

Endogeneidade, Variáveis Instrumentais e Modelos de Equações Estruturais

**Ernesto F. L. Amaral
Magna M. Inácio**

**21 de outubro de 2010
Tópicos Especiais em Teoria e Análise Política:
Problema de Desenho e Análise Empírica (DCP 859B4)**

ENDOGENEIDADE

E

VARIÁVEIS INSTRUMENTAIS

PROBLEMAS DE AUTO-SELEÇÃO

- Indivíduos se auto-selecionam para certos procedimentos ou programas, o que não é uma escolha aleatória:
 - Pessoas decidem por vínculo partidário por diversas razões (individuais e contextuais).
 - Crianças entram em programas de saúde infantil por decisão dos pais.
 - Decisão de cidades por implementar determinadas leis está relacionada com vários fatores não aleatórios.

INCLUSÃO DE NOVAS VARIÁVEIS

- A participação não é determinada de forma aleatória, havendo problemas de **auto-seleção**.
- É preciso incluir outras variáveis na regressão, quando uma variável independente estiver sistematicamente relacionada com fatores não-observados.
- Incluímos fatores relacionados com a variável independente de interesse para corrigir endogeneidade.

USO DE VARIÁVEL INSTRUMENTAL

- Em uma equação de regressão:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- X_i pode ser dependente do termo de erro e_i (ou seja, a variável independente é endógena).
- Por isso, os coeficientes estimados serão viesados.
- Modelos com variáveis instrumentais possibilitam correção das estimativas, disponibilizando parâmetros consistentes.
- Para usar tal modelo, é preciso utilizar uma variável aleatória (Z_i), de modo que:
 - Seja estatisticamente independente do termo de erro (ε_i).
 - X_i e Z_i sejam correlacionadas.
- Em geral, é importante defender teoricamente que seu modelo obedece o pressuposto de exogeneidade.

DUNNING (2008)

- Dunning (2008) chama atenção de que não é somente a exogeneidade do instrumento (Z_i) que permite a estimação do efeito de tratamento (X_i).
- A inferência também depende da definição do modelo causal que deve ser expresso na equação de regressão.
- Sem a definição da equação, não é possível teorizar sobre o termo de erro, a exogeneidade e a inferência causal em modelos com variáveis instrumentais.
- Dunning chama atenção para o pressuposto de que a variação no regressor endógeno (X_i), relacionada à variável instrumental (Z_i), deve ter o mesmo efeito causal que a variação não-relacionada com o instrumento.

MEDINDO ATITUDE POLÍTICAS

- Seja: X_i uma medida de renda; Y_i uma medida de atitudes políticas (tal como opiniões sobre impostos); e uma população limitada aos participantes de uma loteria.
- A renda total do indivíduo i consiste em $X_i \equiv X_{1i} + X_{2i}$, onde:
 - X_{1i} é a renda habitual.
 - X_{2i} mede os ganhos em loterias.
- É provável que a renda total (X_i) seja endógena, porque fatores associados a características familiares influenciam tanto a renda total (X_i) como as atitudes políticas (Y_i).
- Porém, ganhos em loteria (X_{2i}) são correlacionados com a renda total (X_i) e possivelmente são exógenos.

ESPECIFICAÇÃO DO MODELO

- O modelo anterior assume que o aumento marginal de ganhos em loterias tem o mesmo efeito causal em atitudes políticas que o aumento marginal na renda habitual:

$$Y_i = \alpha + \beta(X_{1i} + X_{2i}) + \varepsilon_i$$

- Porém, ganhos em loterias podem ser vistos como fonte de renda muito diferente das demais, o que geraria uma influência específica sobre atitudes políticas:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

- Portanto, é preciso especificar o modelo com variável instrumental, antes de sua estimação.

EX.: CRESCIMENTO ECONÔMICO & CONFLITOS CIVIS

- Suponha que queremos medir o impacto de crescimento econômico em conflitos civis, com base em dados de países da África sub-Saariana.
- Como os conflitos civis também possuem um impacto sobre crescimento econômico, essa relação é endógena.
- Uma variável instrumental para crescimento econômico poderia ser mudanças anuais em chuvas.
- Diferentes fontes de crescimento econômico (agricultura ou indústria) podem ter diferentes impactos em guerras civis.
- Porém, mudanças em chuva podem estar associadas com crescimento na agricultura, mas não na indústria.

É PRECISO USAR CORRETAMENTE INSTRUMENTOS

- Estes problemas não significam uma falha geral em modelos com variáveis instrumentais.
- O que se questiona é a utilização de pressupostos de efeitos parciais homogêneos não condizentes com alguns casos.
- O principal objetivo de Dunning é de ressaltar a necessidade de discutir substantivamente a aplicação de modelos com variáveis instrumentais, ao invés de simplesmente aplicá-los com o pressupostos de efeitos homogêneos.
- Portanto, o uso desta modelagem para identificar efeitos endógenos do regressor (X_i) depende da especificação de modelos em que todas as partes componentes deste regressor tenham o mesmo efeito na variável dependente.

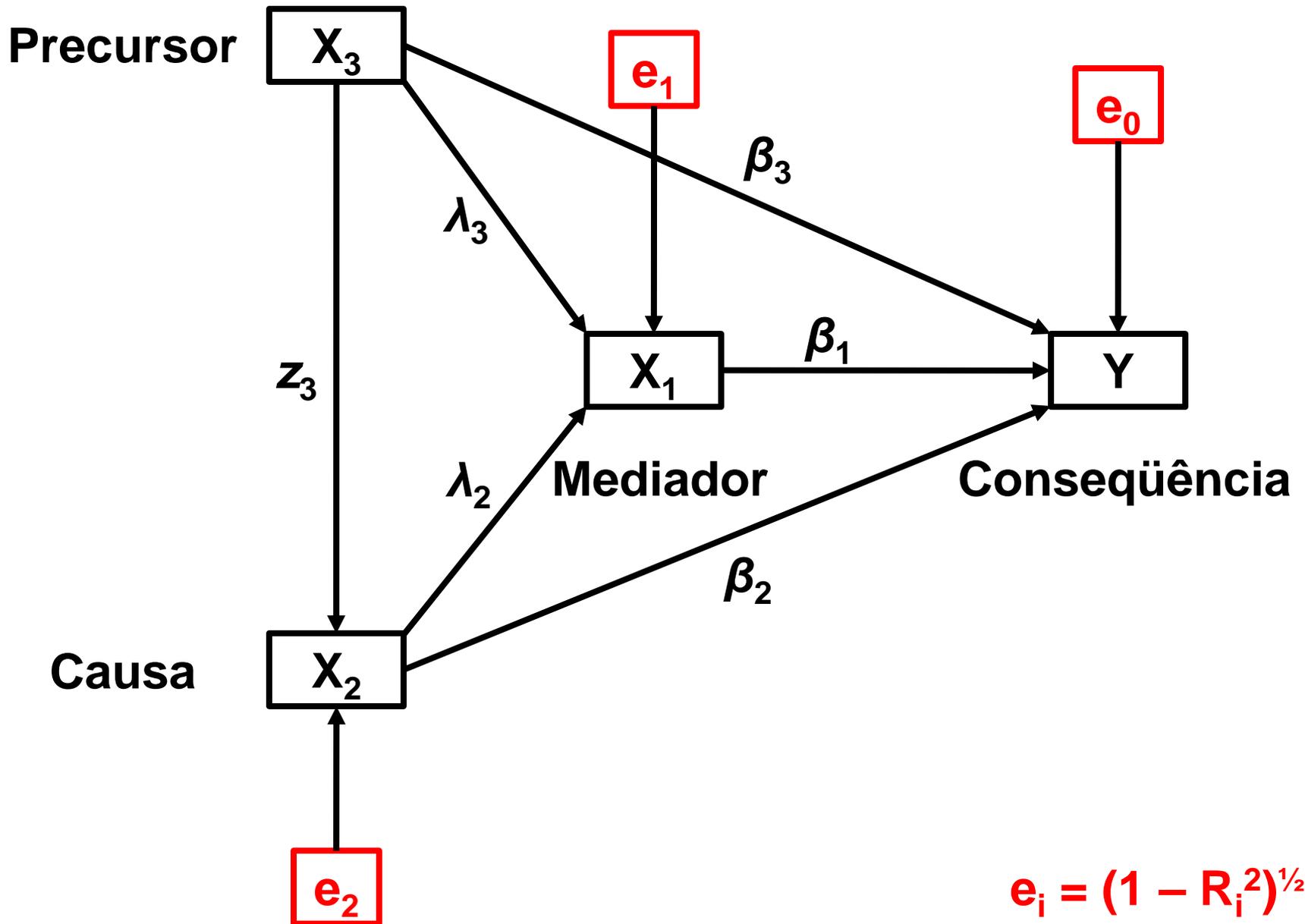
MODELOS DE EQUAÇÕES ESTRUTURAIS

MODELOS DE EQUAÇÕES ESTRUTURAIS

- Análise de trajetória (“path analysis”) é uma abordagem de modelagem causal que explora diversas relações entre variáveis em uma rede específica.

- Esse método é também conhecido como “Structural Equation Modeling” (SEM), “Covariance Structural Equation Modeling” (CSEM), “Analysis of Covariance Structures” ou “Covariance Structure Analysis”.

EXEMPLO DE ESTRUTURA DO MODELO



ANÁLISE DE TRAJETÓRIA (“PATH MODEL”)

– Observa-se cada um dos coeficientes separadamente:

INDEPENDENTES	DEPENDENTES		
	Causa (X_2)	Mediador (X_1)	Conseqüência (Y)
Precursor (X_3)	z_3	λ_3	β_3
Causa (X_2)	—	λ_2	β_2
Mediador (X_1)	—	—	β_1
R^2	R_2^2	R_1^2	R_0^2
$e = (1 - R^2)^{1/2}$	e_2	e_1	e_0

ASSOCIAÇÃO TOTAL, EFEITO TOTAL, EFEITO DIRETO E ¹⁵ EFEITO INDIRETO (via “path model”)

- 1) **Associação total entre “causa” e “conseqüência”:**
coeficiente de correlação entre “causa” e “conseqüência”.
- 3) **Efeito direto da “causa” sobre “conseqüência”:** coeficiente padronizado da “causa” sobre “conseqüência” (β_2).
- 4) **Efeito indireto da “causa” através do “mediador”:**
coeficiente padronizado da “causa” no “mediador” (λ_2)
multiplicado por coeficiente padronizado do “mediador” sobre
“conseqüência” (β_1).
- 2) **Efeito causal total da “causa” sobre “conseqüência”:** efeito direto (β_2) mais efeito indireto ($\lambda_2 * \beta_1$).
- 5) **Associação espúria devido ao “precursor”:** $z_3 * [\beta_3 + (\lambda_3 * \beta_1)]$ ou
associação total menos efeito causal total.

MODELO PROGRESSIVO (“PROGRESSIVE MODEL”)

– A variável dependente é a mesma em todos modelos:

INDEPENDENTES	DEPENDENTE		
	Conseqüência (Y)		
Causa (X_2)	z_2	λ_2	β_2
Precursor (X_3)	—	λ_3	β_3
Mediador (X_1)	—	—	β_1
R^2	R_2^2	R_1^2	R_0^2

ASSOCIAÇÃO TOTAL, EFEITO TOTAL, EFEITO DIRETO E EFEITO INDIRETO (via “progressive model”)

- 1) **Associação total entre “causa” e “conseqüência”:** coeficiente padronizado da “causa” sobre “conseqüência” (z_2).
- 2) **Efeito causal total da “causa” sobre “conseqüência”:** λ_2 .
- 5) **Associação espúria devido ao “precursor”:** associação total (z_2) menos efeito causal total (λ_2).
- 3) **Efeito direto da “causa” sobre “conseqüência”:** β_2 .
- 4) **Efeito indireto da “causa” através do “mediador”:** efeito causal total (λ_2) menos efeito direto (β_2).

OUTRAS INFORMAÇÕES

– **Qual o percentual de associação espúria?**

$$100 * (\text{associação espúria} / \text{associação total})$$

– **Qual o percentual de associação causal?**

$$100 * (\text{efeito causal total} / \text{associação total})$$

– **Qual o percentual do efeito causal total que ocorre através do mediador?**

$$100 * (\text{efeito indireto} / \text{efeito causal total})$$

**MODELOS DE
DOIS ESTÁGIOS DE QUADRADOS LINEARES:
“TWO-STAGE LEAST SQUARES” (2SLS)**

MODELOS HIERÁRQUICOS & MODELOS DE “CLUSTER”

- Huber, Kernell e Leoni (2005) realizam modelo em dois níveis.
- Eles não utilizam **modelos hierárquicos** (que permitem interceptos aleatórios) e nem **modelos de “cluster”** (que ajustam os valores dos erros padrões).
- Estes dois modelos assumem que os erros do primeiro nível (individual) e do segundo nível (países) são independentes (veja equações da página 376... próximo slide).
- Para que esse pressuposto seja satisfeito, seria preciso ter um número de unidades de segundo nível que tendesse ao infinito.
- Porém, o número de unidades (25 países) é menor do que 40 (mínimo exigido).

MODELOS HIERÁRQUICOS (PÁG. 376)

– Para pais (k) e indivíduo (i), há o seguinte modelo “probit”:

$$P(y_i = 1 \mid educ_i, age_i, age_i^2, male_i) \\ = \Phi(\beta_{const}^k + \beta_{educ}^k educ_i + \beta_{age}^k age_i + \beta_{age^2}^k age_i^2 + \beta_{male}^k male_i)$$

– A linha de base do vínculo partidário (β_{const}) e o efeito de educação no vínculo partidário (β_{educ}) são funções de variáveis de contexto:

$$\beta_{const}^k = \delta_{const} + \gamma_{const} z^k + \varepsilon_{const}^k \\ \beta_{educ}^k = \delta_{educ} + \gamma_{educ} z^k + \varepsilon_{educ}^k$$

– O modelo com equações de primeiro e segundo níveis é:

$$P(y_i = 1 \mid educ_i, age_i, male_i, z^k) \\ = \Phi \left[\begin{array}{l} \delta_{const} + \gamma_{const} z^k + \\ (\delta_{educ} + \gamma_{educ} z^k) educ_i + \\ (\delta_{age} + \gamma_{age} z^k) age_i + \\ (\delta_{age^2} + \gamma_{age^2} z^k) age_i^2 + \\ (\delta_{male} + \gamma_{male} z^k) male_i + \\ \varepsilon_{const}^k + \varepsilon_{educ}^k educ_i + \varepsilon_{age}^k age_i + \varepsilon_{age^2}^k age_i^2 + \varepsilon_{male}^k male_i \end{array} \right]$$

MODELOS BAYESIANOS

- Modelos Bayesianos também não são usados por Huber, Kernell e Leoni (2005), porque é difícil definir os valores “a priori” dos coeficientes aleatórios.
- Além disso, é complicado e demorado monitorar a convergência do modelo com os programas computacionais existentes.

MODELOS DE DOIS ESTÁGIOS DE QUADRADOS LINEARES (“2SLS”)

- Huber, Kernell e Leoni (2005) usam modelo de dois estágios de quadrados lineares.
- Primeiro, são estimadas regressões “probit” para cada um dos países no primeiro nível.
- Depois, são estimados os coeficientes de segundo nível, ajustando pelas variâncias dos erros do primeiro nível.
- São utilizados comandos de “loop” para estimar o primeiro nível de cada país e, em seguida, utilizar o resultado para inserir no segundo nível.

ESTIMAÇÃO DE “2SLS”

– Assuma que x_1 é endógeno; x_2 é exógeno; e que z_1 e z_2 são instrumentos para x_1 .

– Estimação de modelos com variáveis instrumentais:

```
ivregress 2sls y (x1=z1 z2) x2
```

– Informe que os resultados da regressão de primeiro estágio sejam mostrados:

```
ivregress 2sls y (x1=z1 z2) x2, first
```

– Uma forma intuitiva para calcular o modelo...

```
ivreg y (x1=z1 z2) x2
```

– ... e que gera os mesmos resultados é:

```
reg x1 z1 z2 x2
predict x1hat
reg y x1hat x2
```

TESTES PARA MODELOS COM INSTRUMENTOS

- Teste para a significância de instrumentos (regra geral é de rejeitar H_0 se $F > 10$):

```
estat firststage
```

- Teste para a validade dos instrumentos. Ou seja, testa se é necessário mais instrumentos do que variáveis endógenas:

```
estat overid
```

- Teste para a exogeneidade de todas variáveis:

```
estat endogenous
```

- Teste para a significância dos instrumentos (regra geral é de rejeitar H_0 se $F > 10$):

```
reg x1 z1 z2
```

```
test z1 z2
```