Variáveis instrumentais

Wooldridge, capítulo 15

Da aula passada

- $$\widehat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{SQT_x R_{xz}^2}$$

  Preço a pagar quando se usa VI quando x e u são não correlacionados: a variância assintótica do estimador de VI é maior e algumas vezes muito maior que a variância assintótica do estimador MQO

- Estimativas de VI podem ter grandes erros-padrão, especialmente se z e x forem fracamente correlacionados
- Correlação fraca entre z e x pode ter consequências ainda mais sérias: o estimador VI pode ter um grande viés assintótico, mesmo se z e u forem só moderadamente correlacionados

Efeitos de instrumentos fracos

- O que acontece se a hipótese Cov(z,u) = 0 é falsa?
- O estimador VI também será inconsistente
- Comparar o viés assintótico VI e MQO
- Preferir VI com base no viés assintótico se Corr(z,u)/Corr(z,x) < Corr(x,u)

$$\text{IV}: \text{plim}\,\hat{\beta}_1 = \beta_1 + \frac{Corr(z,u)}{Corr(z,x)} \bullet \frac{\sigma_u}{\sigma_x}$$

$$\text{OLS}: \text{plim}\,\tilde{\beta}_1 = \beta_1 + Corr(x,u) \bullet \frac{\sigma_u}{\sigma_x}$$

Conclusão

- VI é consistente se z e u são não correlacionados e z e x são correlacionados (positiva ou negativamente)
- Embora VI seja consistente, os estimadores VI podem ter erros padrão grandes, especialmente se z e x forem fracamente correlacionados
- A fraca correlação entre z e x pode ter consequência ainda mais séria: o estimador VI pode ter um grande viés assintótico

Problema com instrumentos: um exemplo

- Estimar retorno da educação: fazer a regressão do logaritmo dos salários nos anos de escolaridade usando dados sobre os indivíduos.
- Se os indivíduos mais aptos são também mais bem sucedidos no mercado de trabalho e frequentam a escola por mais tempo, então anos de escolaridade serão relacionados com a variável omitida aptidão e o estimador de MQO do retorno da educação será viesado.
- Usar regressão de VI. Mas que variável é correlacionada com anos de educação, mas não com o termo de erro na regressão de salários, ou seja, qual variável é um instrumento válido?

Does compulsory school attendance affect schooling and earnings? Angrist e Krueger

- Por causa da obrigatoriedade de frequentar a escola até os 16 anos, a data de aniversário está correlacionada com os anos de escolaridade
- Os alunos nascidos no início do ano , em geral, começam a estudar com mais idade. Portanto, eles atingem o tempo de estudo obrigatório (16 anos na maioria dos estados) com um pouco menos de educação do que os alunos que começaram a estudar com menos idade
- Se a lei requer que o aluno vá à escola até fazer 16 anos e o aluno completa 16 anos em janeiro e ele desiste de estudar ele será menos educado do que um aluno que faz 16 anos em dezembro (este último terá completado o décimo ano enquanto o outro aluno não)

Exemplo

- Maioria dos estados exige que os alunos comecem a escola quando eles vão completar 5 anos
- Aluno J nasceu em 1 de Janeiro e então está prestes a completar 6 anos quando começa a frequentar a escola em setembro. Aluno D nasceu em 1 de Dezembro e então não completou 5 anos quando começa a estudar
- Aluno J começou a escola com 5 anos e 8 meses e fez 16 anos em Janeiro, 10 anos mais tarde. Aluno D começou a escola com 4 anos e 9 meses e fez 16 anos em Dezembro, 11 anos mais tarde
- Ambos querem largar a escola tão logo possam, ou seja, tão logo completem 16 anos, mas D por acaso acabou completando um ano a mais de estudo do que J

- Data do nascimento é bastante aleatória, mimetizando a atribuição aleatória dos experimentos

- Cenário promissor para VI

- Angrist e Krueger usaram as diferenças na escolaridade geradas pelo trimestre de nascimento para obter estimativas VI dos retornos econômicos da educação

- Instrumento: *dummy* (pritrim) igual a 1 se o homem nasceu no primeiro trimestre e zero caso contrário
- Os anos de estudo realmente diferem sistematicamente na população com base no trimestre do nascimento
- O termo de erro (em particular a aptidão) não deve ser relacionada com o trimestre de nascimento.
- Estimativa de VI não difere muito da estimativa de MQO, o que foi interpretado como uma demonstração de que não existe viés de aptidão omitida quando as equações de salários são estimadas por MQO

Problems with instrumental variables estimation when the correlation between the instruments and endogenous explanatory variables is weak Bound, Jaeger e Baker

- John Bound: sabia que instrumentos fracos tornam as estimativas não confiáveis e se preocupou que, apesar da amostra extremamente grande, o instrumento fosse fraco
- Bound e Krueger se encontraram, conversaram sobre o assunto e Krueger sugeriu um jeito de descobrir a verdade: refazer as regressões usando um instrumento realmente irrelevante – substituir o verdadeiro trimestre de nascimento por um trimestre de nascimento falso, gerado aleatoriamente pelo computador – e comparar os resultados usando o instrumento verdadeiro e o falso

- Resultado: não importava se fosse usado o verdadeiro trimestre ou o falso como instrumento, as estimativas eram basicamente as mesmas
- Datas de aniversário não são literalmente atribuídas aleatoriamente; evidências de que o trimestre de nascimento não é independente das características das mães

VI para regressões múltiplas

- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$, onde y_2 é endógena e z_1 é exógena
- Seja z_2 o instrumento, de tal forma que $Cov(z_2,u_1) = 0$ e
- $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2$, onde $\pi_2 \neq 0$
- A equação na forma reduzida regride a variável endógena em todas as exógenas

Resumo das hipóteses

- z_1 e z_2 são não correlacionados com u_1
- u_1 tem média zero
- $E(u_1) = 0, \mathrm{cov}(z_1, u_1) = 0, cov(z_2, u_1) = 0$
- Devido a hipótese de média zero de u_1, as duas últimas hipóteses são equivalentes a
- $\mathrm{E}(z_1, u_1) = 0, E(z_2, u_1) = 0$

Resolver as correspondentes amostrais

$$\sum_{i=1}^{n}\left(y_{i1}-\hat{\beta}_0-\hat{\beta}_1 y_{i2}-\hat{\beta}_2 z_{i1}\right)=0$$

$$\sum_{i=1}^{n} z_{i1}\left(y_{i1}-\hat{\beta}_0-\hat{\beta}_1 y_{i2}-\hat{\beta}_2 z_{i1}\right)=0$$

$$\sum_{i=1}^{n} z_{i2}\left(y_{i1}-\hat{\beta}_0-\hat{\beta}_1 y_{i2}-\hat{\beta}_2 z_{i1}\right)=0$$

- Ainda é necessário que a variável instrumental z_2 seja correlacionada com y_2
- Como acabamos de ver esta hipótese é estabelecida em termos de correlação parcial
- Após considerar os efeitos parciais z_1, z_2 e y_2 são correlacionados

- A adição de mais variáveis explicativas exógenas ao modelo é direta
- $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1 \qquad (1)$

Onde y_2 é correlacionada com u_1

- Definir z_k como uma variável exógena que não pertence à equação

Assume-se que

$E(u_1)$=0, $cov(z_j,u_1)$=0 j=1,...,k (2)

Forma reduzida de y_2:

$y_2 = \pi_0 + \pi_1 z_1 + \cdots . + \pi_{k-1} z_{k-1} + \pi_k z_k + v_2$ (3)

Testar $\pi_k \neq 0$ (4)

- Sob (1)-(4) z_k é uma VI válida de y_2
- As variáveis exógenas servem como suas próprias VIs. $z_1, \dots, z_{k-1}$ são seus próprios instrumentos
- Portanto, a lista de variáveis exógenas com frequência é chamada de lista de variáveis instrumentais
- Hipótese adicional: não há relações lineares perfeitas entre as variáveis exógenas

Notação matricial

- $\hat{\beta}_{MQO} = (X'X)^{-1}X'\mathrm{y}$
- $\hat{\beta}_{VI} = (Z'X)^{-1}Z'\mathrm{y}$
- $\mathrm{var}\widehat{(\beta_{MQO})} = \hat{\sigma}^2(X'X)^{-1}$
- $\mathrm{var}\widehat{(\beta_{VI})} = \hat{\sigma}^2(Z'X)^{-1}Z'Z(X'Z)^{-1}$