Aula Prática - Variáveis Instrumentais

Introdução

Nesta aula prática, replicaremos os principais resultados do Capítulo 12 de Stock e Watson (2020), que trata de **Regressão por Variáveis Instrumentais (VI)**. A aplicação empírica central explora a seguinte pergunta: **qual é a elasticidade-preço da demanda por cigarros nos Estados Unidos?**

Estimar essa elasticidade é uma questão de política pública relevante: se a elasticidade for alta, um aumento de impostos reduz substancialmente o consumo; se for baixa, o mesmo aumento arrecada mais receita, mas pouco altera o comportamento dos fumantes.

O desafio econométrico é que preço e quantidade são determinados *simultaneamente* pelo mercado: preço e quantidade são endógenos. Isso significa que uma simples regressão de MQO de $\ln(Q)$ sobre $\ln(P)$ não estima consistentemente a elasticidade da demanda. A solução é encontrar uma **variável instrumental (VI)**: uma variável que afeta o preço sem afetar diretamente a demanda, ou seja, que desloca a curva de oferta sem afetar a curva de demanda.

Pacotes e Dados

Os pacotes utilizados nesta aula são:

A base de dados `CigarettesSW` está disponível no pacote `AER`. Ela contém dados anuais de 48 estados contíguos dos EUA para os anos de 1985 e 1995.[1]

```
Rows: 96
Columns: 9
$ state      <fct> AL, AR, AZ, CA, CO, CT, DE, FL, GA, IA, ID, IL, IN, KS, KY,…
$ year       <fct> 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985,…
$ cpi        <dbl> 1.076, 1.076, 1.076, 1.076, 1.076, 1.076, 1.076, 1.076, 1.0…
$ population <dbl> 3973000, 2327000, 3184000, 26444000, 3209000, 3201000, 6180…
$ packs      <dbl> 116.4863, 128.5346, 104.5226, 100.3630, 112.9635, 109.2784,…
$ income     <dbl> 46014968, 26210736, 43956936, 447102816, 49466672, 60063368…
$ tax        <dbl> 32.50000, 37.00000, 31.00000, 26.00000, 31.00000, 42.00000,…
$ price      <dbl> 102.18167, 101.47500, 108.57875, 107.83734, 94.26666, 128.0…
$ taxs       <dbl> 33.34834, 37.00000, 36.17042, 32.10400, 31.00000, 51.48333,…
```

Dicionário de variáveis originais

Variável	Descrição
`state`	Estado (sigla)
`year`	Ano (1985 ou 1995)
`cpi`	Índice de preços ao consumidor (CPI)
`population`	População do estado
`packs`	Pacotes de cigarros vendidos *per capita*
`income`	Renda pessoal total do estado (nominal)
`tax`	Impostos estaduais + federais por pacote (em centavos)
`price`	Preço médio por pacote, incluindo todos os impostos (em centavos)
`taxs`	Impostos totais + imposto geral sobre vendas (sales tax) por pacote

[1] Os dados são descritos no Apêndice 12.1 do livro. A base originalmente possui 96 observações (48 estados × 2 anos).

Preparação dos Dados

Construímos as variáveis de interesse seguindo o Apêndice 12.1 do livro. O preço real é deflacionado pelo CPI; a renda é calculada em termos *per capita* e também deflacionada. Os instrumentos são construídos como a componente de imposto geral sobre vendas (*SalesTax*) e o imposto de excise total (*CigTax*), ambos em termos reais.

Como o objetivo é estimar a elasticidade de *longo prazo* da demanda, o livro utiliza **diferenças de 10 anos** (1995 – 1985) para cada estado. Essa abordagem elimina fatores fixos no tempo que variam entre estados (como tradições culturais de fumo), controlando para potenciais variáveis omitidas.[2]

```
Rows: 96
Columns: 12
Groups: state [48]
$ state        <fct> AL, AR, AZ, CA, CO, CT, DE, FL, GA, IA, ID, IL, IN, KS, K…
$ year         <fct> 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985, 1985, 198…
$ packs        <dbl> 116.4863, 128.5346, 104.5226, 100.3630, 112.9635, 109.278…
$ rprice       <dbl> 94.96438, 94.30762, 100.90962, 100.22058, 87.60842, 118.9…
$ rincome      <dbl> 10.763866, 10.468165, 12.830456, 15.713321, 14.326190, 17…
$ salestax     <dbl> 0.7884121, 0.0000000, 4.8052211, 5.6728627, 0.0000000, 8.…
$ cigtax       <dbl> 30.20447, 34.38662, 28.81041, 24.16357, 28.81041, 39.0334…
$ d_ln_packs   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…
$ d_ln_rprice  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…
$ d_ln_rincome <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…
$ d_salestax   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…
$ d_cigtax     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…
```

Análise Exploratória

O problema da causalidade simultânea

A Figura 1 mostra o diagrama de dispersão entre $\Delta \ln(P^{cigarros})$ e $\Delta \ln(Q^{cigarros})$ para os 48 estados. Intuitivamente, esperaríamos uma relação negativa: estados com maiores aumentos de preço deveriam ter maiores quedas no consumo. A linha de MQO mostra justamente isso — mas o estimador de MQO não é consistente neste caso.

[2] Essa estratégia é análoga à regressão em diferenças com $T = 2$ períodos discutida no Capítulo 10 e também pode ser implementada como um modelo de efeitos fixos com *dummies* de estados. A diferença de 10 anos captura o efeito de longo prazo porque cigarros são aditivos: no curto prazo, consumidores respondem pouco a variações de preço, mas no longo prazo a resposta é maior.

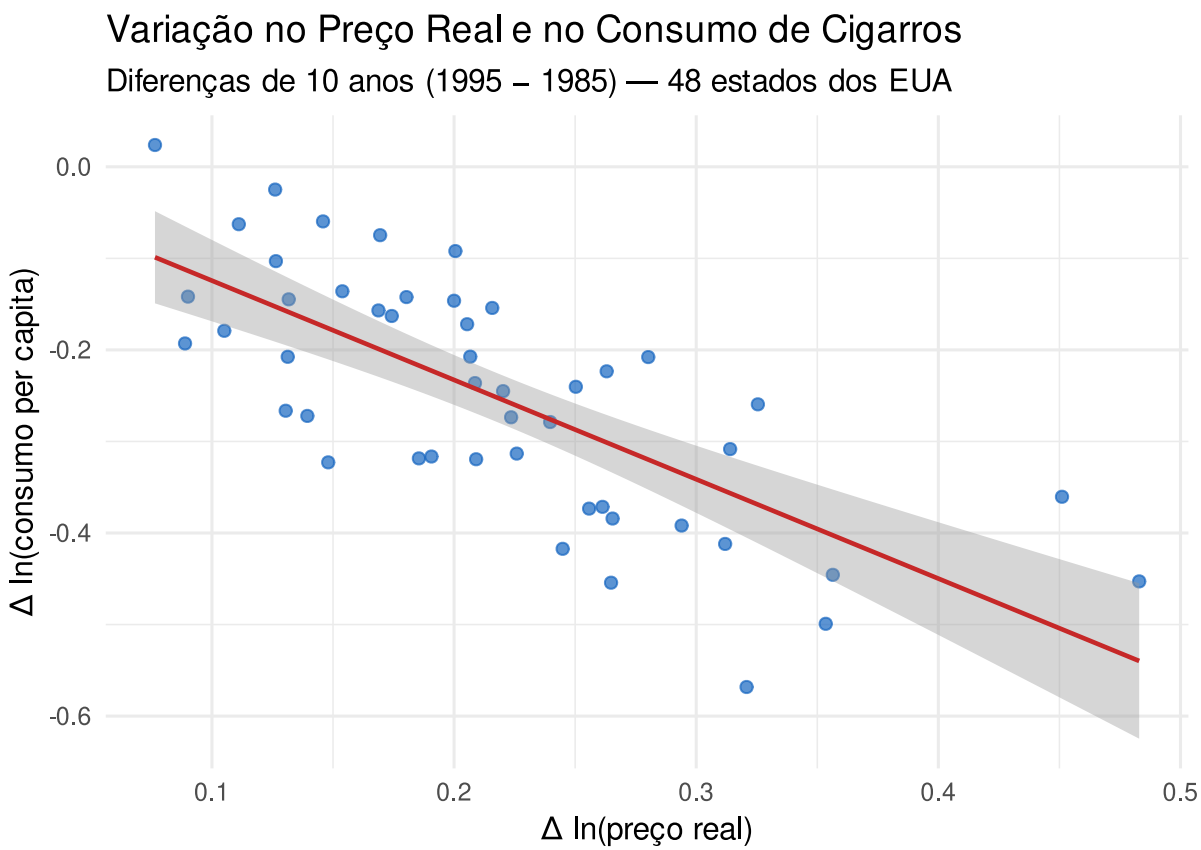


Figura 1: Variação do preço real e do consumo de cigarros entre 1985 e 1995

Por que MQO não funciona? Quando a demanda por cigarros aumenta (por exemplo, em estados onde fumar é mais aceito culturalmente), os produtores respondem com preços mais altos. Isso significa que o preço P_i é correlacionado com o erro u_i da equação de demanda. Com isso, a hipótese de exogeneidade do MQO é violada. O estimador de MQO captura uma *mistura* das elasticidades de demanda e oferta e não apenas a elasticidade da demanda.

Relação dos instrumentos com o preço

A Figura 2 mostra a correlação entre os dois instrumentos e a variação no preço real. Um instrumento válido precisa ser *relevante* (correlacionado com o preço) e *exógeno* (não correlacionado com o erro da demanda).

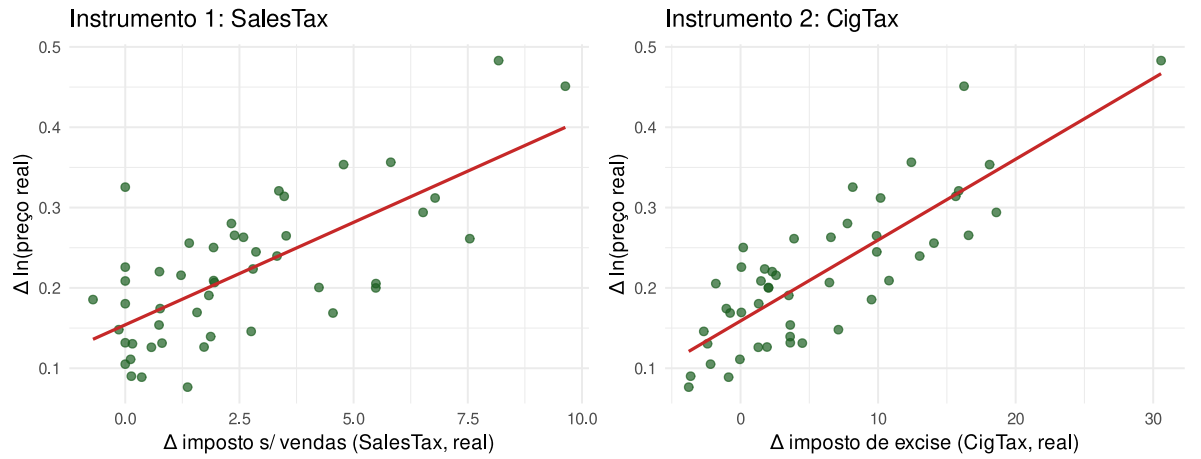


Figura 2: Correlação entre os instrumentos e a variação no preço real

Ambos os instrumentos mostram correlação positiva com o preço: estados com maiores aumentos de impostos tiveram maiores aumentos de preço. Isso é consistente com a condição de **relevância do instrumento**.

Variáveis Instrumentais e MQ2E

Conceito

Quando o regressor X_i é correlacionado com o erro u_i, uma **variável instrumental** Z_i permite isolar a parte de X_i que é exógena. Um instrumento válido deve satisfazer duas condições:

1. **Relevância:** $\text{corr}(Z_i, X_i) \neq 0$: o instrumento deve ser correlacionado com o regressor endógeno.
2. **Exogeneidade:** $\text{corr}(Z_i, u_i) = 0$: o instrumento não deve ser correlacionado com o erro da equação de interesse.

No nosso caso, os impostos sobre cigarros afetam o preço (relevância), e variam principalmente por razões políticas e fiscais alheias à demanda por cigarros (exogeneidade — ao menos como hipótese de trabalho).

Estimador de Mínimos Quadrados em Dois Estágios (MQ2E)

O **MQ2E** (do inglês *Two-Stage Least Squares*) é o estimador mais comum de VI e funciona em dois passos:

Primeiro estágio: Regride o regressor endógeno X_i sobre o(s) instrumento(s) e as variáveis exógenas incluídas:

$$X_i = \pi_0 + \pi_1 Z_i + \pi_2 W_i + v_i$$

Segundo estágio: Substitui X_i pelo seu valor ajustado $\hat{X}_i$ do primeiro estágio e regride Y_i sobre $\hat{X}_i$ e W_i:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 W_i + u_i$$

O valor predito $\hat{X}_i$ contém apenas a parte de X_i explicada pelo instrumento — que, se o instrumento for exógeno, é a parte não correlacionada com u_i. O estimador resultante é consistente.

Em termos formais, com um único regressor e instrumento, o estimador MQ2E tem a forma:

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)} = \beta_1$$

Na prática, utilizamos o pacote `fixest` com a função `feols()`. A sintaxe para VI é:

```
feols(Y ~ W | X ~ Z, data = dados, vcov = "hetero")
```

onde `W` são os controles exógenos, `X` é o regressor endógeno e `Z` é o instrumento.

Estimação em Corte Transversal (1995)

Antes de usar o painel de diferenças, examinamos o corte transversal de 1995. Isso reproduz os resultados da Seção 12.1 e permite comparar MQO com MQ2E.

Primeiro estágio (1995)

O primeiro estágio regride o log do preço real sobre o imposto de vendas (`salestax`). Isso verifica se o instrumento é de fato correlacionado com o preço.

```
t test of coefficients:

             Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 4.6165463  0.0289177 159.6444 < 2.2e-16 ***
salestax    0.0307289  0.0048354   6.3549 8.489e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

O coeficiente de `salestax` é positivo e altamente significativo (0,031, EP = 0,005), e o R^2 do primeiro estágio é de 47.1%. Isso reproduz a Equação (12.9) do livro: estados com alíquota de imposto sobre vendas maior cobram preços mais altos por maços de cigarros.

MQO vs. MQ2E (1995)

Comparamos o estimador de MQO (inconsistente) com o MQ2E (consistente):

Tabela 1: MQO e MQ2E — Corte Transversal 1995

Elasticidade da Demanda por Cigarros — Corte Transversal 1995

Variável dependente: ln(pacotes vendidos *per capita*)

	MQO	MQ2E (s/ renda)	MQ2E (c/ renda)
ln(preço real)	−1.213**	−1.084**	−1.143**
	(0.195)	(0.319)	(0.372)
ln(renda real p.c.)			0.215
			(0.312)
Constante	10.339**	9.720**	9.431**
	(0.935)	(1.528)	(1.259)
Observações	48	48	48

* p < 0.05, ** p < 0.01

Erros-padrão robustos (HC). * p < 0,05; ** p < 0,01. Instrumento: SalesTax.

Leitura dos resultados: O estimador de MQO produz elasticidade de −1,21, enquanto o MQ2E com instrumento *SalesTax* (sem controle de renda) fornece −1,08. Com a inclusão da renda, o MQ2E retorna −1,14 (Equação 12.15 do livro). Note que ao utilizar a função `feols`com variáveis instrumentais os erros-padrão já são ajustados para levar em conta o fato de que usamos os valores preditos do primeiro estágio para obter o coeficiente.[3]

Estimação com Dados em Painel — Diferenças de 10 Anos

Estratégia de identificação

O painel de diferenças elimina fatores fixos entre estados. Os instrumentos são as **variações de 10 anos** nos impostos:

- $\Delta SalesTax_i = SalesTax_{i,1995} - SalesTax_{i,1985}$: variação no imposto geral sobre vendas
- $\Delta CigTax_i = CigTax_{i,1995} - CigTax_{i,1985}$: variação no imposto de excise total sobre cigarros

Esses impostos variam entre estados principalmente por decisões de política fiscal não relacionadas à demanda por cigarros, tornando-os candidatos plausíveis a instrumentos exógenos.

[3] Além disso, a opção `vcov="hetero"` é o mesmo que `vcov="HC1"`, ou seja, é a opção a ser utilizada para estimadores robustos à heterocedasticidade. Veja o manual do pacote `fixest` as opções da função `feols`.

Primeiro estágio e diagnóstico de instrumento fraco

Antes do MQ2E, estimamos o primeiro estágio manualmente para obter o **F-estatístico do primeiro estágio**, que é o teste padrão para instrumentos fracos (ver seção 12.3 do livro).

> **Regra prática:** Um F-estatístico do primeiro estágio abaixo de 10 indica instrumento fraco. Com instrumentos fracos, o MQ2E é viesado em amostras finitas e os intervalos de confiança são não confiáveis.

Tabela 2: Primeiro Estágio — Variação no Log do Preço Real

Primeiro Estágio — MQ2E

Diferenças de 10 anos (1995 − 1985) — 48 estados

	Instrumento: SalesTax	Instrumento: CigTax	Ambos os instrumentos
Δ SalesTax	0.025**		0.013**
	(0.004)		(0.003)
Δ CigTax		0.010**	0.008**
		(0.001)	(0.001)
Δ ln(renda real p.c.)	−0.224	0.029	−0.029
	(0.219)	(0.131)	(0.124)
Constante	0.184**	0.155**	0.144**
	(0.030)	(0.020)	(0.018)
Observações	48	48	48
F-estat. 1º estágio (HC1)	33.7	107.2	88.6

* p < 0.05, ** p < 0.01

Variável dependente: Δ ln(preço real). Erros-padrão robustos (HC1). * p < 0,05; ** p < 0,01.

Os três F-estatísticos (33,7; 107,2; 88,6) estão bem acima de 10, confirmando que os instrumentos não são fracos. O imposto sobre cigarro (*CigTax*) é o instrumento individualmente mais forte: faz sentido, pois é um imposto diretamente sobre cigarros e varia mais entre estados.

Replicando a Tabela 12.1

Estimamos três especificações por MQ2E, correspondendo às três colunas da Tabela 12.1 do livro. O modelo de interesse é:

$$\Delta \ln(Q_i^{cig}) = \beta_0 + \beta_1 \Delta \ln(P_i^{cig}) + \beta_2 \Delta \ln(Inc_i) + u_i$$

onde $\Delta \ln(P_i^{cig})$ é endógeno e instrumentado pelos impostos.

Tabela 3: MQ2E da Demanda por Cigarros — Painel de 48 Estados (1985–1995)

Tabela 12.1 — Estimativas MQ2E da Demanda por Cigarros

Variável dependente: Δ ln(*Q*) = Δ ln(pacotes vendidos *per capita*)

	(1) SalesTax	(2) CigTax	(3) Ambos
Δ ln(preço real)	−0.94**	−1.34**	−1.20**
	(0.21)	(0.23)	(0.20)
Δ ln(renda real p.c.)	0.53	0.43	0.46
	(0.34)	(0.30)	(0.31)
Constante	−0.12	−0.02	−0.05
	(0.07)	(0.07)	(0.06)
Observações	48	48	48
Instrumento(s)	SalesTax	CigTax	SalesTax + CigTax
F-estat. 1º estágio (HC1)	33,7	107,2	88,6
Teste J (p-valor)	—	—	4.84 (1)

Tabela 12.1 — Estimativas MQ2E da Demanda por Cigarros

Variável dependente: Δ ln(*Q*) = Δ ln(pacotes vendidos *per capita*)

	(1) SalesTax	(2) CigTax	(3) Ambos

* p < 0.05, ** p < 0.01

Erros-padrão robustos (HC) entre parênteses. * p < 0,05; ** p < 0,01. Variável dependente: Δ ln(pacotes vendidos per capita). Diferenças de 10 anos (1995 − 1985), 48 estados.

Interpretação dos coeficientes

Elasticidade-preço da demanda: O coeficiente de $\Delta \ln(P)$ é a elasticidade de longo prazo da demanda por cigarros.

- Coluna (1), instrumento *SalesTax*: elasticidade = −0,94
- Coluna (2), instrumento *CigTax*: elasticidade = −1,34
- Coluna (3), ambos os instrumentos: elasticidade = −1,20

Em todos os casos, a demanda por cigarros é **elástica no longo prazo** ($|\beta_1| > 1$): um aumento de preço de 1% reduz o consumo em mais de 1%. Isso pode surpreender dado o caráter aditivo do produto, mas lembre-se que estamos medindo a resposta *de longo prazo.*A hipótese aqui é que dez anos são tempo suficiente para mudanças de comportamento.

Elasticidade-renda: O coeficiente da renda é positivo em todas as especificações (cigarros são um bem normal), mas não é estatisticamente significativo ao nível de 5%.

Verificação da Validade dos Instrumentos

Condição 1 — Relevância

A relevância é verificada pelo F-estatístico do primeiro estágio:

Instrumento	F-estat. 1º estágio	Conclusão
SalesTax	33,7	Não fraco (F > 10) ✓
CigTax	107,2	Não fraco (F > 10) ✓
Ambos	88,6	Não fraco (F > 10) ✓

Condição 2 — Exogeneidade e Teste J

A exogeneidade não pode ser testada diretamente quando o modelo é exatamente identificado (colunas 1 e 2). Quando há mais instrumentos do que variáveis endógenas (*sobreidentificação*), como na coluna (3), podemos aplicar o **Teste J de restrições sobreidentificadoras** (ver seção 12.3 do livro).

Ideia do teste: Se ambos os instrumentos são exógenos, as estimativas de MQ2E usando cada instrumento individualmente devem ser próximas. Se forem muito diferentes, pelo menos um instrumento viola a exogeneidade.

```
Sargan: stat = 4.8380, p = 0.027838, on 1 DoF.
```

O Teste J retorna estatística de 4.84 com p-valor de 1, rejeitando a hipótese nula de que ambos os instrumentos são exógenos ao nível de 5%.

Isso levanta uma questão: as estimativas usando *SalesTax* (−0,94) e *CigTax* (−1,34) divergem bastante. O livro argumenta que o caso para a exogeneidade do *SalesTax* é mais forte, pois o imposto geral sobre vendas é determinado por considerações fiscais gerais do estado, não pela dinâmica específica do mercado de cigarros. Já o *CigTax* pode ser influenciado por lobbies da indústria tabagista. Portanto, a estimativa preferida do livro é a da **coluna (1)**: elasticidade de longo prazo ≈ −0,94.

Visualização: MQO vs. MQ2E

A Figura 3 compara visualmente as estimativas de MQO e das três especificações de MQ2E.

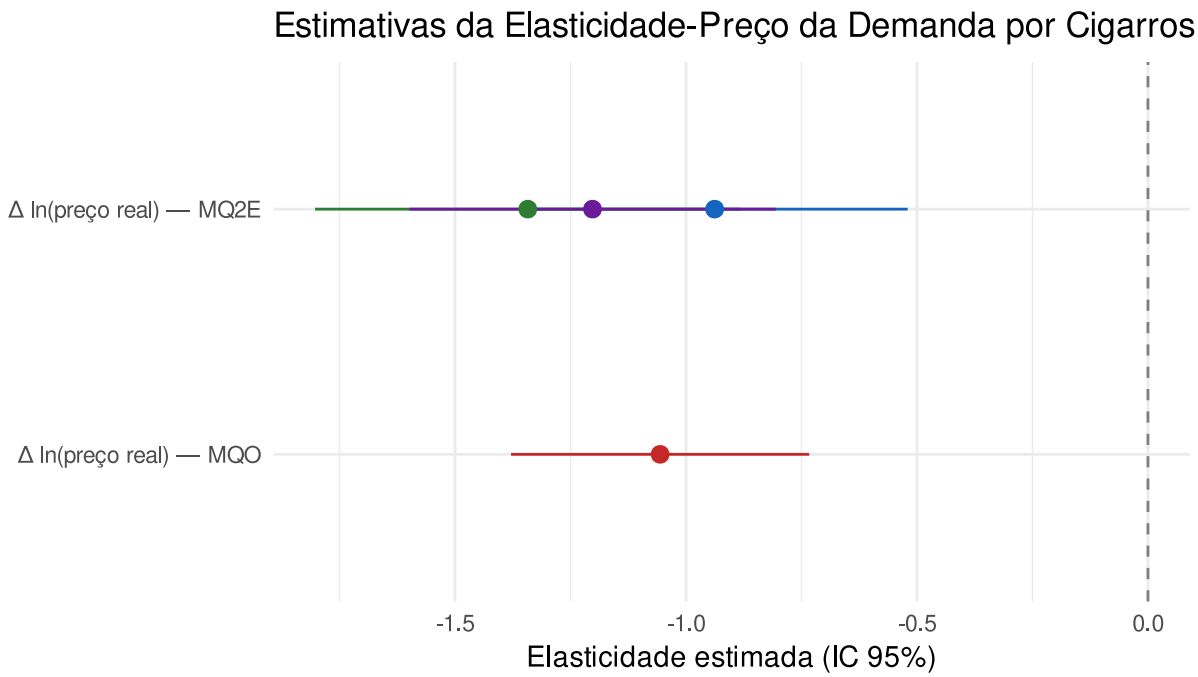


Figura 3: Estimativas da elasticidade-preço: MQO vs. MQ2E

O gráfico ilustra dois padrões importantes: (i) as estimativas de MQ2E são mais negativas do que a de MQO, sugerindo que o MQO subestima (em valor absoluto) a elasticidade; e (ii) os intervalos de confiança do MQ2E são mais largos, refletindo a perda de precisão ao usar apenas a variação exógena nos preços — mas com a vantagem de ter um modelo identificado com coeficiente consistente.

Atividade

Leia as Seções 12.1 a 12.5 do livro e responda às perguntas abaixo:

1. **Causalidade simultânea:** Explique com suas palavras por que o estimador de MQO da elasticidade da demanda por cigarros é inconsistente. Qual é o sentido esperado do viés? (Dica: pense no que acontece com o preço quando a demanda por cigarros sobe inesperadamente em um estado.)
2. **Condições do instrumento:** O imposto geral sobre vendas (*SalesTax*) é um instrumento plausível para o preço dos cigarros?
   a. Argumente em favor da *relevância*: por que o imposto sobre vendas deveria afetar o preço dos cigarros?
   b. Argumente em favor da *exogeneidade*: por que o imposto sobre vendas seria não correlacionado com outros fatores que determinam a demanda por cigarros?
   c. Você consegue pensar em algum cenário em que a exogeneidade seria violada?
3. **Interpretação econômica:** Com base na coluna (1) da Tabela 12.1, a elasticidade estimada é −0,94. Para reduzir o consumo de cigarros em 20%, quanto o preço precisaria subir? E se a elasticidade fosse −0,5 (mais inelástica)?
4. **Teste J e o dilema dos instrumentos:** O Teste J da coluna (3) rejeita a hipótese de que ambos os instrumentos são exógenos (p-valor ≈ 0,026). O livro conclui que *SalesTax* é o instrumento mais crível e descarta *CigTax*. Você concorda com esse argumento? Justifique.