Mínimos quadrados dois estágios

Wooldridge, capítulo 15

- Considerar o modelo estrutural $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$, onde y_2 é uma variável endógena e z_1 é uma variável exógena
- Supor agora que temos duas variáveis excluídas da equação acima: z_2 e z_3

- ◆ Se z_2 e z_3 foram ambas correlacionadas com y_2 , podemos usar cada uma delas como instrumento
- O problema neste caso é que teríamos dois estimadores VI
- Como z_1, z_2 e z_3 são não correlacionados com u_1 , qualquer combinação linear de z_1, z_2 e z_3 também não será correlacionada com u_1

Melhor VI

- ◆ Combinação linear mais altamente correlacionada com y_2
- Fornecida pela equação na forma reduzida

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2$$

$$E(v_2) = 0, Cov(z_1, v_2) = 0, Cov(z_2, v_2) = 0, Cov(z_3, v_2) = 0$$

Melhor VI

- O melhor instrumento é uma combinação linear de todas as variáveis exógenas, $y_2^* = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3$
- Para que esta VI não seja perfeitamente correlacionada com z_1 precisamos que $\pi_2 \neq 0$ ou $\pi_3 \neq 0$
- Hipótese de identificação
- Testar $H_0: \pi_2 = 0$ e $\pi_3 = 0$ contra a alternativa usando um teste F

Usar os dados da amostra

- Estimar y_2^* regredindo y_2 em z_1 , z_2 e z_3 – chamamos isto de primeiro estágio
- $\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3$
- Verificar se z_2 e z_3 são conjuntamente significantes
- Caso sim, usar $\hat{y}_2$ como VI para y_2

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

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

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

- Com instrumentos múltiplos o estimador de VI é chamado estimador de mínimos quadrados de dois estágios (MQ2E) – *two stages least squares* (2SLS)
- Quando usamos $\hat{y}_2$ como VI de y_2 as estimativas são idênticas àquelas obtidas: (i) estimando a equação na forma reduzida por MQO para obter $\hat{y}_2$; (ii) fazendo a regressão de y_1 em $\hat{y}_2$ e z_1 também usando MQO

- No modelo com uma única variável instrumental para y_2 o estimador de VI é idêntico ao estimador MQ2E
- Portanto, quando temos uma VI para cada variável explicativa endógena podemos chamar o método de estimação de VI ou MQ2E

- O método pode ser estendido para o caso onde existem múltiplas variáveis endógenas – necessário ter certeza de que existem pelo menos tantas variáveis excluídas (instrumentos) quando endógenas incluídas na equação estrutural

Hipóteses MQ2E

- 1) Linearidade nos parâmetros
- 2) Amostragem aleatória
- 3) Variáveis instrumentais exógenas

O termo de erro u tem média zero e cada VI é não correlacionada com u

- 4) (i) Não há relações lineares perfeitas entre as variáveis instrumentais; (ii) A condição de classificação de identificação é válida

Obs: Declarar a condição de classificação com duas ou mais variáveis explicativas endógenas exige álgebra matricial

- Teorema 1: Sob as hipóteses 1 a 4 o estimador MQ2E é consistente

5) Homocedasticidade

$$E(u^2 | \mathbf{z}) = \sigma^2$$

- Teorema 2 Sob as hipóteses 1 a 5 os estimadores MQ2E são normalmente distribuídos assintoticamente

Estimadores consistentes da variância assintótica são dados por $\frac{\sigma^2}{\widehat{SQT}_2(1-\widehat{R}_2^2)}$

$\widehat{SQT}_2$ é variação total em $\hat{y}_2$, $\widehat{R}_2^2$ é o R-quadrado da regressão de $\hat{y}_2$ em todas as variáveis exógenas que aparecem na equação estrutural e σ_2 é substituída por $\hat{\sigma}^2 = \sum_{i=1}^n \hat{u}_i^2 / (n-k-1)$ e os $\hat{u}_i$ são os resíduos MQ2E

- Teorema 3 Sob as hipóteses 1 a 5 o estimador MQ2E é assintoticamente eficiente

- Se a hipótese de homocedasticidade não se sustentar, os estimadores MQ2E ainda assim serão assintoticamente normais, mas os erros-padrão e as estatísticas t e F precisarão ser ajustados
- Além disso, o estimador MQ2E não será o estimador eficiente

Teste de endogeneidade: teste de regressão

$$\blacklozenge y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1 \quad (1)$$

- z_3 e z_4
- $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + v_2$
- y_2 será não correlacionado com u_1 se, e somente se, v_2 for não correlacionado com u_1
- Escrever $u_1 = \delta_1 v_2 + e_1$, onde e_1 é não correlacionado com v_2 e tem média zero
- Então u_1 e v_2 serão não correlacionados se, e somente, se $\delta_1 = 0$

- Maneira mais fácil de fazer isso: incluir v_2 como um regressor adicional na equação estrutural e fazer um teste t
- Problema: v_2 não é observado porque ele é o termo de erro na equação da forma reduzida
- Solução: estimar a equação na forma reduzida por MQO e usar os resíduos da forma reduzida, $\hat{v}_2$

- Estimar $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \delta_1 \hat{v}_2 + \text{erro (2)}$ por MQO
- Testar $H_0 : \delta_1 = 0$ usando a estatística t
- Se rejeitarmos H_0 , concluímos que y_2 é endógeno porque v_2 e u_1 são correlacionados

Teste de endogeneidade de uma única variável explicativa

- Salvar os resíduos do primeiro estágio
- Incluir o resíduo na equação estrutural (que obviamente tem a variável y_2)
- Se o coeficiente do resíduo é estatisticamente diferente de zero, rejeitar a nula de exogeneidade

Teste de endogeneidade de mais de uma variável explicativa

- Teste de endogeneidade de múltiplas variáveis explicativas: para cada variável suspeita de ser endógena obtemos os resíduos da forma reduzida. Depois verificamos a significância conjunta desses resíduos na forma estrutural usando um teste F. A significância conjunta indica que pelo menos uma variável explicativa é endógena
- O número de restrições de exclusão testadas é o número de variáveis explicativas suspeitas de serem endógenas

Teste de endogeneidade: comparação estimativas MQO e VI

Testes baseados na comparação entre dois estimadores diferentes são chamados testes de Hausman, em referência a Hausman (1978), ou testes Wu-Hausman ou mesmo testes Durbin-Wu-Hausman em referência a Wu e Durbin que propuseram testes similares

Fórmula geral da estatística de Hausman

- $H = (\hat{\beta}_C - \hat{\beta}_E)' (\widehat{AV(\hat{\beta}_C)} - AV(\hat{\beta}_E))^{-1} (\hat{\beta}_C - \hat{\beta}_E)$
 $\sim \chi_k^2$
- H_0 : exogeneidade, H_1 : endogeneidade
- O estimador $\hat{\beta}_C$ é consistente sob ambas as hipóteses H_0 e H_1 , enquanto o estimador $\hat{\beta}_E$ apenas é consistente sob a hipótese nula
- Por outro lado, sob a hipótese nula $\hat{\beta}_E$ é (assintoticamente) mais eficiente que $\hat{\beta}_C$

No nosso caso de interesse

- $H = (\hat{\beta}_{VI} - \hat{\beta}_{MQO})' (\widehat{AV(\hat{\beta}_{VI})} - \widehat{AV(\hat{\beta}_{MQO})})^{-1} (\hat{\beta}_{VI} - \hat{\beta}_{MQO})$
- Sob a nula (exogeneidade) MQO e VI são consistentes, mas somente MQO é eficiente, sob a alternativa MQO é inconsistente, mas VI é consistente

Para testar se um regressor particular x , que tem o parâmetro β_2 associado, não é endógeno:

$$H_0: \text{Cov}(x, u) = E(ux) = 0 \leftrightarrow \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n x_i u_i \right) = 0$$

Pode-se utilizar a estatística

$$\frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MQO})^2}{V(\hat{\beta}_{2,VI}) - V(\hat{\beta}_{2,MQO})} \rightarrow \chi_1^2$$