Econometria I

O modelo econométrico linear

Luis A. F. Alvarez

23 de março de 2025

Anatomia de um problema

- Nosso ponto de partida é um espaço de probabilidade $(\Omega, \Sigma, \mathbb{P})$, e um par (Y, X), onde Y é uma variável aleatória real, e X é um vetor aleatório de dimensão k.
- A lei de probabilidade $\mathbb{P}_{Y,X}$ induzida pelo par (Y, X) sobre $(\mathbb{R}^{k+1}, \mathcal{B}(\mathbb{R}^{k+1}))$ será interpretada como representativa da distribuição de características (Y, X) numa população de interesse.
- Por exemplo, se estamos interessados na relação entre salário (Y) e anos de estudo (X) na população paulista, o experimento $(\Omega, \Sigma, \mathbb{P})$ representa a amostragem de uma unidade ω desta população, de modo que $\mathbb{P}_{Y,X}$ reflete a distribuição conjunta de salário e anos de estudo nessa população (e $(X(\omega), Y(\omega))$ é o valor dessa característica para o indivíduo ω).

O conceito de população

- Em alguns casos, a noção de "população" é diferente àquela convencional.
- Por exemplo, o que é a população se X representa a taxa de inflação em um mês t, e Y representa a taxa de inflação no mês $t+1$?
  - Nesse caso, podemos pensar no experimento $(\Omega, \Sigma, \mathbb{P})$ como refletindo **incerteza econômica**, de modo que cada elemento $\omega \in \Omega$ representa uma possível realização dos fenômenos econômicos relevantes à determinação da taxa inflação (um possível cenário econômico), e $(X(\omega), Y(\omega))$ as taxas de inflação sob ω.
  - Ideia aqui é que a população consiste em todos os possíveis cenários para taxa de inflação, sobre realizações possíveis da incerteza econômica.
    - É costumeiro denotar a "população" destes casos como "superpopulação".
    - Podemos ver (X, Y) como resultantes de um sorteio de ω realizado pela "natureza".

From this point of view we may consider the total number of possible observations (the total number of decisions to consume A by all individuals) as result of a sampling procedure, which Nature is carrying out, and which we merely watch as passive observers.

(Haavelmo, 1944)

O PROBLEMA PREDITIVO

- Considere o seguinte problema.
  - A natureza ou processo de amostragem sorteia $\omega \in \Omega$, de acordo com a lei $\mathbb{P}$.
  - Você, pesquisador, observa $X(\omega)$, e deve usar essa quantidade para realizar uma previsão de $X(\omega)$.
- No curso de verão vimos que, se $\mathbb{E}[|Y|^2] < \infty$, a regra de comunicação $h : \mathbb{R}^k \mapsto \mathbb{R}$ que minimiza o erro quadrático esperado:

$$\mathsf{EQM}(h) := \mathbb{E}[(Y - h(X))^2]\,,$$

  dentre todas as possíveis funções h t.q. $\mathbb{E}[h(X)^2] < \infty$, é dada pela esperança condicional, i.e. $h^*(X) = \mathbb{E}[Y|X]$.
- Embora utilizar h^* seja o melhor que possamos em termos de erro quadrático médio, na prática, como a lei $\mathbb{P}$ dificilmente é **conhecida**, teremos de realizar inferência estatística sobre h^* com base em uma amostra $\{(X_i, Y_i)\}_{i=1}^n$ em que $(X_i, Y_i) \overset{d}{=} (X, Y)$.
  - Inferência sobre h^* pode ser bastante ruim se adotamos um modelo não paramétrico, permitindo que h^* varie de forma complexa.

Funções preditoras lineares

- Como alternativa ao problema anterior, podemos considerar classes de funções h **mais simples**.
- Por exemplo, se $\mathbb{E}[X_j^2] < \infty$ para $j = 1, \ldots, k$, faz sentido considerar o melhor preditor na classe de funções lineares da forma $\mathcal{H}_{\text{Lin}} = \{h(x) = \alpha + x'\delta : \alpha \in \mathbb{R}, \delta \in \mathbb{R}^k\}$.
- Nesse caso, problema de previsão se torna encontrar um h_* da forma $h_*(x) = \alpha'_* + x'\delta_*$, conhecido como **melhor preditor linear**, tal que:

$$\text{EQM}(h_*) = \min_{h \in \mathcal{H}_{\text{Lin}}} \text{EQM}(h) = \min_{a \in \mathbb{R}, b \in \mathbb{R}^k} \mathbb{E}[(Y - a - b'X)^2]$$

Esperança e variância de vetores aleatórios

- O valor esperado de um vetor (matriz) aleatória U é denotado por $\mathbb{E}[U]$, e dado pelo vetor (matriz) em que a posição i (i, j) é preenchida com $\mathbb{E}[U_i]$ $(\mathbb{E}[U_{i,j}])$.
- Para um vetor aleatório $\boldsymbol{Z}$ de dimensão d tal que $\text{tr}(\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}']) < \infty$, a matriz de variância-covariância, $\mathbb{V}[\boldsymbol{Z}]$, é definida como:

$$\mathbb{V}[\boldsymbol{Z}] = \mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}'] - \mathbb{E}[\boldsymbol{Z}]\mathbb{E}[\boldsymbol{Z}']$$

- Matriz de variância-covariância é $d \times d$, simétrica, positiva semidefinida (por quê?), com entrada (i, j) iguais a:

$$\mathbb{V}[\boldsymbol{Z}]_{i,j} = \text{cov}(\boldsymbol{Z}_i, \boldsymbol{Z}_j)\,.$$

Covariância entre vetores aleatórios

- Seja $\boldsymbol{Z}$ um vetor aleatório em $\mathbb{R}^d$ com $\text{tr}(\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}']) < \infty$, e $\boldsymbol{U}$ um vetor aleatório em $\mathbb{R}^p$ com $\text{tr}(\mathbb{E}[\boldsymbol{U}\boldsymbol{U}']) < \infty$, definimos a matriz de covariância entre $\mathbf{Z}$ e $\boldsymbol{U}$ como:

$$\text{cov}(\boldsymbol{Z}, \boldsymbol{U}) = \mathbb{E}[\boldsymbol{Z}\boldsymbol{U}'] - \mathbb{E}[\boldsymbol{Z}]\mathbb{E}[\boldsymbol{U}'] \,.$$

- Matriz $d \times p$ em que a entrada (i, j) é igual a:

$$\text{cov}(\boldsymbol{Z}, \boldsymbol{U})_{i,j} = \text{cov}(\boldsymbol{Z}_i, \boldsymbol{U}_j)$$

- Note que $\text{cov}(\boldsymbol{Z}, \boldsymbol{U}) = \text{cov}(\boldsymbol{U}, \boldsymbol{Z})'$.

Caracterização do melhor preditor linear

Tome o problema de predição linear com $\mathbb{E}[|Y|^2] < \infty$ e $\mathrm{tr}(\mathbb{E}[XX']) < \infty$.

Proposição

Sejam $\alpha_* \in \mathbb{R}$ e $\delta_* \in \mathbb{R}^k$ constantes tais que $h_*(x) = \alpha_* + \delta_*' x$ é solução ao problema de predição linear. Então, definindo $\epsilon := Y - \alpha_* - \delta_*' X$ como o erro de predição, podemos escrever:

$$Y = \alpha_* + \delta_*' X + \epsilon\,,,$$

com $\mathbb{E}[\epsilon] = 0$ e $\mathrm{cov}(X, \epsilon) = 0$.

Proposição

Se $\mathbb{V}[X]$ tem posto cheio, então o melhor preditor linear de Y como função de X é da forma $h_*(x) = \alpha_* + \delta_*' x$, com

$$\delta_* = \mathbb{V}[X]^{-1}\,\mathrm{cov}(X, Y)$$

$$\alpha_* = \mathbb{E}[Y] - \delta_*'\mathbb{E}[X]$$

Discussão

- Primeira proposição mostra que o problema de predição linear define um **modelo preditivo linear** da forma:

$$Y = \alpha_* + \delta_*' X + \epsilon\,,$$

  onde α_* e δ_* são os coeficientes de melhor predição linear, e ϵ é um erro de predição que, **por construção**, possui média zero e não covaria com ϵ.
- No entanto, o primeiro resultado não garante que haja somente um único par (α_*, δ_*) que produza tal modelo.
- Segundo resultado mostra que, se $\mathbb{V}[X]$ tem posto cheio, o coeficiente do melhor preditor linear **é único**, com uma expressão fechada.
  - Essa condição limita o grau de colinearidade entre os preditores X_j, sobre realizações repetidas da incerteza.
  - De fato, condição de posto cheio equivale a $\mathbb{V}[(\psi' X)] = \psi' \mathbb{V}[X] \psi > 0$, $\forall \psi \in \mathbb{R}^k \setminus \{0\}$, de modo que nenhum preditor pode ser recuperado perfeitamente como função afim dos demais.
  - Pergunta: faria sentido incluir um preditor **perfeitamente** colinear para realizar uma predição?

Melhor aproximação linear a $\mathbb{E}[Y|X]$

- Até agora, motivamos nosso interesse sobre o melhor preditor linear como solução a um problema *restrito* de predição, cuja solução irrestrita $\mathbb{E}[Y|X]$ seria difícil de se inferir empiricamente.
- No entanto, em diversas situações, nosso interesse pode residir sobre $h^*(X) = \mathbb{E}[Y|X]$.
  - No exemplo em que $\mathbb{P}_{Y,X}$ reflete a distribuição conjunta de salário e anos de estudo na população de interesse, podemos querer entender como grupos de diferente nível educacional diferem em sua remuneração média.
    - Por exemplo, poderíamos querer calcular $h^*(x') - h^*(x)$ a diferença salarial média entre indivíduos com x' e x anos de estudo, uma *quantidade descritiva*
- O resultado abaixo nos mostra que qualquer melhor preditor linear nos oferece a melhor aproximação linear a h^*, na distância $L_2(\mathbb{P}_X)$.

Proposição

Seja $h_*(x) = \alpha_* + \delta_*' x$ solução ao problema de predição linear. Então:

$$\mathbb{E}[(h^*(X) - h_*(X))^2] = \min_{h \in \mathcal{H}_{Lin}} \mathbb{E}[(h^*(X) - h(X))^2]$$

Perguntas de natureza causal

- Até agora, nós nos restringimos a perguntas de natureza preditiva ou descritiva.
- No entanto, as perguntas mais relevantes em Economia (e nas Ciências de modo geral) são de natureza causal.
  - Qual é o efeito de um aperto monetário sobre a taxa de inflação?
  - Qual é o efeito de políticas de transferência de renda sobre a decisão de emprego dos beneficiários?
- Os efeitos causais, em Economia, estão associados ao qualificador *ceteris paribus*.
  - Efeito causal de uma variável A sobre uma variável B é definido como o efeito de se manipular o valor de A sobre B, mantidas *todas as demais causas* de B *constantes*.
  - Nesse sentido, uma curva de demanda é um objeto de natureza causal.
- Primeira etapa de uma análise causal é definir o sistema econômico ou modelo causal, explicitando quais variáveis causam o quê e os efeitos causais de interesse.
  - Trata-se de atividade puramente mental, não dependente de amostra e envolvendo conhecimento prévio ou teoria econômica (Heckman e Pinto, 2022).

Causalidade enquanto modo de pensar

The modern understanding of causal inference dates back to the 1930 lectures of Ragnar Frisch, who described causality as a thought experiment in which the researcher hypothetically manipulates one or more inputs affecting an output (Frisch 1930). According to Frisch, causality is not a physical property of the natural world, but rather an exercise of abstract manipulation of inputs causing an output. *(*Heckman e Pinto, 2022*)*

"...we think of a cause as something imperative which exists in the ***exterior world****. In my opinion this is fundamentally* ***wrong****. If we strip the word cause of its animistic mystery, and leave only the part that science can accept, nothing is left except a certain* ***way of thinking****. [T]he scientific ...problem of causality is essentially a problem regarding our way of thinking, not a problem regarding the nature of the exterior world."*

*(*Frisch (1930) *apud* Heckman e Pinto, 2024*)*

Função estrutural causal e efeitos causais

- Consideramos, agora, num espaço de probabilidade $(\Omega, \Sigma, \mathbb{P})$, uma tripla de variáveis aleatórias (Y, X, U).
  - Y representa um resultado cujas causas desejamos estudar.
  - X é um vetor que representa as causas observáveis de Y.
  - U representa as causas não observáveis
- $\mathbb{P}_{Y,X,Z}$ representa a distribuição do desfecho Y e das causas na população (superpopulação) de interesse.
- Uma função estrutural causal de Y é um mapa $h : \mathcal{X} \times \mathcal{U} \mapsto \mathbb{R}$ tal que $h(x, u)$ é o valor que observaríamos para Y caso as causas de um fenômeno fossem fixadas em $x \in \mathcal{X}$ e $u \in \mathcal{U}$.
  - Com base nessa definição temos que a variável aleatória Y efetivamente observável é $Y = h(X, U)$.
  - Ao amostrar uma unidade ω da população, o valor efetivamente observado é $Y(\omega) = h(X(\omega), U(\omega))$.
- O efeito causal de se manipular (X, U) de (x, u) para (x', u') é $h(x', u') - h(x, u)$.
  - Para qualquer unidade (indivíduo/cenário) ω amostrado da população, efeito de uma manipulação hipotética das causas de (x, u) para (x', u') é $h(x', u') - h(x, u)$.

Resultados potenciais e efeitos causais

- Como U é não observado, é costumeiro definir resultados potenciais, $Y(x) = h(x, U)$, fixando-se o valor das causas observáveis em $x \in \mathcal{X}$ e permitindo que U varie de acordo com sua distribuição na população.
  - Note que, $Y = Y(X)$, e, nesse caso, o efeito causal de se manipular X de x a x' é dado pela **variável aleatória** $Y(x') - Y(x)$.
    - Essa variável aleatória representa **o experimento hipotético** de se sortear uma unidade ω da população e perturbar suas causas observadas de x a x', mantidas as causas não observadas em $U(\omega)$.
- Representações causais em termos de funções estruturais causais e resultados potenciais são equivalentes.
  - Fácil ver que, partindo-se de uma representação causal definida a partir de uma delas, pode-se chegar à outra.

Modelo econométrico causal linear

- A função h pode ser bastante complexa, de modo que os efeitos causais $Y(x') - Y(x)$ podem ser bastante complicados.
- Um modelo econométrico causal linear consiste em uma especificação das causas de um fenômeno em que $\mathcal{X} = \mathbb{R}^k$, $\mathcal{U} = \mathbb{R}$, e:

$$h(x, u) = \mu + \beta' x + u \,,$$

em que $\mu \in \mathbb{R}$ e $\beta \in \mathbb{R}^k$.

- No modelo causal linear, os efeitos causais de se manipular X são homogêneos e não dependem de U, i.e:

$$Y(x') - Y(x) = \beta'(x' - x) = h(x', u) - h(x, u) \,, \quad \forall u \in \mathbb{R}$$

- Resultado observável Y toma a forma:

$$Y = Y(X) = h(X, U) = \mu + \beta' X + U$$

Exemplos de modelos causais lineares

Exemplo

Se a produção das firmas em um setor segue uma função de produção $f(u,l) = ul^{\alpha}$, onde $\alpha < 1$ é uma constante, u é a produtivdade da firma e l a quantidade de trabalho, então, denotando por (Y, L, U) a tripla de variáveis aleatórias que representam a produção, demanda por trabalho e produtividade de uma firma amostrada dessa população, temos que:

$$\log(Y) = \alpha \log(L) + \log(U)\,.$$

Exemplo

Considere uma regra de decisão de um banco central, por exemplo:

$$i_t = i^* + \gamma(\pi_{t-1} - \pi^m) + u_t\,,$$

onde i_t representa a taxa de juros nominal em t, i^* é a taxa de juros nominal neutra, π_{t-1} é a taxa de inflação em $t-1$, π_m é a meta de inflação, u_t são os demais fatores que influenciam a política monetária, e γ é o parâmetro de resposta do banco central a desvios da inflação da meta.

Identificação do modelo causal linear

Proposição

Considere um modelo causal linear da forma:

$$Y = Y(X) = \mu + \beta' X + U \,,$$

onde U são causas não observáveis, e $\beta \in \mathbb{R}^k$ é um vetor que representa os efeitos causais das causas observáveis. Se $\mathbb{V}[X]$ tem posto cheio e $\text{cov}(X, U) = 0$, *então é possível recuperar β como:*

$$\beta = \mathbb{V}[X]^{-1} \, \text{cov}(X, Y)$$

- Proposição acima nos mostra que se (1) não há colinearidade perfeita entre as causas observadas e (2) não há associação sistemática entre causas observáveis e não observáveis na populalação de interesse, os efeitos causais são identificáveis a partir da distribuição dos observáveis (X, Y).

Modelo linear preditivo vs. modelo linear causal

- Na aula de hoje, vimos dois tipos de modelos lineares.
- No modelo linear preditivo, temos uma relação:

$$Y = \alpha + \delta X + \epsilon\,,$$

onde (α, δ) são **definidos** como o melhor preditor linear de Y como função de X e, por esse motivo, vale **por construção** que $\text{cov}(X, \epsilon) = 0$.
- Outro tipo de modelo é o linear causal. Nele, tem-se relação da forma:

$$Y = \mu + \beta' X + U\,,$$

onde U representam as causas não observadas de um fenômeno, e β é **definido** como o vetor de efeitos causais das causas observáveis X.
  - Nesse modelo, hipótese de que $\text{cov}(X, U) = 0$ **não** vale por construção, e sua plausibilidade deve ser argumentada pelo pesquisador, visto que, sem informações adicionais, não é possível avaliá-la empiricamente.
- Primeira etapa de qualquer análise econométrica consiste em postular a pergunta de interesse, e a partir dela o modelo relevante.

Efeitos causais heterogêneos

- Em diversas situações, a restrição de homogeneidade dos efeitos causais embutida no modelo linear econométrico pode ser bastante restritiva. Nesses casos, é interessante trabalhar com efeitos heterogêneos.
- Como exemplo, considere uma causa X binária, e os resultados potenciais $(Y(1), Y(0))$ correspondentes ao experimento hipotético de fixar externamente a causa X em 0 ou 1, para um indivíduo ω amostrado da população de interesse.
- O resultado observado é da forma

$$Y = Y(X) = XY(1) + (1 - X)Y(0) = Y(0) + (Y(1) - Y(0))X$$

- Se o modelo causal é linear, então sabemos que $Y(1) - Y(0)$ é constante na população de interesse. Do contrário, $Y(1) - Y(0)$ é uma variável aleatória não constante.

Modelo linear derivado do modelo causal heterogêneo

- Partindo do modelo causal heterogêneo, e definindo $\kappa = \mathbb{E}[Y(0)]$ e $\phi = \mathbb{E}[Y(1) - Y(0)]$, podemos escrever:

$$Y = \kappa + \phi X + V$$

onde $V = Y(0) - \mathbb{E}[Y(0)] + X(Y(1) - Y(0) - \mathbb{E}[Y(1) - Y(0)])$.

- Sob as hipóteses de que $\mathbb{V}[X] > 0$ e $\text{cov}(X, Y(1)) = \text{cov}(X, Y(0)) = 0$, o melhor preditor linear identificaráca ϕ, visto que, nesse caso:

$$\phi = \frac{\text{cov}(X, Y)}{\mathbb{V}[X]}$$

  - Melhor preditor linear identifica o efeito causal esperado ou médio de uma manipulação hipotética de X de 0 a 1.

Referências

Haavelmo, Trygve (1944). "The Probability Approach in Econometrics". Em: *Econometrica* 12, pp. iii–115. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1906935 (acesso em 06/08/2024).

Heckman, James e Rodrigo Pinto (2024). "Econometric causality: The central role of thought experiments". Em: *Journal of Econometrics* 243.1, p. 105719. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2024.105719. URL: https://www.sciencedirect.com/science/article/pii/S0304407624000654.

Heckman, James J. e Rodrigo Pinto (2022). "The Econometric Model for Causal Policy Analysis". Em: *Annual Review of Economics* 14.Volume 14, 2022, pp. 893–923. ISSN: 1941-1391. DOI: https://doi.org/10.1146/annurev-economics-051520-015456. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-economics-051520-015456.