Avaliação de Impactos

HO 731 – Econometria

Prof. Alexandre Gori Maia

Instituto de Economia - UNICAMP

Ementa

inferência Causal

Contrafactual

Viés de Seleção

Mínimos Quadrados em Dois Estágios

Bibliografia

Angrist, J.; Pischke, J. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Caps. 1-4, 2009.

World Bank (2013). Impact Evaluation in Practice.

Inferência Causal

Suponha que desejamos avaliar o impacto de um programa de assistência técnica (*T*=1) sobre a renda (*Y*) de um conjunto de produtores agrícolas.

A relação causal entre a participação no programa e a renda do produtor i seria dada pela diferença entre a renda desse produtor após receber a assistência -$(Y_i|T=1)$ - e a renda desse mesmo produtor caso não tivesse recebido a assistência -$(Y_i|T=0)$:

$$\Delta = (Y_i|T=1) - (Y_i|T=0)$$

O resultado de Y_i quanto $T=0$ é denominado de **contrafactual**, ou seja, qual seria o resultado para um participante do programa na ausência do programa.

O problema é que não é possível observar dois resultados ao mesmo tempo para a mesma pessoa i. Ou seja, para os produtores que participaram do programa ($T=1$), apenas é possível observar a renda na presença do programa, não sendo possível calcular $(Y_i|T=0)$.

Contrafactual

As duas principais estratégias para estimar contrafacturais são:

1) Comparar os resultados de um mesmo grupo, antes e depois da realização do programa. O problema é não poder afirmar que a variação na renda do grupo deva-se exclusivamente à participação no programa. Por exemplo, a renda média do grupo poderia ter aumentado no período, mesmo na ausência do programa, devido a outros fatores, como aumento do preço de venda do produto agrícola. Todos esses fatores precisam ser devidamente controlados no modelo de regressão;

2) Comparar um grupo que participou do programa (grupo de *tratamento*) com um grupo que não participou do programa (grupo de *controle*): o ideal seria ter grupos de tratamento e controle que representem seleções puramente aleatórias da população (a aleatoriedade garantiria características semelhantes). Mas isso dificilmente ocorre, e os grupos podem diferir em função de fatores que não podem ser controlados pelo modelo de regressão, como atitudes e habilidades (viés de seleção).

Uma das estratégias para contornar o viés de seleção da estratégia (2) é o uso de variáveis instrumentais.

Viés de Seleção - Definição

Suponha que o objetivo seja avaliar o impacto de um tratamento *T*, controlando por um conjunto de variáveis observáveis (**X**):

$$\mathbf{y} = \mathbf{X\beta} + \delta T + \mathbf{e}$$

O problema é que a participação no tratamento *T* pode depender de fatores não observáveis que também influenciam o resultado (erros **e**). Por exemplo, os produtores do grupo de tratamento podem possuir habilidades agrícolas que não estão presentes no grupo de controle. Isso ocorre quando a seleção no tratamento não é aleatória, como é comum em ações de políticas públicas. A participação é usualmente direcionada a um grupo específico da população ou é voluntária. Nesses casos, não há como garantir que a seleção seja aleatória.

Caso os fatores não observáveis que influenciam a participação no tratamento também influenciem o resultado *Y*, as estimativas de MQO serão tendenciosas e inconsistentes, pois os erros estarão associados ao regressor *T*, ou seja:

$$E(\mathbf{e}|\mathrm{T}, \mathbf{X}) \neq 0 \Rightarrow E(\mathbf{y}|T, \mathbf{X}) \neq \mathbf{X\beta} + \delta T$$

Viés de Seleção - Exemplo

Seja um programa, chamado HISP (World Bank, 2016), que objetiva subsidiar a atenção básica à saúde da população carente. Foi aplicado um piloto em comunidades rurais selecionadas. A participação era voluntária. Após 2 anos de teste piloto, foram entrevistados 2.965 participantes (*T*=1) e 2.964 não participantes (*T*=0). Para avaliar a efetividade do programa na saúde, podemos estimar o impacto da participação no HISP (*T*=1) sobre os gastos com saúde (*Y*):

$$Y_i = \alpha + \delta T_i + \beta_1 educ_i + \beta_2 age_i + e_i$$

As variáveis *educ* e *age* representam, respectivamente, escolaridade e idade da pessoa responsável pelo domicílio. As estimativas de MQO, que ignoram a seletividade do grupo de tratamento (*T*=1) e controle (*T*=0), seriam:

$$Y_i = 14{,}52 - 12{,}13 T_i + 0{,}14 educ_i + 0{,}12 age_i + \hat{e}_i$$

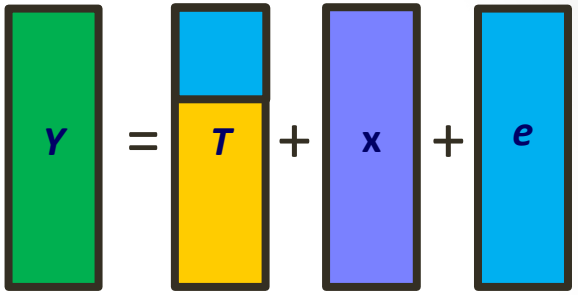
O gasto médio com saúde do grupo de tratamento foi 12,1 mil inferior ao do grupo de controle. Mas não podemos afirmar seja impacto do HIPS. Como a adesão foi voluntária, o grupo de participantes pode, por exemplo, ser formado por pessoas com maior preocupação com a atenção básica à saúde. Essas teriam menos problemas e menos gastos com saúde.

Variáveis Instrumentais

Desejamos analisar: $Y_i = \alpha + \delta T_i + \mathbf{x}_i\boldsymbol{\beta} + e_i$

Mas temos: $Cov(T_i, e_i) \neq 0$

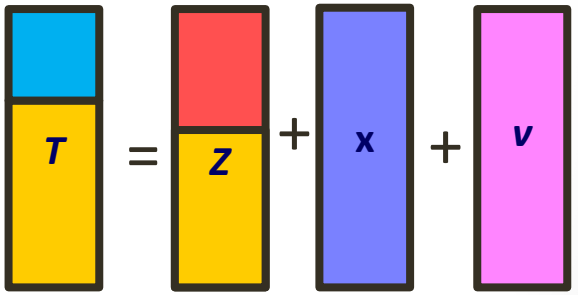
A participação no tratamento T está associada aos erros do modelo e os estimadores de MQO serão inconsistentes



Desejamos encontrar um instrumento Z não associado aos erros e : $Cov(Z_i, e_i) = 0$ e $Cov(Z_i, T_i) \neq 0$

A parcela de Z que interessa é aquela associada a T, que será estimada por:

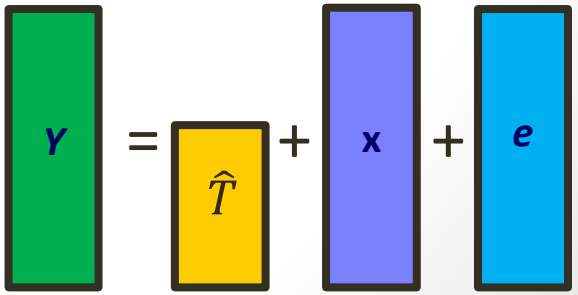
$$\hat{T}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i + \mathbf{x}_i\hat{\boldsymbol{\pi}}_2$$



Um estimador consistente para δ será dado por:

$$Y_i = \alpha + \delta\hat{T}_i + \mathbf{x}_i\boldsymbol{\beta} + e_i$$

Embora consistente, o estimador obtido com o uso de VI tende a ser **viesado para amostras pequenas**.



Mínimos Quadrados 2 Estágios

Passos para avaliar o impacto de *T* com estimadores de VI por MQ2E:

1) **Identificação:** é necessário pelo menos uma variável instrumental *Z* que esteja associada à participação no treinamento e não aos erros;
2) **Equação de seleção**: a participação no tratamento é modelada por um função de todas os regessores (**x**) mais o(s) instrumento(s) *Z*;
3) **1º estágio**: estimar valor previsto da participação no tratamento por MQO;
4) **2º estágio**: substituir a variável *T* do modelo para o resultado *Y* pelo valor previsto na etapa 3 e aplicar novamente MQO;

O modelo para o resultado :

$$Y = \alpha + \delta T + \mathbf{x\beta} + e \quad (1)$$

Z é o instrumento para *T*

O modelo para a seleção: (2)

$$T = \pi_0 + \pi_1 Z + \mathbf{X\pi} + u$$

MQO

$$\hat{T} = \hat{\pi}_0 + \hat{\pi}_1 Z + \mathbf{x\hat{\pi}} \quad (3)$$

$$Y = \alpha + \delta \hat{T} + \mathbf{x\beta} + e \quad (4)$$

Importante

Os estimadores de MQ2E são consistentes mas podem ser tendenciosos para amostras pequenas

MQ2E - Exemplo

Para estimular a adesão ao HISP, o governo local promoveu uma série de campanhas de promoção em comunidades selecionadas aleatoriamente. A variável binária *Z* assume 1 quando o domicílio está localizado em uma comunidade que recebeu a campanha promocional. Espera-se que a campanha tenha um impacto positivo na participação - $Cov(Z,T) \neq 0$. Por outro lado, a campanha foi aleatória nas comunidades e não estaria associada a fatores não observados que afetam os gastos com saúde, como a preocupação com a saúde básica - $Cov(Z,e) = 0$.

A variável Z pode então ser o instrumento para a participação no programa (*T*). O modelo de seleção do 1º estágio será:

$$T_i = \pi_0 + \pi_1 Z_i + \pi_2 educ_i + \pi_3 age_i + v_i$$

E as estimativas de MQ2E seriam:

$$Y_i = 12{,}40 - 8{,}95\hat{T}_i + 0{,}19esc_i + 0{,}14idade_i + \hat{e}_i$$

A estimativa de MQ2E para o impacto do HISP no gasto com saúde é inferior à de MQO, indicando uma redução de 8,95 mil. Ou seja, a estimativa de MQO estaria superestimando o impacto do HISP, provavelmente porque aqueles com mais preocupação com a saúde são mais propensos à participar do programa.