

AULAS 01, 02 E 03 CAUSALIDADE

Ernesto F. L. Amaral

05, 07 e 12 de março de 2013

Técnicas Avançadas de Avaliação de Políticas Públicas (DCP 098)

Fonte:

Curso “Técnicas Econométricas para Avaliação de Impacto” do “International Policy Centre for Inclusive Growth” (IPC-IG) da “United Nations Development Programme” (UNDP) (<http://www.ipc-undp.org/evaluation>).

ESTRUTURA DA AULA

- Análise de causalidade.
- Problema da avaliação de impacto.

CORRELAÇÃO E CAUSALIDADE

- “Que correlação não é causalidade é talvez a primeira coisa que deve ser dito” (Barnand, 1982).
- Na economia, o debate sobre análise de causalidade tem como pano de fundo a decisão entre políticas.
- Para tomar uma decisão, os chamados policymakers devem entender minimamente como as coisas funcionam e, principalmente, qual o efeito que algumas mudanças podem acarretar.

PONDERAÇÕES

- Contudo, muitos estudos que visam a proposição de políticas usam métodos inapropriados para inferir sobre a chamada “causalidade” entre dois eventos.
- Confusão entre o que é uma associação (correlação) entre variáveis e o que pode ser inferido como uma relação causal.
- Frequentemente, verificamos em alguns estudos os seguintes avisos:
 - Apesar da análise parecer estar estritamente ligada à ideia de causalidade, ela, na realidade, não está.
 - A análise apresentada tem o objetivo de ser meramente descritiva.
 - Nenhuma inferência causal pode ser feita com base nos resultados.

CUIDADO COM REGRESSÕES

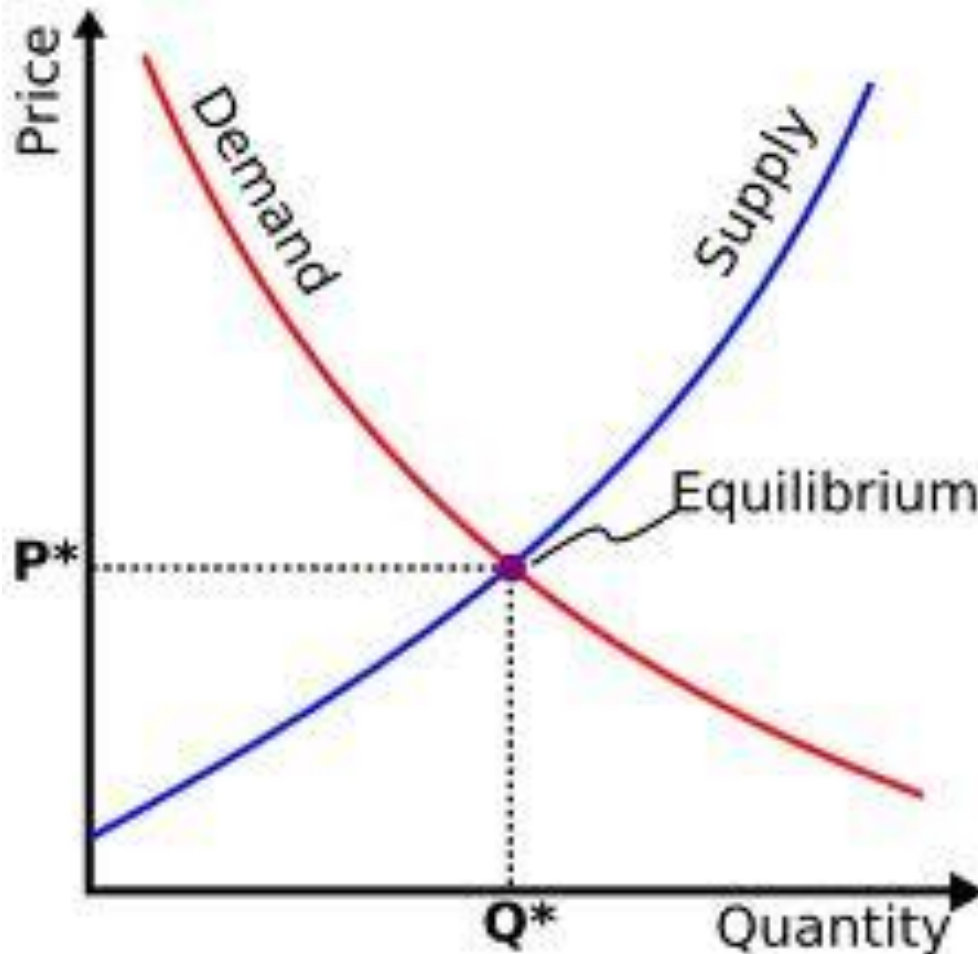
- É óbvio que análises descritivas apresentadas na forma de correlações, condicionais ou incondicionais, entre variáveis possuem o seu valor.
- Contudo, há um risco em ir muito longe com análises descritivas por meio de regressões ou outros métodos que estimam correlações condicionais. Elas são facilmente interpretadas como causalidades.
- Os próprios pesquisadores, muitas vezes, examinam suas regressões pensando na causalidade entre eventos, sem especular sobre os mecanismos implícitos na correlação.
- Exemplo: salários no setor público e no setor privado.

CAUSALIDADE NA ECONOMIA

- Atualmente, muito dos principais avanços na econometria é incentivado pela discussão sobre inferência causal (Imbens e Wooldridge).
- Quando se fala em modelos estatísticos de causalidade, geralmente os associamos ao trabalho de Rubin (1974).
- Contudo, a ideia frequentemente tida como original foi a de Neyman (1923, 1935).
- Na economia, os modelos de causalidade estão principalmente ligados aos nomes de Heckman e McFadden, ganhadores do Prêmio Nobel de 2000.

CURVA DE OFERTA E DEMANDA

- Oferta (*supply*): disposição para vender
- Demanda (*demand*): disposição para comprar



DEFINIÇÕES

- Dowd e Town (2002) apontam 5 conceitos de causalidade.
- **Causa:** quando um resultado deriva da ocorrência de um evento.
- **Determinação:** quando uma variável influencia o resultado de outra variável.
- **Antecedente:** quando um evento precede outro evento.
- **Razão:** processo relacionando a causa ao resultado.
- **Ocasão:** quando o evento que causa um resultado é acionado por uma conjuntura.

AVALIAÇÃO DE IMPACTO

- Avaliações de impacto se interessam pelas duas primeiras definições.
- Há análise do ponto de vista da variação no tratamento, mantendo os demais fatores constantes (*ceteris paribus*, variação controlada).
- A terceira definição está implícita na a ideia de predição através da antecedência (Granger, 1969; Sims, 1972).
- A quarta é de interesse em modelos estruturais (Heckman 2000, 2008).

MODELOS DE CORRELAÇÃO OU ASSOCIAÇÃO

- Em análises de correlação ou associação, o pesquisador está interessado e satisfeito em saber como os valores de uma variável de interesse Y estão associados com os valores de outras variáveis X .
- Probabilidades, esperanças e esperanças condicionais da variável Y podem ser calculadas sobre uma amostra, sem necessariamente inferir causalidade.
- Os parâmetros são determinados pela distribuição conjunta de Y e X , $\Pr(Y = y, X = x)$: proporção da amostra na qual $Y = y$ e $X = x$.

PROBABILIDADE CONDICIONAL

- A distribuição condicional de Y dado X descreve como a distribuição dos valores de Y muda quando X varia:

$$\Pr(Y = y \mid X = x) = \Pr(Y = y, X = x) / \Pr(X = x)$$

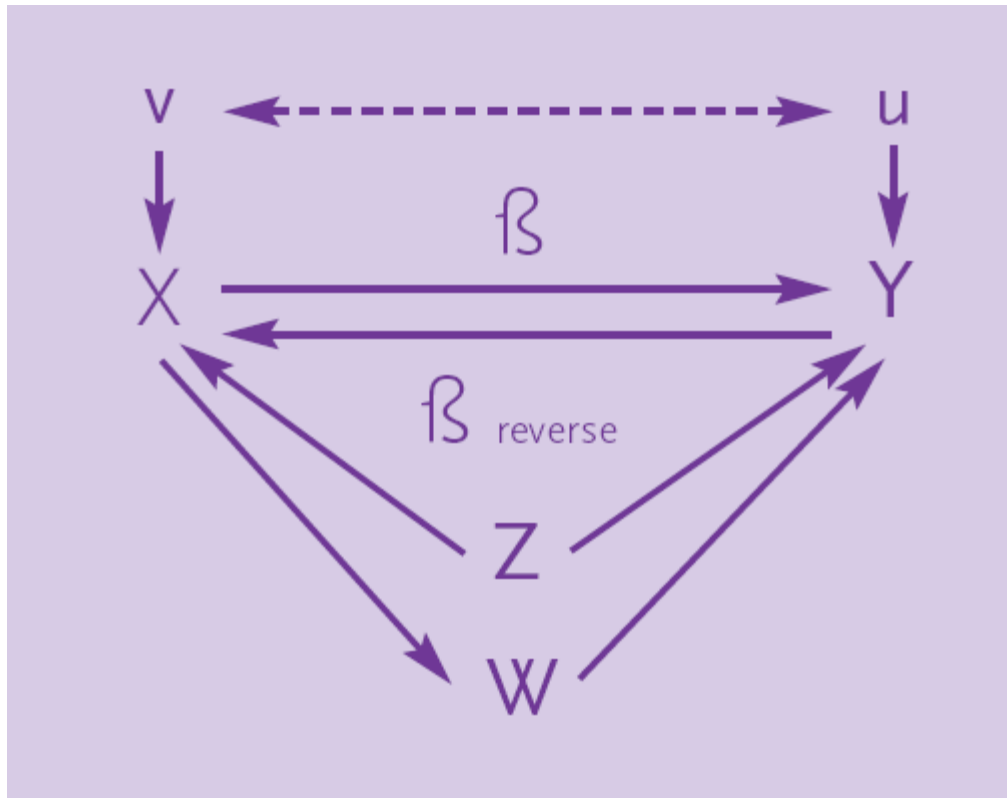
- No caso da regressão de Y sobre X ($Y = \beta_0 + \beta_1 X + u$), o parâmetro de associação pode ser descrito como uma variação em Y , dado uma variação em X :

$$\beta = E[Y \mid X = x] - E[Y \mid X = x - 1]$$

- Neste caso, a inferência é simplesmente uma análise descritiva.

CORRELAÇÃO NÃO IMPLICA CAUSALIDADE

- No mundo real, por trás de uma correlação entre Y e X , podemos ter a seguinte situação:



- X , Y , W e Z são variáveis observáveis e u e v representam características não-observáveis.
- A omissão da variável W pode não ser um problema, pois ela representa uma das formas na qual X causa Y e isso pode não ser de interesse do pesquisador.

PROBABILIDADE CONDICIONAL

- A omissão da variável Z é um problema, pois resulta na estimação de uma correlação espúria entre X e Y .
- Isto gera o chamado viés por omissão de variáveis, onde a variável X pode ser denominada como **endógena**.
- Outro problema na estimação é a causalidade reversa, que caracteriza X como **fortemente endógena**.
- Na causalidade de Granger, uma variável **endógena** é aquela correlacionada com os resíduos passados da variável Y .
- Já uma variável **fortemente endógena** é aquela correlacionada com os valores presentes e futuros de Y .

VARIÁVEIS OMITIDAS

- Se existe um Z que causa Y e este Z não está incluído no modelo, Z causa u .
- Se Z também causa X , u estará correlacionado com X .
- Intuitivamente, Z impõe um nível para X e outro para Y . A consequência é uma associação entre X e Y que não é necessariamente derivada de uma causalidade entre X e Y .
- A direção do viés depende se os efeitos de Z sobre X e Y são positivos ou negativos.

PENSANDO SOBRE VARIÁVEIS OMITIDAS

- A solução para estes casos são modelos que incorporam a correlação entre resíduos das funções de X e Y .
- A função u seria decomposta em uma parte explicada e outra não-explicada por X , tal que esta última atende à exigência de não ser correlacionada com X .
- Outra solução é o uso de modelos com efeitos fixos, estimados por meio de um painel. Contudo, se Z não é fixo (atributo), o modelo não corrige o viés.
- Modelos de efeitos fixos não resolvem todos problemas, mas resolvem problemas de correlação entre resíduos: transversal (*cross-section*), espacial, temporal.
- A variável W , interpretada como uma razão ou uma forma na qual X causa Y , não causa viés na análise. Mas implica uma limitação na extrapolação dos resultados.

SOLUÇÕES PARA VARIÁVEIS OMITIDAS

- 1) Coletar informações adicionais.
- 2) Manipular as variáveis independentes (X) de forma que seus efeitos sobre a variável dependente (Y) não estejam sendo influenciados por outras variáveis não observadas.
- 3) Modelar a correlação entre os termos de erro (v, u) das equações de X e Y , como parte do processo de estimação.

1) COLETA DE INFORMAÇÕES ADICIONAIS

- A coleta de informações adicionais, antes não observadas, pode ser uma solução.
- Contudo, isso pode impor custos adicionais no levantamento das informações sem um retorno esperado.
- Por mais que nossa ignorância sobre o mundo diminua, ainda resta muita coisa para se conhecer.

SATURAÇÃO DE MODELOS

- Além disso, saturar o modelo incluindo um número maior de variáveis pode não ser a solução, por três razões:
 - O modelo perde em graus de liberdade, o que reduz a eficiência dos estimadores.
 - Uma das novas variáveis incluídas pode ser um W que acabará explicando parte do efeito de X sobre Y .
 - Uma das variáveis incluídas pode derivar de uma causalidade reversa em relação a Y , enviesando completamente os demais parâmetros.

2) MANIPULAÇÃO DE X

- Manipulação de X pode ocorrer de diferentes formas:
 - Experimento.
 - Variável instrumental.

EXPERIMENTO

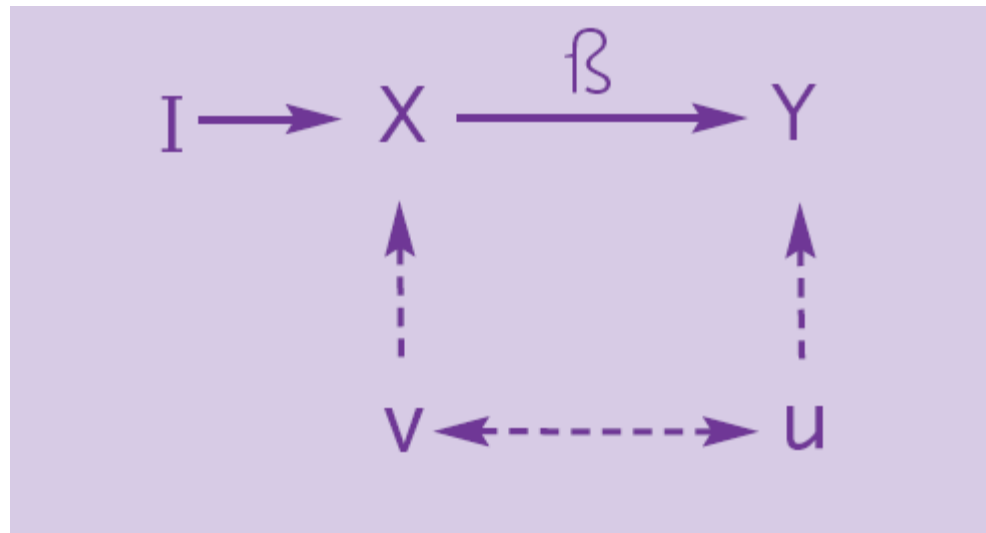
- O melhor mecanismo de manipulação de X seria a escolha aleatória dos grupos de tratamento e controle (experimento)
- Isso evitaria que outras variáveis não observáveis explicassem X e Y ao mesmo tempo.
- A escolha dos grupos não estaria associada com outros fatores que afetam Y.
- Problema ocorreria se indivíduos se recusassem a participar do experimento ou saíssem do experimento por razões ligadas a Y.

First, it often is impossible, as well as unethical, to assign subjects randomly to different values of the X variable

(Dowd e Town, 2002, p.18)

VARIÁVEL INSTRUMENTAL (*INSTRUMENTAL VARIABLE – IV*)

- Outra forma de manipular X seria o de identificar um instrumento (I) que seja correlacionado com X , mas que não tenha efeito direto sobre Y , além das mudanças induzidas em X .



- Pressuposto é que I afeta X , mas não está correlacionado com u , o que é difícil de verificar.
- É difícil identificar variáveis que afetam X , mas que não afetam Y .

AVALIANDO O EFEITO DO TRATAMENTO

- Ao aplicar valores de X em um processo de pseudo-aleatorização, é preciso escolher um método para avaliar o efeito do tratamento.
- Há duas principais formas:
 - Método de diferença em diferenças.
 - Modelos multivariados.

MÉTODO DE DIFERENÇA EM DIFERENÇAS

- Comparar efeito de X em Y para sujeitos com diferentes valores da variável instrumental (I).
- Ser do grupo de tratamento ou controle (X) não é uma decisão aleatória, mas sim um resultado do efeito da variação do instrumento (I).
- Por isso, nem todos sujeitos terão mesmo risco de serem alocados aos diferentes valores de X (tratamento ou controle).
- Esta pseudo-aleatorização resulta no efeito de X sobre Y para os sujeitos “marginais” (aqueles com o risco de receberem diferentes valores de X).

MODELOS MULTIVARIADOS

- Utilização de modelos multivariados com variáveis instrumentais (I) e outras variáveis independentes (R) para explicar X , com erro aleatório (v):

$$X = R\gamma + I\alpha + v$$

- Valores preditos de X :

$$X \text{ predito} = R\gamma^* + I\alpha^*$$

- O erro aleatório (v) não aparece acima porque há o pressuposto que tenha média zero [$E(v)=0$].
- O valor predito de X não tem o problema de correlação entre os erros aleatórios (v, u) das equações de X e Y .
- O valor predito de X é usado para estimar o efeito causal (β) em Y , em procedimento chamado de dois estágios de mínimos quadrados (*two-stage least squares* – 2SLS).

3) MODELAR CORRELAÇÃO ENTRE v E u

- Outra solução para variáveis omitidas é incorporar a correlação entre os termos de erro (v , u) na estimação do parâmetro causal (β).
- O problema essencial é que a média do termo de erro (u) não será igual a zero para indivíduos com diferentes valores em X .
- Há duas possibilidades principais:
 - Estimação em dois passos.
 - Estimação simultânea.

ESTIMAÇÃO EM DOIS PASSOS

- Quando X é dicotômica, são estimados modelos de seleção amostral (*sample selection models*).
- 1º passo:
 - Estimação da equação que explica X .
 - Resíduo é estimado, representando o termo de erro (u), no contexto de auto-seleção amostral.
- 2º passo:
 - Termo de erro é adicionado na equação que explica Y para corrigir o fato de que este erro, devido à auto-seleção da amostra, não tem média zero.
- Este tipo de modelagem é chamado de informação limitada de máxima verossimilhança (*limited information maximum likelihood* – LIML).

ESTIMAÇÃO SIMULTÂNEA

- Neste caso, há a estimação simultânea do modelo de auto-seleção amostral (X) e da equação da variável dependente principal (Y), utilizando estimador de máxima verossimilhança (*maximum likelihood estimator*).
- Este tipo de modelagem é chamado de informação completa de máxima verossimilhança (*full information maximum likelihood* – FIML).
- A variável X pode ser dicotômica ou contínua.
- Esse modelo utiliza o pressuposto de distribuição conjunta dos termos de erro (v, u).

CAUSALIDADE REVERSA

- Como no caso de variáveis omitidas, a causalidade reversa resulta em correlação de X com termo de erro (u).
- Relembremos as soluções para variáveis omitidas:
 - 1) Coletar informações adicionais:
 - Causalidade reversa não pode ser solucionada com coleta adicional de dados no decorrer do tempo.
 - 2) Manipular variáveis independentes (X):
 - Possível de ser aplicado para causalidade reversa.
 - 3) Modelar correlação entre termos de erro:
 - Causalidade reversa não pode ser solucionada com esta modelagem, porque viés ocorre mesmo se termos de erro não estão correlacionados entre as equações.

SOLUÇÃO PARA CAUSALIDADE REVERSA

- Por não ser causada pela omissão de variáveis, um modelo de efeitos fixos não corrige este tipo de viés.
- Precisamos manipular as variáveis independentes, com experimento ou variáveis instrumentais.
- Quando os termos de erro (v, u) não são correlacionados, um parâmetro β consistente é estimado por meio de modelos com duas equações (2SLS).
- Quando os termos de erro (v, u) são correlacionados, é preciso estimar um modelo de três estágios de mínimos quadrados (*three-stage least squares* – 3SLS):
 - Os dois primeiros estágios corrigem o viés em β .
 - O terceiro estágio corrige os erros padrão dos coeficientes, ao considerar a correlação entre os termos de erro (v, u).

ESTIMAÇÃO DE MODELOS CONTRAFACTUAIS

- A formulação de problemas de causalidade, em qualquer área, é baseada em análises contrafactuais.
- Contrafactuais são possíveis resultados em diferentes estados hipotéticos da natureza.
- Uma análise contrafactual busca contrastar resultados em diferentes estados, onde só a ocorrência do evento em questão diferencia estes estados.
- Mesmo um modelo de regressão linear simples pode ser interpretado como uma análise contrafactual, desde que o coeficiente sobre X seja interpretado como a diferença que ocorreria em Y se houvesse uma mudança exógena em X .

MAIS SOBRE MODELOS CONTRAFACTUAIS

- Heckman coloca que a definição de um conjunto de hipóteses ou, portanto, de contrafactuais é um exercício lógico.
- Um modelo contrafactual será mais aceito quanto mais aceito forem as regras que o define.
- Ou seja, isso depende da aceitação das premissas e da utilização de regras lógicas e matemáticas consistentes.
- Esta análise deve fornecer ainda a base para a especificação de fatores que possam ser creditados como exógenos ao resultado de interesse.

SUTVA

- Rubin (1986) aponta que uma condição necessária para identificação de um contrafactual é a Suposição de Valor Estável da Unidade de Tratamento (*Stable-Unit-Treatment-Value Assumption, SUTVA*).
- O fato de uma unidade receber o tratamento não afeta o resultado potencial de uma unidade que não o recebeu.
- Quando exposto a um tratamento (s), pressuposto é que o resultado Y de um indivíduo será o mesmo, não importando o mecanismo de seleção e o tratamento das outras unidades: $Y(0) \perp s$
- SUTVA pode ser violado quando existem outras versões não representadas de tratamento ou quando há interação entre os indivíduos.
- SUTVA: suposição de não-confundimento/ignorabilidade.

MAIS PRESSUPOSTOS

- Esse pressuposto leva a outros pressupostos também necessários, os quais são mais difíceis de satisfazer (Holland, 1986):
 - Estabilidade temporal e transitoriedade causal.
 - Homogeneidade das unidades investigadas.
 - Independência do tratamento.
 - Efeito constante.
- Outro pressuposto necessário para identificação de um contrafactual é que exista um grupo de comparação ou controle.

PROBLEMA FUNDAMENTAL DA AVALIAÇÃO DE IMPACTO

- O problema é que uma análise contrafactual, quase que por definição, só pode ser conduzida sobre um conjunto informacional incompleto:
 - Cada indivíduo é observado em somente um dos grupos (tratamento ou controle).
 - Seu contrafactual não existe (*missing*).
- Além disso, existe o problema de seleção, quando somente parte dos indivíduos é observada sob determinado tratamento:
 - O resultado pós-tratamento só é observado entre os que escolheram ou foram escolhidos para receber o tratamento.

PRECISAMOS DE PRESSUPOSTOS

- O resultado representativo de um grupo de tratamento pode não convergir com o resultado representativo da população como um todo.
- Conseqüentemente, sem a aplicação de suposições ou restrições (não-testáveis), a construção empírica de um contrafactual é impossível.
- Outro exemplo clássico é o modelo de Roy (1951), em que o trabalhador só escolhe o setor com maior renda.
- O problema de seleção sempre surgirá quando os dados são gerados de acordo com a escolha dos agentes.
- Sejam eles os próprios indivíduos ou os gestores de um programa social.

NEGLIGENCIANDO FATORES NÃO-OBSERVÁVEIS

- Negligenciar fatores não-observados significa supor que os mesmos não possuem efeito sobre a diferença nos possíveis resultados para um mesmo indivíduo.
- Isto também pode ser chamado de seleção sobre variáveis observáveis.
- Uma condição necessária para a identificação de causalidade em um modelo de seleção sobre variáveis observáveis (X) é uma versão condicional da SUTVA, onde:

$$Y(0) \perp s \mid X$$

- Isso implica uma independência condicional de $Y(0)$ e o tratamento.

INDEPENDÊNCIA DA MÉDIA CONDICIONAL

- Há ainda a suposição de independência da média condicional.
- O valor de Y é semelhante entre o grupo de tratamento [$D(s)=1$] e o grupo de controle [$D(s)=0$], controlando pelos valores de X :

$$E[Y(0) | D(s) = 1, X] = E[Y(0) | D(s) = 0, X] = E[Y(0) | X]$$

- Além disso, é necessário que para cada valor de X , existe tanto um caso tratado pela política (s) quanto um caso não-tratado por s :

$$0 < \Pr(D(s) = 1 | X) < 1$$

MODELOS DE EQUAÇÕES ESTRUTURAIS

- Estimações na forma reduzida são suficientes e exigem menos suposições que os modelos estruturais.
- Estimativas não-enviesadas são mais factíveis, ao custo de não aprender muito sobre o processo investigado.
- Muitos modelos causais são incompletos por não especificarem os mecanismos por trás da causalidade.
- A crítica é particularmente voltada aos estudos experimentais e análises de séries temporais, afirmando que seus resultados não podem ser extrapolados (eventos futuros, ocorrência em outros contextos).
- O objetivo dos modelos estruturais é investigar a caixa preta do modelo causal, estudando não só o “efeito das causas” mas também a “causa dos efeitos”.

Avaliação:

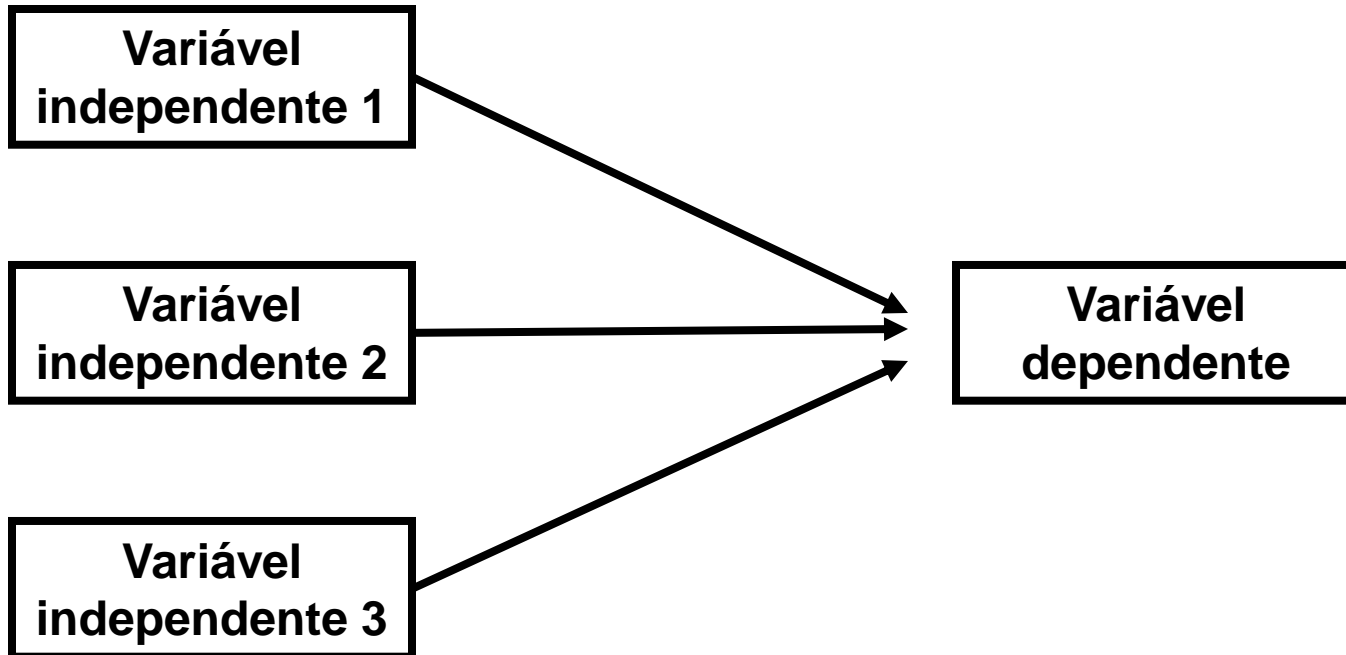
conceito e especificidade

Cohen e Franco, 2000: 72-84 (capítulo 4)

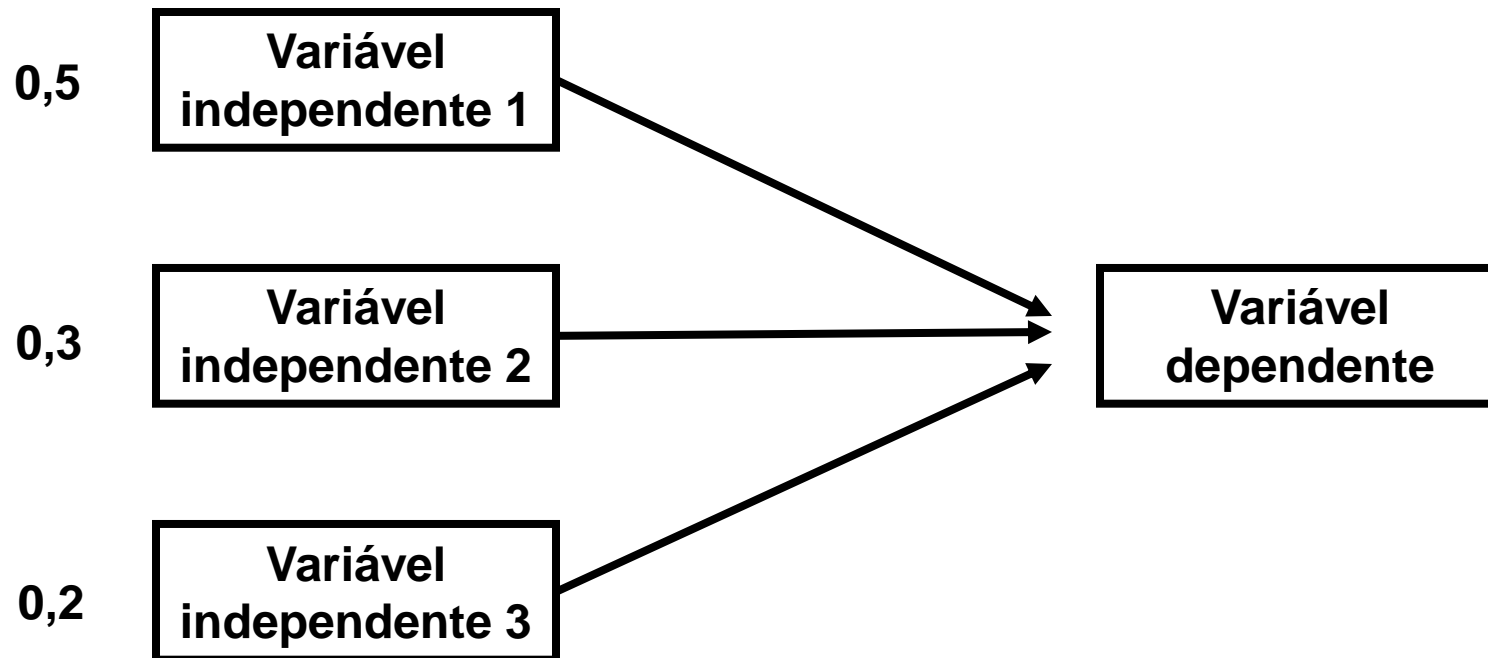
MODELOS EXPLICATIVOS

- O importante na pesquisa é estabelecer os modelos de inter-relação entre as variáveis.
- O modelo explicativo (modelo causal) é um roteiro para alocação de recursos do projeto.
- Cada variável instrumental supõe ações específicas que requerem insumos com diferentes custos.
- Exemplos de tipos de modelos:
 - 1) Modelo de variáveis independentes equiponderadas.
 - 2) Modelo de variáveis independentes desigualmente ponderadas.
 - 3) Modelo de variáveis independentes que são estatisticamente dependentes.
 - 4) Modelo de retroalimentação dinâmica.

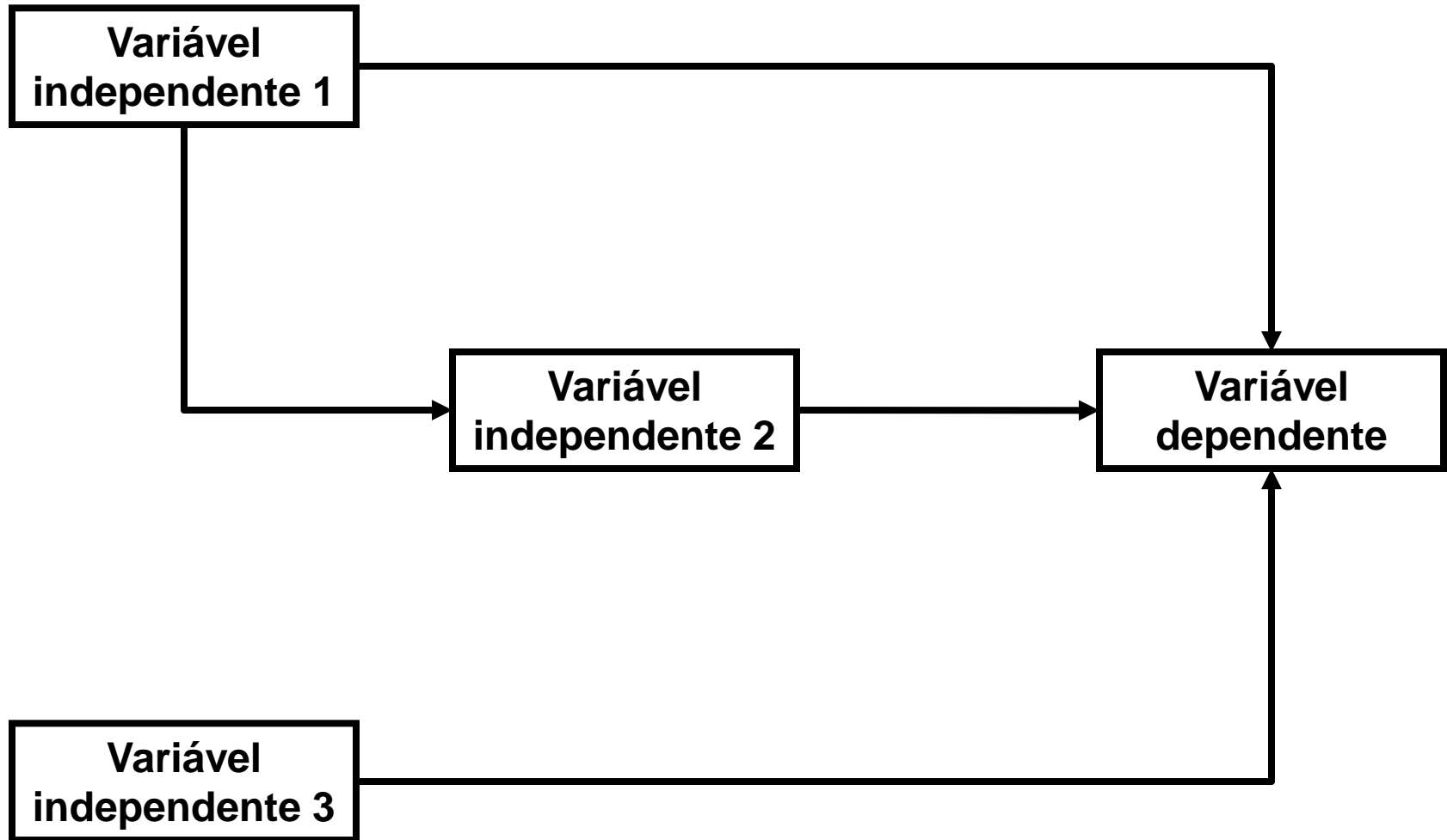
MODELO DE VARIÁVEIS INDEPENDENTES EQUIPONDERADAS



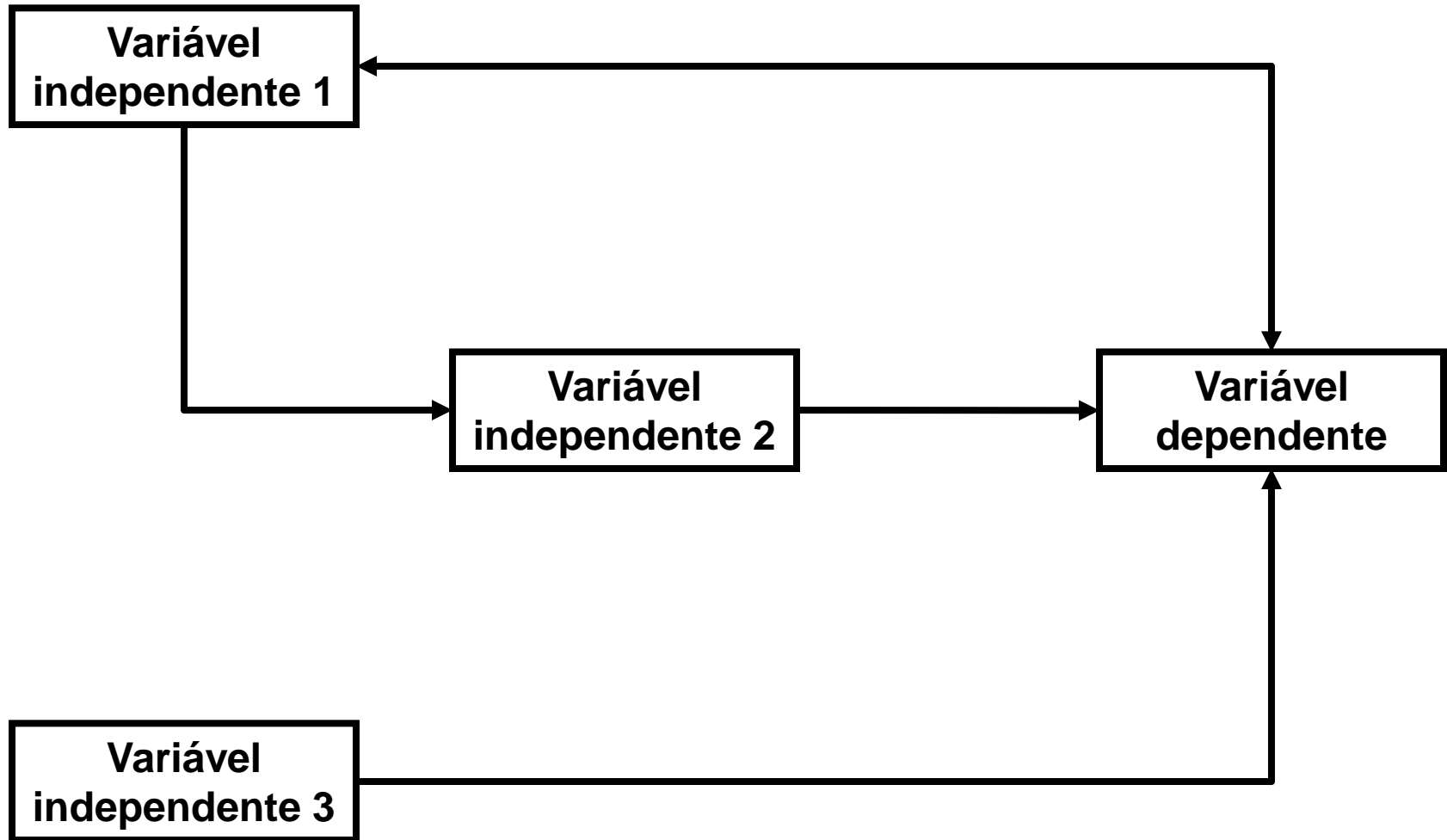
MODELO DE VARIÁVEIS INDEPENDENTES DESIGUALMENTE PONDERADAS



MODELO DE VARIÁVEIS INDEPENDENTES QUE SÃO ESTATISTICAMENTE DEPENDENTES



MODELO DE RETROALIMENTAÇÃO DINÂMICA



Tipos de desenhos de pesquisa

Babbie, 1999: 93-111 (capítulo 4)

TIPOS DE DESENHOS DE PESQUISA

- Há uma tendência a considerar a análise de dados de *survey* como mais desafiadora e interessante do que o desenho do *survey* e a coleta de dados.
- Durante a análise, começa-se a ganhar entendimento do assunto estudado e pode-se compartilhar as descobertas com colegas.
- No entanto, é preciso que haja a correta elaboração do desenho de pesquisa e sua execução, para que não haja problemas na etapa da análise.
- Portanto, o desenho é igualmente desafiador e importante.
- Há vários tipos de pesquisa de *survey*, com objetivos, custos, tempo e escopo diferenciados.

FINALIDADES DA PESQUISA DE *SURVEY*

- Uma pesquisa de *survey* pode ser realizada para atender diferentes objetivos:
 - Intenção de voto.
 - Venda de produtos.
 - Projetar infra-estrutura.
 - Modificar programa social.
- De uma forma geral, três objetivos gerais permeiam os interesses de pesquisas de *survey*:
 - Descrição.
 - Explicação.
 - Exploração.

DESCRIÇÃO

- *Surveys* são frequentemente realizados para descobrir a distribuição de certos traços e atributos da população, permitindo análise descritiva.
- Há preocupação em **como** é a distribuição de determinada variável (descrição), e não no **porquê** (explicação).
- Exemplos de dados descritivos:
 - Distribuições de candidatos por idade e sexo.
 - Taxas de aprovação do governo.
 - Percentual de políticos na escala esquerda/direita.
- Além de descrever a amostra total (e inferir para a população total), pesquisadores descrevem subgrupos.

EXPLICAÇÃO

- Muitas pesquisas de *survey* têm o objetivo adicional de fazer análises explicativas sobre a população, procurando relações de causa e efeito.
- Explicar quase sempre requer análise multivariada, que é a análise do impacto simultâneo de duas ou mais variáveis (independentes) sobre outra variável (dependente).
- Há uma preocupação em explicar o porquê de determinada distribuição da variável, geralmente utilizando modelos estatísticos.
- Por exemplo, podemos querer explicar escala esquerda/direita de deputados por partido de pertencimento, Estado de residência, partido do governador...

EXPLORAÇÃO

- Pesquisadores podem fazer pesquisa sobre um tema que está começando a ser investigado, e não se sabe exatamente suas causas e consequências.
- Primeiramente, pode ser realizada entrevista em profundidade (qualitativa) para captar o máximo de informações possíveis do público em estudo:
 - Não é necessário ter amostra representativa.
 - Não é preciso coletar dados com questionário padronizado.
- Posteriormente, as informações são analisadas para melhor entender os indivíduos, sem preocupação explicativa, mas visando a elaboração de questionário padronizado.

Econometria

Wooldridge, 2008: 1-17 (capítulo 1)

ECONOMETRIA

- A econometria evoluiu como uma disciplina separada da estatística matemática, porque enfoca problemas inerentes à coleta e à análise de dados econômicos não-experimentais.
- **Dados não-experimentais** não são acumulados por meio de experimentos controlados de indivíduos, firmas ou segmentos da economia.
- Dados não-experimentais são também chamados de **dados observacionais** para enfatizar o fato de que o pesquisador é um coletor passivo de dados.
- **Dados experimentais** são frequentemente coletados em ambientes de laboratório nas ciências naturais, mas são muito mais difíceis de serem obtidos nas ciências sociais.
- O método de análise da **regressão múltipla** é utilizado por econometristas e estatísticos matemáticos, mas o foco e interpretação pode diferir significativamente.

ANÁLISE ECONÔMICA EMPÍRICA

- Os métodos econométricos são usados para testar uma teoria econômica ou para analisar relações que apresentam importância para análises de políticas públicas.
- Uma análise empírica usa dados para testar uma teoria ou estimar uma relação.
- O primeiro passo em qualquer análise empírica é a formulação cuidadosa da questão de interesse, a qual pode ser a de testar efeitos de uma política governamental ou, até mesmo, de testar hipóteses e teorias.
- O modelo econômico formal consiste em equações matemáticas que descrevem relações para testar teorias.

MICROECONOMIA

- Os indivíduos fazem escolhas para maximizar seu bem-estar (**maximização da utilidade**), sujeitas às restrições de recursos.
- Isso oferece um arcabouço para criar modelos econômicos para fazer previsões entre variáveis.
- A maximização da utilidade leva a um conjunto de **equações de demanda**, no contexto das decisões de consumo.
- Em uma equação de demanda, a quantidade demandada de cada produto depende do seu próprio preço, do preço dos bens substitutos e complementares, da renda do consumidor e das características individuais que influem no gosto.

MODELO ECONÔMICO

- O modelo econômico é a formulação teórica de uma relação entre variáveis econômicas.
- A quantidade de tempo gasto na atividade criminosa é uma função de vários fatores (Gary Becker 1968):

$$y=f(x_1, x_2, x_3, x_4, x_5, x_6, x_7),$$

y = horas gastas em atividades criminosas.

x_1 = “salário” por hora ocupada em atividade criminosa.

x_2 = salário-hora em emprego legal.

x_3 = renda de outras atividades que não o crime ou um emprego legal.

x_4 = probabilidade de ser capturado.

x_5 = probabilidade de ser condenado se capturado.

x_6 = sentença esperada se condenado.

x_7 = idade.

MODELO ECONOMETRICO

- Após elaborar o modelo econômico, é especificado um modelo econométrico, que será aplicado a dados existentes.
- A forma da função $f(.)$ deveria ser especificada antes de realizar uma análise econométrica.
- Se uma variável não pode ser obtida, é possível utilizar uma variável que se aproxima desta que se quer medir (**proxy**).
- Outros fatores são considerados no termo de erro u (ou termo de distúrbância):
 - **Erro amostral** é a diferença entre o resultado amostral e o verdadeiro resultado da população (devidos ao acaso).
 - **Erro não-amostral** ocorre quando os dados amostrais são coletados, registrados ou analisados incorretamente.
- Modelo econométrico de Becker (1968):

$$\textit{crime} = \beta_0 + \beta_1 \textit{salário} + \beta_2 \textit{outrenda} + \beta_3 \textit{freqpris} + \beta_4 \textit{freqcond} + \beta_5 \textit{sentmed} + \beta_6 \textit{idade} + u$$

MODELO ECONOMÉTRICO NA PRÁTICA

- Na maioria dos casos, a análise econométrica começa pela especificação de um modelo econométrico, sem consideração de detalhes da criação do modelo econômico.
- É comum começar com um modelo econométrico e usar o raciocínio econômico e conhecimentos científicos como guias para escolher as variáveis.
- Após a especificação do modelo econométrico, várias hipóteses podem ser formuladas em termos das direções e influências dos parâmetros desconhecidos (independentes) sobre a variável de interesse (dependente).
- Após os dados terem sido coletados, os métodos econométricos são usados para estimar os parâmetros do modelo econométrico e para testar as hipóteses de interesse.