AULA 2 - Regressão: Fundamentos Conceituais

Econometria I

Área quantitativa - IE/UFRJ

Relação contábil, modelo teórico e econométrico

- Relação contábil: refletem identidades entre magnitudes econômicas.

  Exemplo: PIB= C + I + G + X - M ou
  PIB = PIBagr + PIBind + PIBser

- Modelos teóricos: expressam as leis econômicas sem a necessidade de especificá-los de forma matemática.

  Exemplo: PIB=f(i) onde f é a função e i é a taxa de juros

- Modelo econométrico: tem especificação da forma matemática, além de conter um termo residual com a finalidade de considerar variáveis e outros elementos que não puderem ser considerados explicitamente.

  Exemplo: $PIB = \alpha + \beta i + u$ onde u é o termo aleatório.

Fundamentação: definindo o problema

- Em econometria, estamos em geral interessados em identificar (empiricamente) a relação funcional entre variáveis econômicas, políticas, sociais...
- Considere um caso simples: queremos identificar a relação funcional entre duas variáveis, Y (variável dependente) e X (independente).
- Como já discutimos, por exemplo, queremos saber...
  - Como a inflação se relaciona com a taxa de juros?
  - Como o salário de um trabalhador se relaciona com sua escolaridade?
  - Como crescimento econômico se relaciona com o sistema político?
  - Como a taxa de desmatamento responde a políticas de fiscalização?

Fundamentação: definindo o problema

- O primeiro passo é supor (e aqui começa a abstração) que existe um modelo populacional por trás da relação entre Y e X.
- Modelo populacional? Um mecanismo gerador de dados, não observado, e que determina a relação entre Y e X.
- Mais do que isso: vamos supor que a relação populacional entre Y e X é linear, e definida da seguinte forma:

$$Y = \alpha + \beta X + u$$

  onde:
  - Os termos α e β são parâmetros populacionais: são 2 números, 2 constantes! Para um dado X: $\alpha + \beta X$ é um termo determinístico.
  - E o termo u é um componente estocástico, um termo de erro ou perturbação: são outros fatores que influenciam Y que não X.

Fundamentação: definindo o problema

- Fundamental ter muito claro até aqui:
  - Sabemos (ou melhor, assumimos por hipótese) que o modelo que define a relação entre Y e X existe, e que essa relação é linear.
  - ... mas não observamos, e nunca observaremos, o modelo populacional. Dito de outra forma: nunca observaremos u, tampouco α e β.
- Qual é nosso problema então:
  - Identificar o modelo populacional (descobrir quem são α e β populacionais), essa construção abstrata e não observável, e que é formada em parte por um termo de erro não observado u
  - ... a partir do comportamento de Y e X que observamos no mundo real... ou seja, a partir de uma amostra da população.

Fundamentação: definindo o problema

- Vejamos um exemplo para fixar ideias: temos 4 indivíduos em que criaremos os dados
- Cada indivíduo tem um determinado nível educacional (X anos de escolaridade) e ganha um salário mensal (Y reais).
- Como fomos nós que criamos o mundo, sabemos que o criamos de forma tal que o salário e a renda seguem exatamente esta relação na população:

$$Y = \alpha + \beta X + u$$

- Não apenas isso, como fomos nós que criamos o mundo, sabemos também qual é o valor de α e β (digamos, 1.000 e 160).

Fundamentação: definindo o problema

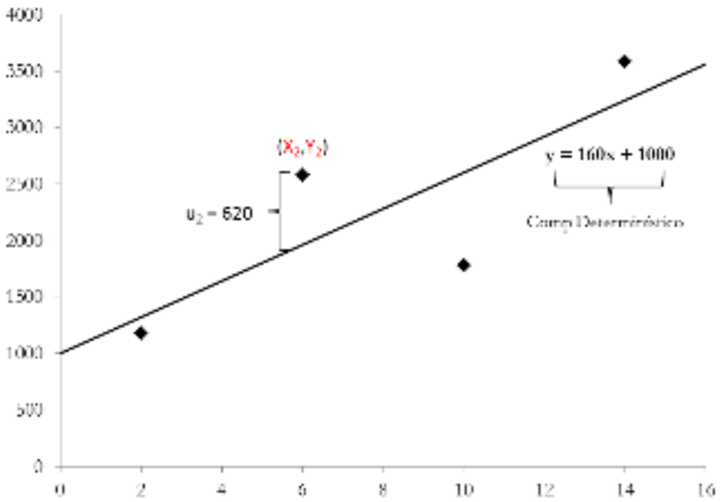
- Por fim, ao criar o mundo, escolhemos aleatoriamente, para cada indivíduo, um valor para u (que sintetiza a influência de outras variáveis, que não X, sobre Y). Em particular, deverão valer:

$$E(u) = 0 \quad \text{e} \quad E(u|X) = 0$$

- Assim, para um dado valor de X, a realização do sorteio de u determina Y.
- Logo, para uma distribuição inicial de X, uma determinada realização dos u_s gera 4 pontos em um gráfico no R^2 : um par (X, Y) para cada indivíduo.
- Por exemplo, seja $\alpha = 1000, \beta = 160$ e a realização de um sorteio descrita pelo vetor $u = \{-140, 620, -820, 340\}$. Logo temos:

Fundamentação: definindo o problema

- Graficamente, o mundo criado por nós é este...



Fundamentação: definindo o problema

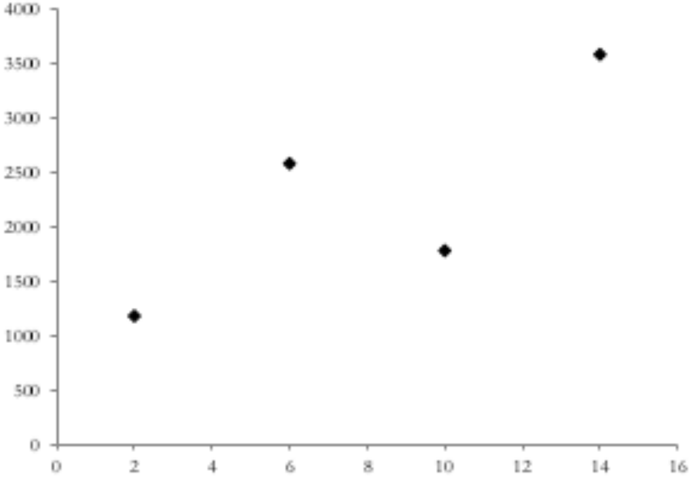
- Antes de redefinir o nosso problema, neste contexto específico, um último exercício de abstração:
  - Imaginem que somos capazes de sortear o vetor u uma infinidade de vezes.
  - Mais do que isso: suponha que u está variando continuamente na população.
  - Vamos pensar então em cada realização de u como sendo um retrato dessa população: tiramos o retrato, e a partir dele somos capazes de plotar 4 pontos no gráfico anterior.
  - Logo, para cada retrato, teremos uma nuvem distinta de pontos no plano.

- Vamos chamar esse retrato de uma amostra da população.

Fundamentação: definindo o problema

- Qual é então o problema do econometrista?
  - Nosso objetivo consiste em identificar o modelo populacional não observado (ou seja, descobrir quem são os parâmetros α e β não observados) a partir de uma amostra da população...
  - ... ou seja, a partir de uma única nuvem de pontos, ou um único retrato da população.
  - Logo, para cada ponto no gráfico, para um dado X, não somos capazes de decompor o Y em parte determinística e parte estocástica.
- Graficamente, a única coisa que observamos é...

Fundamentação: definindo o problema



Fundamentação: definindo o problema

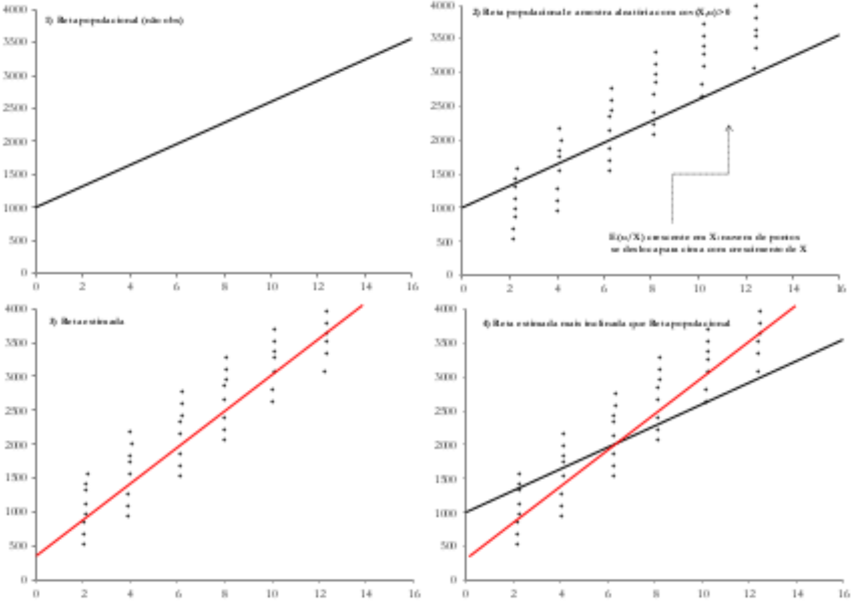
- Qual é então o problema do econometrista?
  - Na prática: traçar através de uma nuvem de pontos uma reta que seja uma boa aproximação da reta populacional...
  - ... onde traçar uma reta no plano significa encontrar um coeficiente de intercepto e um de inclinação.
  - ... e onde a cada amostra, pela natureza estocástica de u, teremos uma nuvem de pontos distinta.
  - ... e provavelmente traçaremos uma reta distinta.

- Quem são α e β populacionais? Nunca saberemos. O melhor que fazemos é buscar por uma boa aproximação destes parâmetros ao traçar uma reta através de uma nuvem de pontos.

Fundamentação: importância de $E(u|X) = 0$

- Antes de seguir adiante, vamos entender melhor a importância da validade da hipótese $E(u|X) = 0$
  - Suponha que não vale $E(u|X) = 0$; mais especificamente, suponha que u e X são positivamente correlacionados na população. Logo, $Cov(X, u) > 0$
  - ... por exemplo, ao aumentarmos escolaridade X, $E(u|X)$ tende a aumentar.
  - Assim, valores mais altos para u tenderão a deslocar a nuvem de pontos (nossa amostra aleatória, ou o que observamos) para cima da reta populacional para valores mais altos de X...
- Consequência: como observamos apenas a nuvem de pontos, e não os erros u, tenderemos a traçar uma reta mais inclinada que a reta populacional.

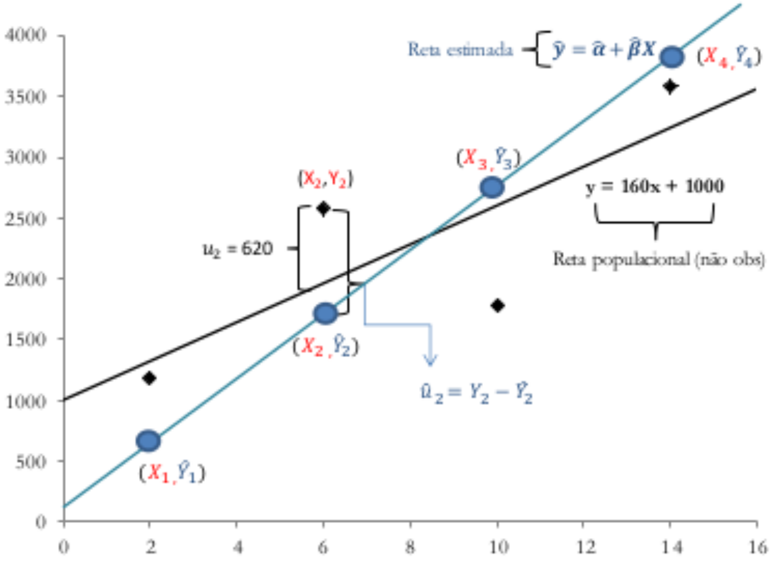
Fundamentação: importância de $E(u|X) = 0$



Estimação: Preliminares

- Já estamos quase prontos para começar o tópico de estimação: afinal, veremos que traçar uma reta é estimar um coeficiente de intercepto e um de inclinação, um exercício algébrico bastante simples.
- Antes, contudo, uma definição importante:
  - Suponha que temos uma amostra aleatória (os quatro pontos do mundo hipotético) e que traçamos uma reta qualquer por esses pontos.
  - Vamos chamar essa reta de reta estimada, e denotá-la por $\hat{Y} = \hat{\alpha} + \hat{\beta}X$.
  - Ou seja, estimamos os coeficientes $\hat{\alpha}$ e $\hat{\beta}$. Logo, para um dado X, encontramos $\hat{Y}$, chamado de Y ajustado.
- Vamos definir então por resíduo, denotado $\hat{u}$, a diferença $\hat{u} = Y - \hat{Y}$. Graficamente...

Estimação: Preliminares



Estimação: Preliminares

- Note, portanto, que ao traçar (estimar) uma reta, encontramos os resíduos: um para cada observação da amostra. No exemplo anterior encontramos um vetor $\hat{u}$, de dimensão 4×1.
- Fundamental saber que: naturalmente, para cada reta estimada, encontramos um vetor distinto de resíduos.
- Mais do que fundamental é ter muito clara a distinção entre u e $\hat{u}$ ...
  - O termo u é uma definição teórica que sintetiza a influência de outros fatores sobre Y que não X... um componente não observável de erro estocástico do modelo populacional.
  - O termo $\hat{u}$ é uma definição operacional (algébrica), portanto observável; ele é calculado a partir do momento em que estimamos uma reta qualquer com base em uma amostra da população.

Estimação: Preliminares

- Vamos então aprender a estimar o modelo populacional. Vamos antes reunir tudo o que supomos até aqui, e mais um pouco...
- Suponha que valem as seguintes hipóteses:
  - Modelo populacional existe, com linearidade nos parâmetros: $Y = \alpha + \beta X + u$
  - $E(u) = 0$ e $E(u|X) = 0$
  - Temos uma amostra aleatória da população para X e Y: $\{(X_i, Y_i), i = 1, 2, ..., n\}$, tamanho n.
- Temos que aprender então como usar esses dados (a amostra) para estimar os parâmetros populacionais α e β não observados.

Estimação: Preliminares

- Mas que dados são esses? No nosso exemplo com 4 observações, seriam os 4 pares (X, Y). Isso é tudo o que observamos.
- Podemos então escrever um modelo de regressão baseado na amostra (X_i, Y_i):

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

- Onde o par (X_i, Y_i) é observado para todo i, subscrito para o indivíduo $i = 1, 2, ..., n$; e onde, naturalmente, u_i não é observado.
- Vamos agora usar esses dados para estimar uma reta.

Estimação: Preliminares

- Existem distintos métodos de estimação em econometria, por exemplo:
  - Método de Momentos.
  - Método dos Mínimos Quadrados Ordinários, MQO.
  - Método da Maximaverossimilhança (modelos não lineares, necessária hipótese sobre distribuição).
  - Método de Momentos Generalizado (método geral, métodos acima são casos particulares).