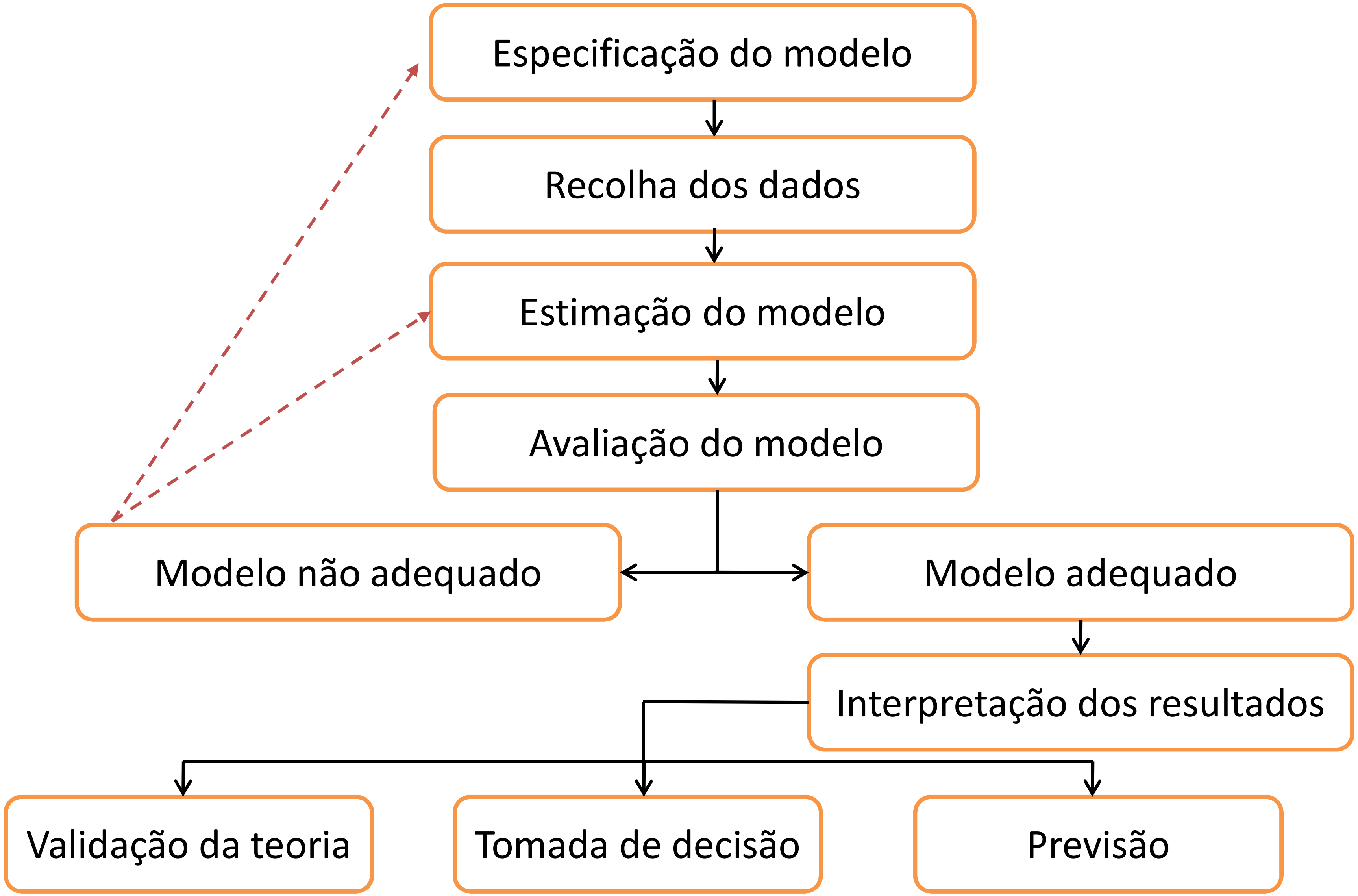
Econometria

- Noção:
  - Aplicação de técnicas estatísticas à análise de dados económicos, financeiros, sociais, etc. com o objectivo de estimar relações entre uma determinada variável dependente e um conjunto de variáveis explicativas
  - Exemplos:
    - Consumo= f(rendimento disponível)
    - Salário=f(escolaridade, experiência, idade, género)

- Finalidade:
  - Testar a validade de teorias
  - Efectuar previsões
  - Fundamentar quantitativamente políticas

Metodologia



Tipos de Dados

Dados seccionais:

- N indivíduos
- 1 observação por indivíduo

Dados temporais:

- 1 indivíduo
- T observações por indivíduo

Dados de painel

- N indivíduos
- T observações por indivíduo

Causalidade e análise ceteris paribus

Em econometria estudam-se frequentemente os **determinantes** de uma variável de interesse:

- verifica-se se as variáveis explicativas consideradas são **estatisticamente significativas** para explicar a variável de interesse, isto é, verifica-se se existe **relação de causalidade**

- Obtém-se o **efeito parcial / marginal** de cada variável explicativa sobre a variável de interesse, ceteris paribus, isto é, assumindo que tudo o resto se mantém constante. Pode interessar:
  - sinal do efeito (se positivo / negativo existe uma relação directa / inversa com a variável dependente)
  - magnitude

Objectivo / Motivação do Modelo de Regressão Linear Múltipla (MRLM)

Objectivo genérico: explicar $E(Y|X)$

Y: variável dependente / variável de interesse

X: variáveis explicativas / regressores

$E(Y|X)$: valor esperado para Y dado o valor de X

$E(Y|X)$ é função de um conjunto de parâmetros β que serão estimados

Especificação do MRLM

Especificação do Modelo:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i \qquad (i = 1, \cdots, N)$$

$$E(Y|X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

u: termo erro – contém todos os factores explicativos de Y, para além dos incluídos em X

β: parâmetros – serão estimados

k: nº variáveis explicativas

$k+1$: nº parâmetros (em modelos incluindo contante)

N: nº observações

Especificação do MRLM

Exemplo: factores incluídos em u_i

$$salário\_hora_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$$

u_i poderá incluir factores como género, região onde trabalha, motivação, …

Valores ajustados, resíduos e estimação OLS

Valores ajustados: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}$

Resíduos: $\hat{u}_i = Y_i - \hat{Y}_i$

$\hat{u}$: resíduo

$\hat{Y}$: estimador de $E(Y|X)$

$\hat{\beta}$: estimador de β

Estimação do Modelo:

- Método dos Mínimos Quadrados (OLS): $min \sum_{i=1}^{N} \hat{u}_i^2$

Interpretação dos coeficientes

Em geral, a interpretação de um coeficiente β_j consiste em avaliar a reacção do valor esperado condicional de Y dado X, $E(Y|X\ )$, a uma variação da variável explicativa X_j associada a β_j, mantendo tudo o resto constante

Modelo linear sem transformação das variáveis,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i$$

Tem-se que

$$\Delta X_j = 1 \rightarrow \Delta E(Y|X\ ) = \beta_j, \textit{ ceteris paribus}$$

Interpretação dos coeficientes

Resumo:

Modelo linear nos parâmetros: $Y^* = X^*\beta + u$

Y^*	X^*	Interpretação
Y	X_j	$\Delta X_j = 1 \rightarrow \Delta E(Y\|X) = \beta_j$
$\ln(Y)$	X_j	$\Delta X_j = 1 \rightarrow \Delta E(Y\|X) = 100\beta_j\%$
Y	$\ln(X_j)$	$\Delta X_j = 1\% \rightarrow \Delta E(Y\|X) = \frac{\beta_j}{100}$
$\ln(Y)$	$\ln(X_j)$	$\Delta X_j = 1\% \rightarrow \Delta E(Y\|X) = \beta_j\%$
Y	X, X^2	$\Delta X_j = 1 \rightarrow \Delta E(Y\|X) = \beta_x + 2\beta_{x^2}X$

Interpretação dos coeficientes

Exemplo: considere o modelo

$$preço_i = \beta_0 + \beta_1 area_i + \beta_2 quartos_i + u_i$$

onde o preço de uma casa, em milhares de dólares, depende da área (m2) e do número de quartos. Estimou-se

$$\widehat{preço}_i = -19.286 + 1.384 area_i + 15.121 quartos_i$$

Admitindo tudo o resto constante:

- por cada variação unitária de área, isto é, por cada m^2 adicional, é de esperar que o preço médio das casas aumente 1.384 milhares de dólares
- por cada variação unitária de quartos, isto é, por cada quarto adicional, é de esperar que o preço médio das casas aumente 15.121 milhares de dólares

Interpretação dos coeficientes

Output em excel

SUMÁRIO DOS RESULTADOS						
Estatística de regressão						
R múltiplo	0,794906945					
Quadrado de R	0,63187705					
Quadrado de R ajustado	0,623215334					
Erro-padrão	63,04837628					
Observações	88					
ANOVA						
	gl	SQ	MQ	F	F de significância	
Regressão	2	579971,1994	289985,6	72,95056	3,58672E-19	
Residual	85	337883,3088	3975,098			
Total	87	917854,5083				
	Coeficientes	Erro-padrão	Stat t	valor P	95% inferior	95% superior
Interceptar	-19,28550028	31,0475285	-0,62116	0,536156	-81,0163048	42,44530424
quartos	15,12133684	9,488597692	1,593632	0,11473	-3,744537442	33,98721112
area	1,38360615	0,148943494	9,28947	1,4E-14	1,08746658	1,67974572

Interpretação dos coeficientes

Output em stata

```
. regress preco quartos area

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   72.95
       Model |  579971.198     2  289985.599           Prob > F      =  0.0000
    Residual |  337883.308    85  3975.09774           R-squared     =  0.6319
-------------+------------------------------           Adj R-squared =  0.6232
       Total |  917854.506    87  10550.0518           Root MSE      =  63.048

------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     quartos |   15.12134   9.488598     1.59   0.115    -3.744538    33.98721
        area |   1.383606   .1489435     9.29   0.000     1.087467    1.679746
       _cons |   -19.2855   31.04753    -0.62   0.536    -81.0163     42.4453
------------------------------------------------------------------------------
```

Interpretação dos coeficientes

Exemplo (continuação): retome o modelo

$$\widehat{preço}_i = -19.286 + 1.384area_i + 15.121quartos_i$$

Note-se que:

- o aumento estimado no preço médio pelo acrescento de um quarto (separação de um quarto em dois), mantendo tudo o resto contante (não alterando a área), é 15.121 milhares de dólares
- o aumento estimado no preço médio pelo acrescento de um quarto que aumenta a área da casa em 13m2 é

$$\Delta\widehat{preço}_i = 1.384\Delta area_i + 15.121\Delta quartos_i$$

$$= 1.384 * 13 + 15.121 * 1 = 33.108 \text{ milhares de dólares}$$

Interpretação dos coeficientes

Exemplo: considere outro modelo alternativo

$$\widehat{ln(preço_i)} = 1.289 + 0.810ln(area_i) + 0.038quartos_i$$

Admitindo tudo o resto constante:

- uma variação da área de 1% gera um aumento no preço médio das casas de 0.810%
- por cada quarto adicional, o preço médio das casas aumenta 3.8%

Note-se que ln(.) apenas é definido para valores positivos das variáveis, por isso não se aplicou a *quartos já que quartos* pode assumir o valor 0

Interpretação dos coeficientes

Output em stata

```
. generate lpreco=ln(preco)
. generate larea=ln(area)
. regress lpreco quartos larea

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   54.47
       Model |  4.50364223     2  2.25182112           Prob > F      =  0.0000
    Residual |  3.51396129    85  .041340721           R-squared     =  0.5617
-------------+------------------------------           Adj R-squared =  0.5514
       Total |  8.01760352    87  .092156362           Root MSE      =  .20332

------------------------------------------------------------------------------
      lpreco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     quartos |   .0376464   .0303446     1.24   0.218    -.0226868    .0979795
       larea |   .8100637   .0987611     8.20   0.000     .6137002    1.006427
       _cons |    1.28929   .4666125     2.76   0.007     .3615395    2.217041
------------------------------------------------------------------------------
```

Decomposição da variação e coeficiente de determinação

- Quadro de análise de variância: decomposição da variação total de Y

	SS	DF	MS
Explicados	$\sum(\hat{y}_i - \bar{y})^2$	k	MSE
Residual	$\sum \hat{u}_i{}^2$	N-k-1	$MSR = \hat{\sigma}^2$
Total	$\sum(y_i - \bar{y})^2$	N-1	s_y^2

- Coeficiente de determinação:

$$R^2 = \frac{SSE}{SST}, \; 0 \leq R^2 \leq 1$$

Mede a proporção da variação de Y explicada pela regressão

Coeficiente de determinação

- Formas alternativas do coeficiente de determinação e comparação de modelos:

$R^2 = \frac{SSE}{SST}$	$\bar{R}^2 = 1 - (1 - R^2)\frac{N-1}{N-k-1}$
Mod. com constante	Mod. com constante
Mod. com a mesma variável dependente	Mod. com a mesma variável dependente
Mod. com k igual	

- Tipicamente, o software apresenta R^2 e a sua versão ajustada $\bar{R}^2$
- Aumentando o número de regressores, R^2 aumenta necessariamente – limitação na sua utilização
- Um R^2 elevado (baixo) não significa que o modelo descreva (não descreva) convenientemente os dados: será necessário testar a validade do modelo

Pressupostos do MRLM e propriedades do estimador OLS

Pressupostos:

1. Modelo linear nos parâmetros: $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i$
2. Amostra aleatória
3. $E(u|X\ ) = 0$
4. Ausência de colineariedade perfeita
5. Homoscedasticidade: $Var(u|X) = \sigma^2$
6. Normalidade do erro: $u \sim Normal(0, \sigma^2)$

Pequenas amostras	Propriedades assimptóticas
1-4: estimadores centrados	1-4: estimadores consistentes
1-5: estimadores centrados e eficientes	1-5: estimadores consistentes, eficientes e normalmente distribuídos
1-6: estimadores centrados, eficientes e normalmente distribuídos	

Variáveis explicativas omitidas e variáveis explicativas irrelevantes

Aquando da escolha dos regressores a incluir no modelo, deve-se ter em conta que:

- Regressores em excesso geram redução de eficiência

- Omissão de regressores (tendo em conta que os regressores omitidos estão incluídos no erro u) gera:
  - Inconsistência se $E(U|X) \neq 0$ $\rightarrow$ endogeneidade
  - Consistência se $E(U|X) = 0$ $\rightarrow$ exogeneidade

Multicolinearidade

Problema: $\sigma_{\beta_j}, j = 1, \dots k$ vem inflacionado, gerando-se estimadores com pouca precisão. Contudo, o estimador OLS de β continua consistente

Causa: decorre do facto de dois ou mais regressores estarem altamente correlacionados. Dado que se mostra que

$$\sigma_{\beta_j}^2 = \frac{\sigma^2}{\left(1 - R_j^2\right) \sum_{i=1}^{N} \left(X_{ij} - \bar{X}_j\right)^2}$$

onde R_j^2 é o coeficiente de determinação da regressão auxiliar de X_j sobre os restantes regressores, verifica-se que a variância de β_j aumenta quando :

- A variância condicional do erro, σ^2, aumenta
- a variação total de X_j diminui
- R_j^2, que mede o grau de associação entre X_j e os restantes

Distribuição por amostragem do estimador OLS, estatística t e intervalo de confiança para um parâmetro

Com $U \sim N(0, \sigma^2)$, $\mathrm{Y}|\mathrm{X} \sim N(X\beta, \sigma^2 I)$ e $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$ sendo

$$\hat{\beta}_j \sim N\left(\beta_j, \sigma^2_{\beta_j}\right)$$

Estandardizando $\hat{\beta}_j$ tem-se

$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{\beta_j}} \dot{\sim} N(0,1)$$

que não se usa directamente, pois σ_{β_j} é desconhecido. Utilizando $\hat{\sigma}^2_{\beta_j}$, tem-se

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\beta_j}} \sim t(N - k - 1)$$

Intervalo de confiança para um parâmetro

Tendo em conta o rácio t

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\beta_j}} \sim t(N-k-1)$$

Obtém-se o intervalo de confiança para β_j

$$\left(\hat{\beta}_j - t_{N-k-1}^{\alpha/2}\hat{\sigma}_{\beta_j}; \hat{\beta}_j + t_{N-k-1}^{\alpha/2}\hat{\sigma}_{\beta_j}\right)$$

com $(1-\alpha)100\%$ de confiança, β_j estará contido neste intervalo

Estatística t: principais variantes

Teste à significância (individual) estatística de um regressor:

Hipóteses	t_j	Nota
$H_0: \beta_j = 0$ $H_1: \beta_j \neq 0$	$\dfrac{\hat{\beta}_j}{\hat{\sigma}_{\beta_j}}$	Apresentando automaticamente nos output da regressão

Teste ao sinal de um coeficiente:

Hipóteses	t_j	Nota
$H_0: \beta_j = 0$ $H_1: \beta_j > 0$ $(H_1: \beta_j < 0)$	$\dfrac{\hat{\beta}_j}{\hat{\sigma}_{\beta_j}}$	Região de rejeição unilateral direita (esquerda)

Teste para um valor particular de um coeficiente (c):

Hipóteses	t_j	Nota
$H_0: \beta_j = c$ $H_1: \beta_j \neq c$ $(H_1: \beta_j > c$ ou $H_1: \beta_j < c)$	$\dfrac{\hat{\beta}_j - c}{\hat{\sigma}_{\beta_j}}$	Região de rejeição bilateral (unilateral à direita ou esquerda)

Testes sobre uma combinação linear de coeficientes

Pretende-se testar por exemplo

$H_0\colon \beta_1 = \beta_2$

A estatística de teste é então

$$t_{\hat{\delta}} = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 0}{\hat{\sigma}_{\hat{\beta}_1 - \hat{\beta}_2}} \sim \mathrm{t}(\mathrm{N} - \mathrm{k} - 1)$$

.

Testes sobre uma combinação linear de coeficientes

Exemplo: Testes de significância individual (retomar output Excel)

$$\widehat{preço}_i = -19.286 + \underset{(0.149)}{1.384}\ area_i + \underset{(9.489)}{15.121}\, quartos_i$$

SUMÁRIO DOS RESULTADOS						
Estatística de regressão						
R múltiplo	0.794906945					
Quadrado de R	0.63187705					
Quadrado de R ajustado	0.623215334					
Erro-padrão	63.04837628					
Observações	88					
ANOVA						
	gl	SQ	MQ	F	F de significância	
Regressão	2	579971.1994	289985.6	72.95056	3.58672E-19	
Residual	85	337883.3088	3975.098			
Total	87	917854.5083				
	Coeficientes	Erro-padrão	Stat t	valor P	95% inferior	95% superior
Interceptar	-19.28550028	31.0475285	-0.62116	0.536156	-81.0163048	42.44530424
quartos	15.12133684	9.488597692	1.593632	0.11473	-3.744537442	33.98721112
area	1.38360615	0.148943494	9.28947	1.4E-14	1.08746658	1.67974572

A 5% de significância, apenas área é individualmente significativa

Testes sobre uma combinação linear de coeficientes

Exemplo (continuação): Retoma-se o output do STATA

```
. regress preco area quartos

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   72.95
       Model |  579971.198     2  289985.599           Prob > F      =  0.0000
    Residual |  337883.308    85  3975.09774           R-squared     =  0.6319
-------------+------------------------------           Adj R-squared =  0.6232
       Total |  917854.506    87  10550.0518           Root MSE      =  63.048

------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        area |   1.383606   .1489435     9.29   0.000     1.087467    1.679746
     quartos |   15.12134   9.488598     1.59   0.115    -3.744538    33.98721
       _cons |   -19.2855   31.04753    -0.62   0.536     -81.0163    42.4453
------------------------------------------------------------------------------
```

Testes sobre uma combinação linear de coeficientes

Exemplo: no âmbito do modelo anterior, testar se o efeito de área e quartos é igual

$$\widehat{preço}_i = -19.286 + \underset{(0.149)}{1.384}\ area_i + \underset{(9.489)}{15.121}\ quartos_i$$

$H_0 : \beta_1 = \beta_2$

```
. test area==quartos
 ( 1)  area - quartos = 0
       F(  1,    85) =    2.06
            Prob > F =    0.1548
```

Não se rejeita a hipótese de igualdade dos dois efeitos

Testes sobre várias combinações lineares de coeficientes

Hipóteses em teste

1.Teste à **significância conjunta de alguns regressores**:

$$H_0: \beta_{p+1} = \cdots = \beta_k = 0 \left(Y_i = \beta_0 + \beta_1 X_{i1} + \cdots \beta_p X_{ip} + u_i\right) \rightarrow R_*^2$$

$$H_1: \exists \beta_j \neq 0, j = p+1, \ldots k \left(Y_i = \beta_0 + \beta_1 X_{i1} + \cdots \beta_p X_{ip} + \beta_{p+1} X_{ip+1} + \cdots \beta_k X_{ik} + u_i\right) \rightarrow R^2$$

2. Teste de **significância global**:

$$H_0: \beta_1 = \cdots = \beta_k = 0 (Y_i = \beta_0) \rightarrow R_*^2 = 0$$

$$H_1: \exists \beta_j \neq 0, j = 1, \ldots k \ (Y_i = \beta_0 + \beta_1 X_{i1} + \cdots \beta_k X_{ik} + u_i) \rightarrow R^2$$

Notas:

- Pode-se considerar $H_0: R^2 = 0$ (ausência de ajustamento)
- A magnitude do R^2 não é interpretável: deve-se fazer o teste F de ajustamento global

Testes sobre várias combinações lineares de coeficientes

Estatística de teste:

1. Teste à significância conjunta de alguns regressores:

$$F = \frac{(R^2 - R_*^2)/m}{(1 - R^2)/(N - K - 1)} \sim F(m, N - k - 1)$$

onde m é o # restrições (# β's em teste sob H_0)

.

2. Teste de significância global:

$$F = \frac{(R^2)/m}{(1 - R^2)/(N - K - 1)} \sim F(m = k, N - k - 1)$$

Região de rejeição: aba direita da distribuição *F*

Testes t e F

Exemplo: escolha entre os dois modelos seguintes

$$preço_i = \beta_0 + \beta_1 area_i + \beta_2 quartos_i + \beta_3 lote_i + u_i$$
$$preço_i = \beta_0 + \beta_2 quartos_i + e_i$$

$H_0: \beta_1 = \beta_3 = 0\ (mod. restrito)$
$H_1: não\ H_0$
$\rightarrow R_*^2 = 0.258$
$\rightarrow R^2 = 0.672$

$$F = \frac{(0.672 - 0.258)/2}{(1 - 0.672)/(88 - 3 - 1)} = 53.06$$

A 5% de significância, $F(2,84) \simeq 18$ .Rejeita-se H_0: escolhe-se o modelo alargado

Exemplo (continuação):

Output STATA

```
. regress preco area lote quartos

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  3,    84) =   57.43
       Model |  617018.847     3  205672.949           Prob > F      =  0.0000
    Residual |  300835.658    84  3581.37688           R-squared     =  0.6722
-------------+------------------------------           Adj R-squared =  0.6605
       Total |  917854.506    87  10550.0518           Root MSE      =  59.845
                                             .
------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        area |   1.322524   .1426449     9.27   0.000     1.038859    1.606189
        lote |   .0222365   .0069137     3.22   0.002     .0084878    .0359852
     quartos |    13.7864   9.015998     1.53   0.130    -4.142903     31.7157
       _cons |  -21.72645   29.47964    -0.74   0.463    -80.34994    36.89704
------------------------------------------------------------------------------

. test área quartos
(…)
```

Testes t e F

Exemplo: teste ao ajustamento global / significância conjunta dos regressores

$$preço_i = \beta_0 + \beta_1 area_i + \beta_2 quartos_i + \beta_3 lote_i + u_i$$

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ (sem significância conjunta)
$H_1: não\ H_0$
$\rightarrow R^2 = 0.672$

$$F = \frac{0.672/3}{(1-0.672)/(88-3-1)} = 57.43$$

A 5% de significância, $F(3{,}84) \simeq 8{,}0$

Rejeita-se H_0: evidência de ajustamento global / os regressores são conjuntamente significativos

A estatística de teste e respectivo p-value é dada no output de *regress* no STATA (ver resultados a verde)

Notas sobre testes assimptoticos

Quando a amostra é grande, dispensa-se o pressuposto (6), relativo à normalidade de u

- Teste a coeficientes individuais:

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\beta_j}} \sim N(0,1)$$

Embora também se possa utilizar $t_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\beta_j}} \sim t(N-k-1)$

- Teste a conjuntos de coeficientes:
  - Utiliza-se o teste F, apesar da distribuição ser apenas aproximada
  - Pode-se aplicar o **teste LM** (Lagrange multiplier), que se executa de forma muito fácil no caso do **teste de ajustamento global**:

$$LM = NR^2 \sim X_k^2$$

  - **Região de rejeição:** aba direita da distribuição *Qui-quadrado*

Notas sobre testes assimptoticos

Exemplo: testar a significância conjunta dos regressores pelo teste LM no âmbito do modelo

$$preço_i = \beta_0 + \beta_1 area_i + \beta_2 quartos_i + \beta_3 lote_i + u_i$$

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ (ausência de significância conjunta)
$H_1: não\ H_0$

.

$$LM = 88 * 0.6722 = 59.154$$

A 5% de significância, o valor crítico é $X_3^2 = 7.815$. Rejeita-se H_0

Introdução de determinantes qualitativos

Existem factores explicativos do valor esperado condicionado da variável dependente que são de natureza qualitativa:

- Preco de imóveis = f(área, nº quartos, qualidade da localização, tipo de imóvel)
- Salario = f(idade, experiencia profissional, tipo de profissão, género, região local trabalho)

A introdução dos factores qualitativos é feita por meio de variáveis artificiais, designadas de variáveis dummy, que têm a particularidade de apenas poderem assumir o valor 1 ou 0

$$d = \begin{cases} 1 \; se \; atributo \; observado \\ 0 \; se \; atributo \; não \; observao \end{cases}$$

Introdução de determinantes qualitativos

Representação de duas categorias: requer a inclusão de uma dummy

$$d = \begin{cases} 1 \; se \; atributo \; observado \; (ex: homem) \\ 0 \; se \; atributo \; não \; observado \; (ex: mulher) \end{cases}$$

Representação de M categorias: requer a introdução de M-1 dummies

$$d_1 = \begin{cases} 1 \; se \; atributo \; 1 \; observado \; (ex: região \; Sul) \\ 0 \; se \; atributo \; não \; observado \; (ex: outras \; regiões) \end{cases}$$

$$d_2 = \begin{cases} 1 \; se \; atributo \; 2 \; observado \; (ex: região \; Centro) \\ 0 \; se \; atributo \; não \; observado \; (ex: outras \; regiões) \end{cases}$$

Em ambos os casos a interpretação do efeito sobre o valor esperado é feita por referência ao atributo omitido: no primeiro exemplo o ser mulher e no segundo o pertencer à região Norte

Introdução de determinantes qualitativos

Exemplo:

$$Y_i = \beta_0 + \beta_1 masc_i + \beta_2 sul_i + \beta_3 centro_i + u_i$$

β_1: diferença no valor de Y de um homem relativamente a uma mulher;

β_2: diferença em Y de um individuo que vive no Sul, relativamente a um individuo que vive no Norte

β_3: diferença em Y de um individuo que vive no Centro, relativamente a um individuo que vive no Norte

Ajustamento para a interpretação quando o modelo é log-lin:

$$\left(e^{\beta_j} - 1\right)100\%$$

Importante quando β_j é de grande magnitude. Por exemplo

- $(e^{0.457-0.063} - 1)100\% = 48.3\%$
- $(e^{-0.063} - 1)100\% = -6,1\ \%$

Introdução de determinantes qualitativos

Variável de interacção: resulta da multiplicação de uma dummy por outra variável

Exemplo:

$$salario_i = \beta_0 + \beta_1 masc_i + \beta_2 educ_i + \beta_3 masc * educ_i + u_i$$

Escrevendo o modelo para cada grupo:

- homens (masc=1): $salario_i = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) educ_i + u_i$
- mulheres (masc=0): $salario_i = \beta_0 + \beta_2 educ_i + u_i$

Logo:

- β_0: termo independente relativo às mulheres
- β_1: diferença no termo independente entre os homens e as mulheres
- β_2: acréscimo que se verifica no salário das mulheres por cada ano de escolaridade adicional
- β_3: diferença que se verifica no efeito anterior entre os homens e as mulheres.

Introdução de determinantes qualitativos

Teste de Chow para a Quebra de Estrutura:

- Contexto:
  - Dois grupos de indivíduos / empresas: G_A, G_B
  - Suspeita-se que o comportamento dos dois grupos no que diz respeito à variável dependente tenha diferentes determinantes
- Implementação:
  - Gerar a variável binária $D = \begin{cases} 1 & \text{se a observação pertence a } G_A \\ 0 & \text{se a observação não pertence a } G_A \end{cases}$
  - Estimar o modelo original de forma 'duplicada':
    $Y = \theta_0 + \theta_1 X_1 + \cdots + \theta_k X_k + \gamma_0 D + \gamma_1 D X_1 + \cdots + \gamma_k D X_k + v$
  - Aplicar um teste F para a significância das variáveis envolvendo D:

    $H_0: \gamma_0 = \gamma_1 = \cdots = \gamma_k = 0$ (sem quebra de estrutura)

    H_1: Não H_0 (com quebra de estrutura)

Teste à forma funcional - RESET

Modelo em teste: $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i$

Teste F para a significância de um sub-conjunto de parâmetros, onde o modelo sob H_1 é uma regressão auxiliar que acrescenta ao original potências de $\hat{Y}$ (isto é, de $X\hat{\beta}$): $\hat{Y}^2 = (X\hat{\beta})^2$, $\hat{Y}^3 = (X\hat{\beta})^3$…

.

H_0: $\gamma_1 = \cdots = \gamma_k = 0$ $(Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i) \rightarrow$ FF correcta $\rightarrow R_*^2$

H_1: $n\ H_0$ $(Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \gamma_1 \hat{Y}_i^2 + \gamma_2 \hat{Y}_i^3 \ldots + v) \rightarrow$ FF incorrecta $\rightarrow R^2$

Nota: a rejeição de H_0 requer que se procure outra forma funcional alternativa - o modelo alternativo de H_1 é uma regressão auxiliar, não é candidato a modelo alternativo

Exemplo: testar a forma funcional de

$$preço_i = \beta_0 + \beta_1 area_i + \beta_2 quartos_i + u_i$$

```
 regress preco quartos area

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   72.95
       Model |  579971.198     2  289985.599           Prob > F      =  0.0000
    Residual |  337883.308    85  3975.09774           R-squared     =  0.6319
-------------+------------------------------           Adj R-squared =  0.6232
       Total |  917854.506    87  10550.0518           Root MSE      =  63.048
------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     quartos |   15.12134   9.488598     1.59   0.115    -3.744538    33.98721
        area |   1.383606   .1489435     9.29   0.000     1.087467    1.679746
       _cons |   -19.2855   31.04753    -0.62   0.536     -81.0163     42.4453
------------------------------------------------------------------------------

. predict precohat
(option xb assumed; fitted values)

. generate precohat2=precohat^2

. generate precohat3=precohat^3
```

Exemplo (continuação):

```
. regress preco quartos area precohat2 precohat3

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  4,    83) =   41.17
       Model |  610249.039     4  152562.26            Prob > F      =  0.0000
    Residual |  307605.467    83  3706.08996           R-squared     =  0.6649
-------------+------------------------------           Adj R-squared =  0.6487
       Total |  917854.506    87  10550.0518           Root MSE      =  60.878

------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     quartos |  -58.37904   38.71904    -1.51   0.135    -135.3897    18.63158
        area |  -5.680895   3.613211    -1.57   0.120    -12.86743    1.505637
   precohat2 |   .0133394   .0076821     1.74   0.086    -.0019399    .0286187
   precohat3 |  -.0000109   7.20e-06    -1.52   0.133    -.0000252    3.40e-06
       _cons |   675.0476   328.2222     2.06   0.043     22.22683    1327.868
------------------------------------------------------------------------------
```

$H_0: \gamma_1 = \gamma_2 = 0 \rightarrow$ FF correcta

$$F = \frac{(0.6649 - 0.631 *)/2}{(1 - 0.6649)/(88 - 5)} = 4.08$$

A 5% de significância, $F(2{,}83) =$

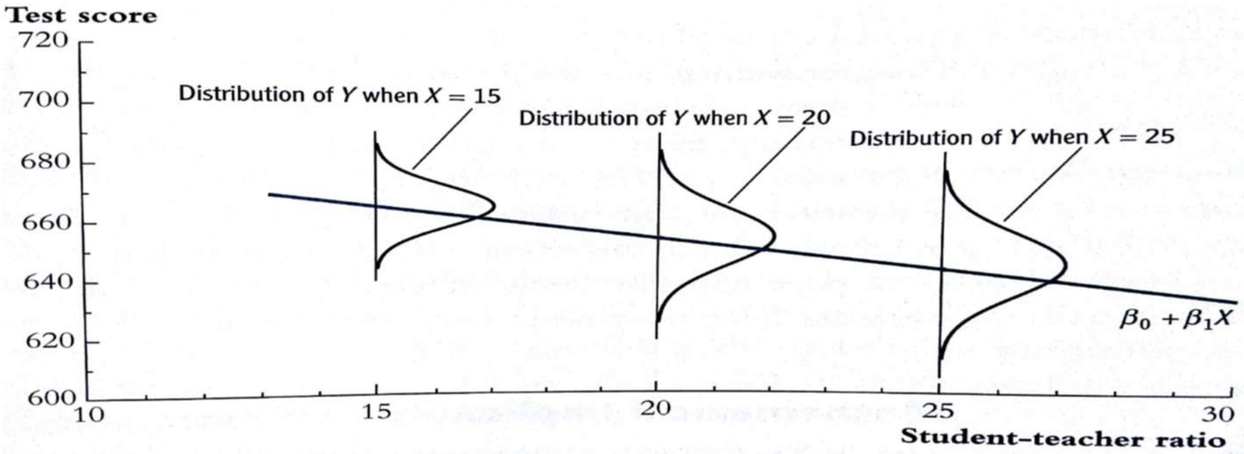
Rejeita-se H_0: a FF não é válida

Heteroscedasticidade: motivação e definição

Variância condicional do erro (skedastic function): $Var(u|X)$

Pressuposto (5): $Var(u|X) = \sigma^2$ (homocedasticidade)
Falha do pressuposto (5): $Var(u|X) = \sigma^2 h(X)$ (heteroscedasticidade)

- Ilustração:



O pressuposto (5) garante a eficiência e a normalidade assimptótica dos estimadores OLS, mas não coloca em causa o seu não enviesamento ou consistência. Assim, **os estimadores dos MQ continuam a ser centrados e consistentes mas perdem a eficiência e normalidade assimptotica**. Além disto, a fórmula da variância deixa de ser válida

Heteroscedasticidade: estimação robusta da matriz de covariâncias

Dado que os estimadores OLS continuam a ser centrados e consistentes, corrige-se apenas a sua variância

Variância standard (assume $Var(U|X) = \sigma^2\mathrm{I}$)

$$Var(\hat{\beta}) = (X'X)^{-1}X'Var(U|X)X(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

Variância robusta (assume $Var(\dot{U}|X) = \Sigma$)

$$Var(\hat{\beta}) = (X'X)^{-1}X'Var(U|X)X(X'X)^{-1}$$
$$= (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

onde $\Sigma = diag(\sigma_1^2, \sigma_2^2, \dots \sigma_n^2)$ para u_i independentes mas com variâncias σ_i^2

A implementação é simples mas o método é apenas válido assimptoticamente e os estimadores não são eficientes.

Heteroscedasticidade: estimação robusta da matriz de covariâncias

Exemplo (continuação): estimação de variâncias robustas (os coeficientes serão iguais aos do OLS)

```
. regress preco quartos area, robust
Linear regression                                      Number of obs =      88
                                                       F(  2,     85) =   27.22
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.6319
                                                       Root MSE      =  63.048
------------------------------------------------------------------------------
             |               Robust
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     quartos |   15.12134    8.96599     1.69   0.095    -2.705452    32.94813
        area |   1.383606   .2111629     6.55   0.000     .9637578    1.803455
       _cons |   -19.2855   41.54017    -0.46   0.644    -101.8785    63.30749
------------------------------------------------------------------------------
```

$$\widehat{preço}_i = -19.286 + \underset{\substack{(0.149)\\[0.211]}}{1.384}\, area_i + \underset{\substack{(9.489)\\[8.966]}}{15.121}\, quartos_i$$

Heteroscedasticidade: testes para detecção

Há vários testes, todos baseados em regressões auxiliares com variável dependente $\hat{u}^2$, onde a significância global dos parâmetros é testada através dum teste F / LM.

Modelo base:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i$$

H_0:Homocedasticidade → $Var(u|X) = \sigma^2$ (usar OLS)
H_1:Heteroscedasticidade→ $Var(u|X) = \sigma^2 h(X)$ (usar variâncias robustas

1. Procedimento comum: estimar o modelo base por OLS e obter $\hat{u}^2$

Heteroscedasticidade: testes para detecção

2. Estimar a regressão auxiliar:

- Breusch Pagan: $\hat{u}^2 = \gamma_0 + \gamma_1 X_1 + \cdots + \gamma_k X_k + e$

e obter $R^2_{\hat{u}^2}$

3. Estatística de teste e distribuição:

$$F = \frac{R^2_{\hat{u}^2}/k}{\left(1 - R^2_{\hat{u}^2}\right)/(N - k - 1)} \sim F(k, N - k - 1)$$

ou

$$LM = NR^2_{\hat{u}^2} \sim X^2_k$$

Heteroscedasticidade: testes para detecção

Exemplo: testar heteroscedasticidade no âmbito do modelo

$$ln(preço_i) = \beta_0 + \beta_1 ln(area_i) + \beta_2 quartos_i + u_i$$

```
gen larea=log(area)
gen lpreço=log(preco)
regress  lpreço larea quartos
      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   54.47
       Model |  4.50364223     2  2.25182112           Prob > F      =  0.0000
    Residual |  3.51396129    85  .041340721           R-squared     =  0.5617
-------------+------------------------------.          Adj R-squared =  0.5514
       Total |  8.01760352    87  .092156362           Root MSE      =  .20332
------------------------------------------------------------------------------
      lpreço |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       larea |   .8100637   .0987611     8.20   0.000     .6137002    1.006427
     quartos |   .0376464   .0303446     1.24   0.218    -.0226868    .0979795
       _cons |    1.28929   .4666125     2.76   0.007     .3615395    2.217041
------------------------------------------------------------------------------
predict uhat, resid
gen uhat2=uhat^2
```

Heteroscedasticidade: testes para detecção

Exemplo (continuação):

Teste BP:

```
regress uhat2 larea quartos
      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =    1.93
       Model |  .026033827     2  .013016913           Prob > F      =  0.1516
    Residual |  .573679783    85  .006749174           R-squared     =  0.0434
-------------+------------------------------           Adj R-squared =  0.0209
       Total |   .59971361    87   .00689326           Root MSE      =  .08215
------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t.    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       larea |  -.0607216   .0399045    -1.52   0.132    -.1400625    .0186193
     quartos |   .0227115   .0122608     1.85   0.067    -.0016662    .0470892
       _cons |   .2744371   .1885353     1.46   0.149    -.1004216    .6492957
------------------------------------------------------------------------------
* Teste LM
display 88*0.0434
3.8192
display chi2tail(2,3.8192)
.14813963
```

Não se rejeita H_0: existe evidencia de homocedasticidade (tanto pelo teste F como pelo LM)

Dados de Painel:

- N empresas / indivíduos: $i = 1, \ldots, N$
- T observações por empresa / indivíduo: $t = 1, \ldots, T$

var7[19]

	cidade	ano	pop	est	renda	rendime
1	1	80	75211	15303	197	11537
2	1	90	77759	18017	342	19568
3	2	80	106743	22462	323	19841
4	2	90	141865	29769	496	31885
5	3	80	36608	11847	216	11455
6	3	90	42099	10265	351	21202
7	4	80	36640	13825	267	14682
8	4	90	46209	18173	588	29044
9	5	80	62134	8175	475	31300
10	5	90	110330	18205	925	56307
11	6	80	118550	12223	276	16291

Dados seccionais / painel / agrupados:

- Dados seccionais: indivíduos diferentes ⇒ observações independentes
- Dados de painel: mesmos indivíduos ⇒ observações dependentes ao longo do tempo
- Dados agrupados: indivíduos diferentes observados em diferentes momentos ⇒ observações independentes

Algumas vantagens:

- Análise da dinâmica temporal dos comportamentos individuais
- Ganhos de eficiência, dado o aumento da amostra
- Lidar facilmente com endogeneidade resultante da omissão de variáveis que não se alteram no tempo

Tipos de dados

Painéis curtos:

- Amostra composta por muitos indivíduos ($N \longrightarrow \infty$) mas com um horizonte temporal curto (T pequeno)
- Assume-se correlação temporal das observações para cada indivíduo, mas independência entre os indivíduos.

Painéis equilibrados:

- Todos os indivíduos são observados em todos os momentos ($T_i = T, \forall i$)

Painéis desequilibrados:

- Alguns indivíduos não são observados nalguns períodos de tempo ($T_i \neq T$)
- Principal causa: alguns indivíduos desistem de dar informação ao fim de alguns períodos $\longrightarrow$ problema conhecido como 'attrition'
- Muitos estimadores funcionam com painéis desequilibrados

Modelos para dados de painel

Introdução de time-dummies e interações

Objectivo: estudar o efeito do tempo

Para T anos , toma-se o primeiro ano como referência e incluem-se $(T-1)$ dummies, uma para cada um dos restantes anos

- incluem-se essas dummies, sendo que os seus coeficientes informam sobre a alteração de Y relativamente à base, controlados os factores explicativos incluidos no modelo
  - exemplo, o coeficiente de D2017 informa quanto variou Y entre o ano base e 2017, devido a factores que não os regressores

- se os efeitos parciais dos regressores se alteram no tempo acrescentam-se também termos de interação
  - pode-se usar o teste Chow para quebra de estrutura

Modelos para dados de painel

Variabilidade no tempo e por individuos:

- Com dados de painel, pode-se decompor a variância de Y_{it}:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\bar{Y})^2=\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\bar{Y}_i+\bar{Y}_i-\bar{Y})^2$$

$$=\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\bar{Y}_i)^2+\sum_{i=1}^{N}(\bar{Y}_i-\bar{Y})^2$$

Variabilidade dentro do grupo i   Variabilidade entre os grupos

- Obtém-se a variância 'within', 'between' e total dividindo por, respectivamente, $NT-1$, $N(T-1)$ e $N-1$

Modelos para dados de painel

Modelo Base - Modelo de Efeitos Individuais:

$$Y_{it} = \alpha_i + x'_{it}\beta + u_{it} \qquad (i = 1, \dots, N; t = 1, \dots, T)$$

- α_i: efeitos individuais, não dependentes do tempo e não observáveis
- x_{it} - variáveis explicativas, incluindo:
  - x_{it}: varia de indivíduo para indivíduo e entre os vários períodos de tempo
  - x_i: invariante no tempo
  - d_t: *dummy* temporal para o momento *t*
  - $d_t.x_{it}$: termos de interacção
- u_{it}: erro idiossincrático - difere de indivíduo para indivíduo e de período para período de forma aleatória

Modelos para dados de painel

O modelo de efeitos individuais pode ser reescrito como:

$$Y_{it} = x_{it}'\beta + (\alpha_i + u_{it})$$

- O termo erro passa a ter duas componentes, assumindo-se que α_i pode ou não estar relacionado com as variáveis explicativas

Efeitos Aleatórios:

- Pressuposto: α_i e x_{it} não estão correlacionados → x_{it} é exógeno
- Estimadores em análise: "Pooled" e Efeitos Aleatórios

Efeitos Fixos:

- Pressuposto: α_i e x_{it} podem estar correlacionados → x_{it} é endógeno (relativamente à parte do erro invariante no tempo)
- Estimadores em análise: Efeitos Fixos ou "Within" e Diferenças

Estimador pooled

- Equação:

$$Y_{it} = \alpha + x_{it}'\beta + \underbrace{(\alpha_i - \alpha + u_{it})}_{v_{it}}$$

- Estimação:
  - Método dos mínimos quadrados com opção cluster ou similar para a variância

Stata

regress $Y X_1 \dots X_k$, vce(cluster *clustvar*)

Estimador de efeitos aleatórios

Equação:

$$Y_{it} = \alpha + x_{it}'\beta + (\alpha_i - \alpha + u_{it})$$

com $Var(\alpha_i) = \sigma_\alpha^2$ e $Var(u_{it}) = \sigma_u^2$

Explorando $cor(u_{it}, u_{is}) = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_u^2)$, estes estimadores são mais eficientes. Os estimadores anteriores não exploravam a natureza de painel dos dados no cálculo dos estimadores (apenas na parte da inferência)

Estimação: método dos mínimos quadrados generalizados

$$Y_{it} - \hat{\theta}_i\bar{Y}_i = \left(1 - \hat{\theta}_i\right)\alpha + \left(x_{it} - \hat{\theta}_i\bar{x}_i\right)'\beta + v_{it}$$

onde $\hat{\theta}_i = 1 - \sqrt{\hat{\sigma}_u^2/(T_i\hat{\sigma}_\alpha^2 + \hat{\sigma}_\alpha^2)}$ e $v_{it} = \left(1 - \hat{\theta}_i\right)\alpha_i + \left(u_{it} - \hat{\theta}_i\bar{u}_i\right)$

Stata

xtreg $Y X_1 \ldots X_k$, vce(cluster *clustvar*)

Modelos para dados de painel
Estimador de efeitos fixos

Equação:

$$Y_{it} - \bar{Y}_i = (x_{it} - \bar{x}_i)'\beta + (u_{it} - \bar{u}_i)$$

Estimação: método dos mínimos quadrados com opção cluster ou similar para a variância

Stata

xtreg $Y X_1 \ldots X_k$, fe vce(cluster *clustvar*)

Apesar da robustez, dado o não requisito de efeitos aleatórios, tem a desvantagem de eliminar do modelo

- As variáveis explicativas que não se alteram no tempo
- As variáveis explicativas que se alteram de forma constante no tempo (ex.: idade), se o modelo incluir dummies temporais

Estimador das diferenças

Equação:

$$y_{it} - y_{i,t-1} = \left(x_{it} - x_{i,t-1}\right)'\beta + \left(u_{it} - u_{i,t-1}\right) \Leftrightarrow$$
$$\Delta y_{it} = \Delta x'_{it}\beta + \Delta u_{it}$$

Stata
regress $D.Y$ $D.X_1$ … $D.X_k$, vce(cluster *clustvar*)

Tem as mesmas desvantagens do estimador de EF

Igual ao estimador de EF para T=2

Teste de Hausman:

$H_0: E(\alpha_i x_{it}) = 0$ (EA e EF consistentes, EA eficiente)

$H_1: E(\alpha_i x_{it}) \neq 0$ (EF consistentes, EA inconsistente)

$$H = (\hat{\beta}_{EF} - \hat{\beta}_{EA})'\left[V(\hat{\beta}_{EF}) - V(\hat{\beta}_{EA})\right]^{-1}(\hat{\beta}_{EF} - \hat{\beta}_{EA}) \sim \chi^2_k$$

<u>Stata</u>

(modelos têm de ser estimados sem opções robust ou cluster)

xtreg $Y X_1 \ldots X_k$, fe
estimates store *ModeloEF*
xtreg $Y X_1 \ldots X_k$
estimates store *ModeloEA*
hausman *ModeloEF ModeloEA*

Análise de política com dados de painel de dois períodos de tempo

Considera-se uma amostra em condições similares da "quasi experiment": os indivíduos são observados duas vezes, uma antes do programa de intervenção e outra depois, e devem ser de dois tipos, afectados pela experiencia (tratados) e não afectados (controls)

Modelo geral:

$$y_{it} = \alpha + \delta d2 + \beta prog_{it} + \alpha_i + u_{it}$$

onde $prog = 1$ se recebeu formação

Modelo às diferenças:

$$\Delta y_{it} = \delta + \beta prog_{it} + \Delta u_{it}$$

Efeito de interesse: $\beta = \overline{\Delta y}_{trat} - \overline{\Delta y}_{cont}$

Análise de política com dados de painel de dois períodos de tempo

Exemplo: Wooldridge

Pretende-se apurar se a taxa de produtos defeituosos (% de produtos defeituosos inutilizados na produção total da empresa), *scrap*, se altera em função da participação num programa de formação, (*Grant=1* se participa), ocorrido em 1988. Os dados são de painel, para 1987 e 1988 e contêm indivíduos que receberam e indivíduos que não receberam formação profissional

Modelo estimado

$$\Delta \widehat{ln(scrap)} = -\underset{(0.097)}{0.057} - \underset{(0.164)}{0.317}\ \Delta grant, n = 54, R^2 = 0.067$$

- A formação reduziu a taxa de rejeição em $(e^{0.317} - 1)100\% = 27.2\%$
- A taxa de rejeição reduziu-se em $(e^{0.057} - 1)100\% = 5.9\%$ por factores que não a formação profissional

Modelos para dados de painel dinâmicos: introdução

Modelos com a variável dependente desfasada, $Y_{i,t-1}$, $Y_{i,t-2}$, entre as variáveis explicativas

- Exemplo – $AR(1)$ model:

$$Y_{it} = \gamma_1 Y_{i,t-1} + \alpha_i + u_{it}$$

- Todos os estimadores do modelo estático apresentados anteriormente são inconsistentes

Modelos para dados de painel dinâmicos: introdução

Modelo básico para painéis dinâmicos:

$$\Delta Y_{it} = \gamma_1 \Delta Y_{i,t-1} + \Delta x'_{it}\beta + \Delta u_{it}, t = 3, \ldots, T$$

Pressuposto: u_{it} não tem autocorrelação $\Longrightarrow$ Δu_{it} autocorrelação de ordem 1:

$$Cov(\Delta u_{it}, \Delta u_{i,t-1}) = -Cov(u_{i,t-1}, u_{i,t-1}) \neq 0$$

Principais estimadores (estimadores de variáveis instrumentais):

- Anderson-Hsiao (1981)
- Arellano-Bond (1991) – 'Difference GMM'
- Blundell-Bond (1998) – 'System GMM'

Modelos para dados de painel dinâmicos: introdução

Anderson-Hsiao (1981):

2 tipos de VI:

- $Y_{i,t-2}$

Stata
ivregress gmm $D.Y(DL.Y = L2.Y)\ D.X_1 \ldots D.X_k$

- $\Delta Y_{i,t-2}$ (perde-se 1 observação mas em geral gera estimadores mais eficientes)

Stata
xtivreg $\mathrm{D}.Y(\mathrm{DL}.Y = \mathrm{DL2}.Y)\ \mathrm{D}.X_1 \ldots \mathrm{D}.X_k$

or

xtivreg $Y(\mathrm{L}.Y = \mathrm{L2}.Y)\ X_1 \ldots X_k$, fd

Modelos para dados de painel dinâmicos: introdução

Arellano-Bond (1991):

- Sugere o uso de todos os lags de $Y_{i,t}$ como VI:
  - $t = 3: Y_{i,1}$
  - $t = 4: Y_{i,2}, Y_{i,1}$
  - …
  - $t = T: Y_{i,T-2}, \dots, Y_{i,2}, Y_{i,1}$
- # VI= $(T-1)(T-2)/2$
- Pode-se usar apenas um subconjunto das VI disponiveis
- Mais eficiente do que os estimadores Anderson-Hsiao's (1981)

Stata
$\text{xtabond } YX_1 \dots X_k, \text{ maxldep(\#) twostep vce(robust)}$

Modelos para dados de painel dinâmicos: introdução

Blundell-Bond (1998):

- Sugere o uso das VI: $\Delta Y_{i,2}, \ldots, \Delta Y_{i,T-1}$
- # VI= $\frac{(T-1)(T-2)}{2} + (T-2)$
- Pode-se usar apenas um subconjunto das VI disponiveis
- Mais eficiente do que o estimador Arellano-Bond's (1991), mas requer mais pressupostos

<u>Stata</u>

xtdpdsys $Y X_1 \ldots X_k$, maxldep(#) twostep vce(robust)

Modelos para dados de painel dinâmicos: introdução

Testes principais a aplicar:

- Teste J de Hansen para sobreidentificação

Stata
(after xtabond or xtdpdsys, with variance estimated in a standard way)
estat sargan

- Teste de autocorrelação

Stata
(after xtabond or xtdpdsys, with variance estimated in a standard way)
estat abond, artests(3)