hasta 5 personas. La calificación es sobre 20 puntos

Usted cuenta con la base de datos card.dta, correspondiente al paper “ Using Geographic Variation in College Proximity to Estimate the Return to Schooling” de David Card, 1993. Su primera tarea es leer el artículo, esto facilitará el desarrollo de esta tarea

Sobre la base de los datos y lo visto en clases, responda lo siguiente (2 puntos cada una):

1. Realice una descripción de las características de los individuos de la muestra. ¿Qué información nos provee el promedio de la dependiente en la interpretación de los coeficientes? ¿Cómo se interpretan los coeficientes en este caso?
2. Regresione usando OLS el siguiente modelo en Stata

```
reg lwage educ c.exper##c.exper age black south smsa, smsa66, reg661-reg668
```

 ¿Qué ocurre con edad? ¿Por qué? Ahora saque edad del modelo anterior y regresione de nuevo. Interprete el estimado obtenido para educación, y compárelos con la Tabla 2, Columna 2, del paper
3. Vea la Tabla 2 de Card (1993). ¿Qué se encuentra al inspeccionar el coeficiente estimado de educación? ¿Es creíble este coeficiente? ¿Por qué? ¿Qué 2 posibles fuentes de inconsistencia podrían ocurrir como *resultado de la omisión de variables*?
 Use el marco visto en clase (las expresiones que forman parte de $\text{Inconsistencia} = \gamma \frac{\text{Cov}(x, a)}{\text{Var}(x)}$) para responder y mencione la dirección esperada del sesgo.
4. Estime la ecuación de primera etapa ¿Qué debemos ver en esta regresión? ¿Qué podría decir respecto de la correlación parcial existente entre educ y nearc4?
5. Estime el modelo usando el estimador de Variables Instrumentales, usando nearc4 como un instrumento para la variable educ. Dibuje el DAG correspondiente y explique por qué se espera que esta estrategia funcione. Comente sus resultados y compárelos con los obtenidos en el punto 2.



PUCP

Facultad de
Ciencias Sociales

6. Testee la presencia de instrumentos débiles usando i) el estadístico F; ii) el estadístico de Cragg y Donald (1993); iii) las tablas de Stock y Yogo (2005) con respecto a la medida del test de Wald
7. Use nearc2 con nearc4 como instrumentos para educ. Primero estime la primera etapa para educ y analice cuál de los dos instrumentos está más fuertemente relacionada parcialmente con educ. Después use el estimador de IV usando por separado nearc2 y nearc4 como instrumento para educ. Luego use el de 2SLS (incluyendo ambos instrumentos). Discuta sus resultados. Verifique que el estimador 2SLS es válido.
8. Hemos asumido hasta ahora homoscedasticidad ¿Es razonable esto? ¿Qué permite el estimador de GMM que no permite el de IV en 2 etapas bajo heterosk. Compare los resultados (del coeficiente educación) del estimador 2SLS con el del estimador GMM bajo heterocedasticidad
9. Sigamos con el caso donde se tienen dos instrumentos. Dado lo hallado en términos de si los instrumentos por separado son débiles o no en la pregunta 7, ¿qué concluye ud. a partir del resultado del test de sobreidentificación? ¿Con qué instrumento o instrumentos se quedaría al final? Realícelo asumiendo heterocedasticidad para el estimador GMM y 2SLS (HINT: piense 2 veces)
10. ¿Deberíamos usar MCO y aceptar la inconsistencia o usar en vez IV a pesar de que es un estimador consistente pero ineficiente? ¿Hay forma de testear esto? Si sí, Llévela a cabo para el modelo sobre-identificado. ¿Qué concluye?

Dr. Juan Manuel del Pozo Segura
November 2024