Capítulo 7

- **7. Modelo de regressão com dados temporais: Análise de Regressão Básica**
  - 7.1 A natureza dos dados de séries temporais
  - 7.2 Modelos dinâmicos
  - 7.3 Propriedades do OLS em amostras finitas
  - 7.4 Tendência e sazonalidade
  - 7.5 Modelação de acontecimentos
  - 7.6 Exemplos com aplicações empíricas

7.1. A natureza dos dados de séries temporais

- As observações dizem respeito ao valor de uma variável num dado período

Existe uma ordem/sequência das observações de acordo com a sequência temporal → as observações não podem ser reordenadas

- Características típicas:
  - observações correlacionas / não independentes
  - Tendências
  - Sazonalidade
  - Estacionaridade

7.1. A natureza dos dados de séries temporais

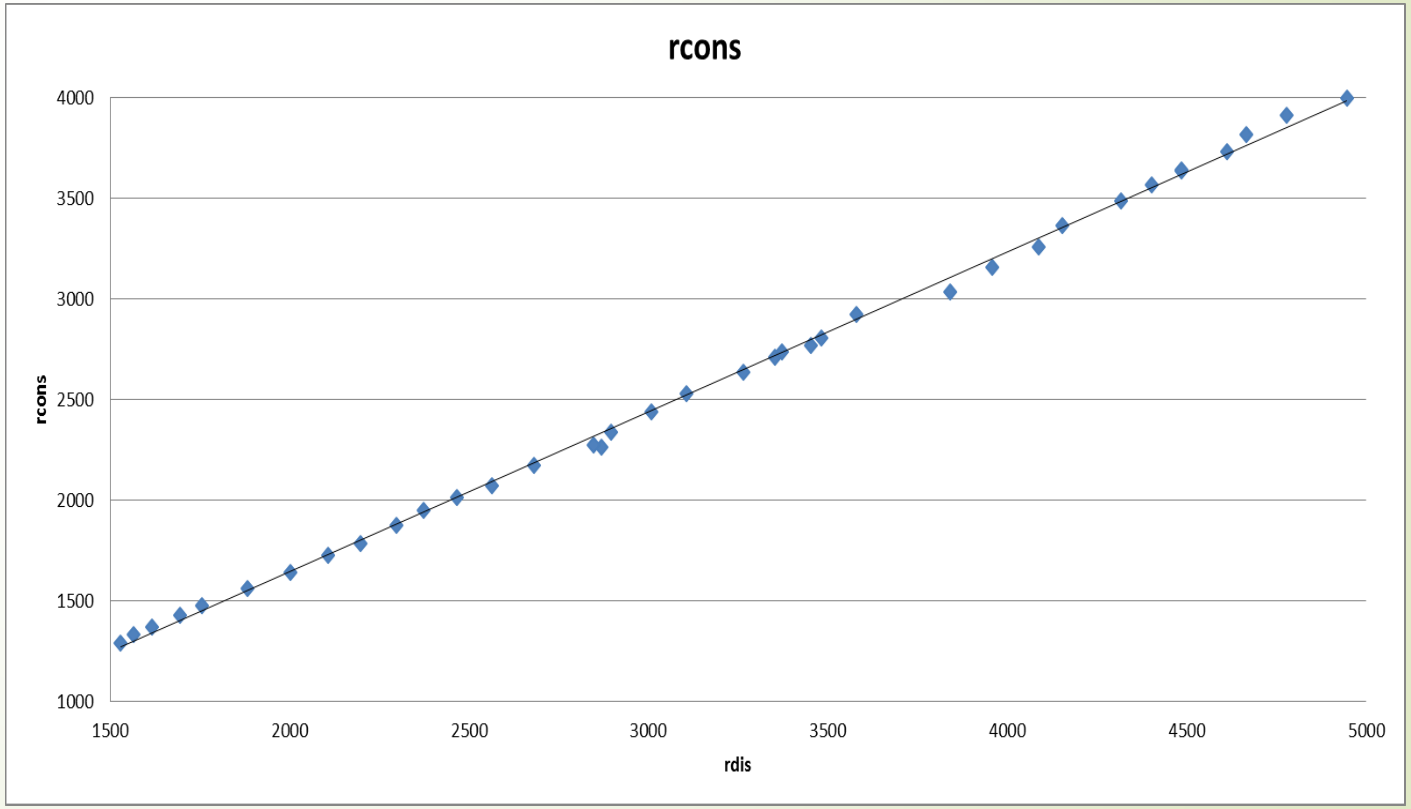
- **Conceito de aleatoriedade, amostra e população em dados temporais**
  - Os dados temporais são **processos estocásticos** (sequências de variáveis aleatórias)
  - A aleatoriedade das variáveis não é devido a um processo de amostragem de uma população ➡ <u>não há amostragem</u>
  - A aleatoriedade é devido à própria natureza incerta das variáveis (pois hoje não sabemos qual vai ser, por exemplo, o valor exato do PIB amanhã) ➡ só se conhecem os valores de uma variável depois de realizados/observados
  - A <u>amostra</u> pode ser considerada como uma das possíveis sequências que o processo estocástico (variável) realizou dentro de todas as sequências possíveis

7.1. A natureza dos dados de séries temporais

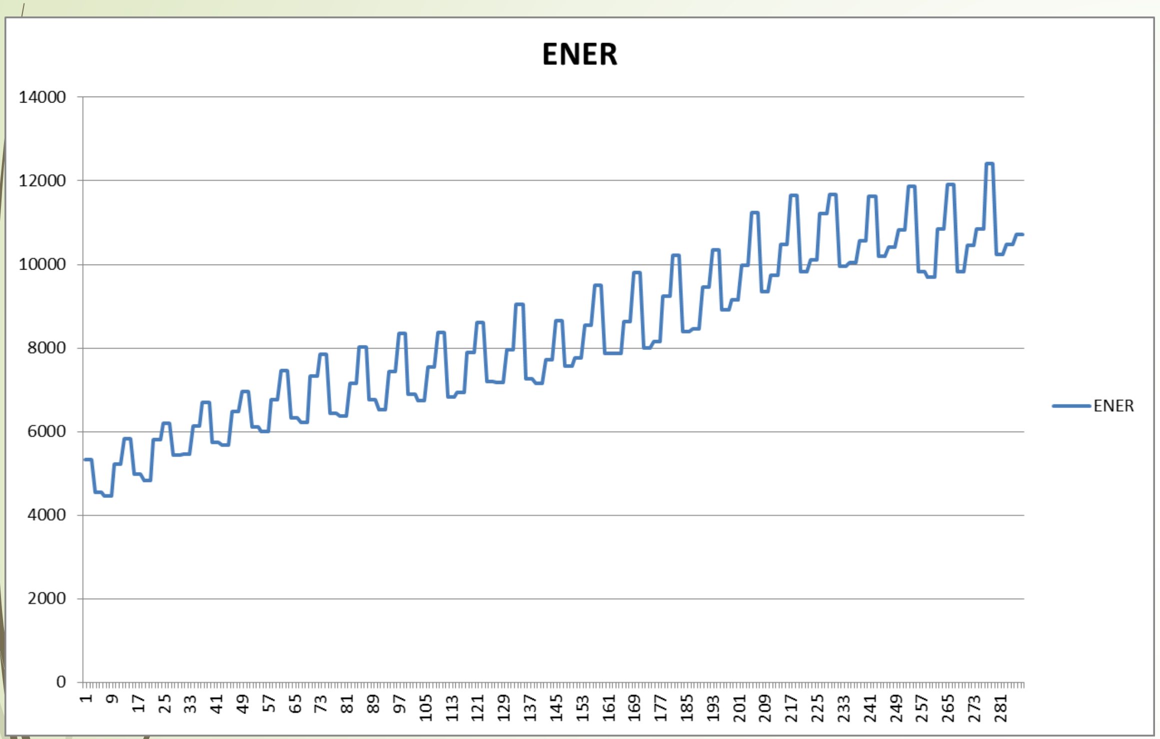
Exemplos:

year	rdis	rcons
1959	1530.1	1293.7
1960	1565.4	1332.8
1961	1615.8	1373.2
1962	1693.7	1429.9
1963	1755.5	1478.7
1964	1881.9	1560.9
1965	2002.2	1643.9
1966	2106.6	1730.2
1967	2198.4	1789
1968	2298.2	1876.5
1969	2373.6	1949.4
1970	2465.6	2014.5
1971	2564	2072.4

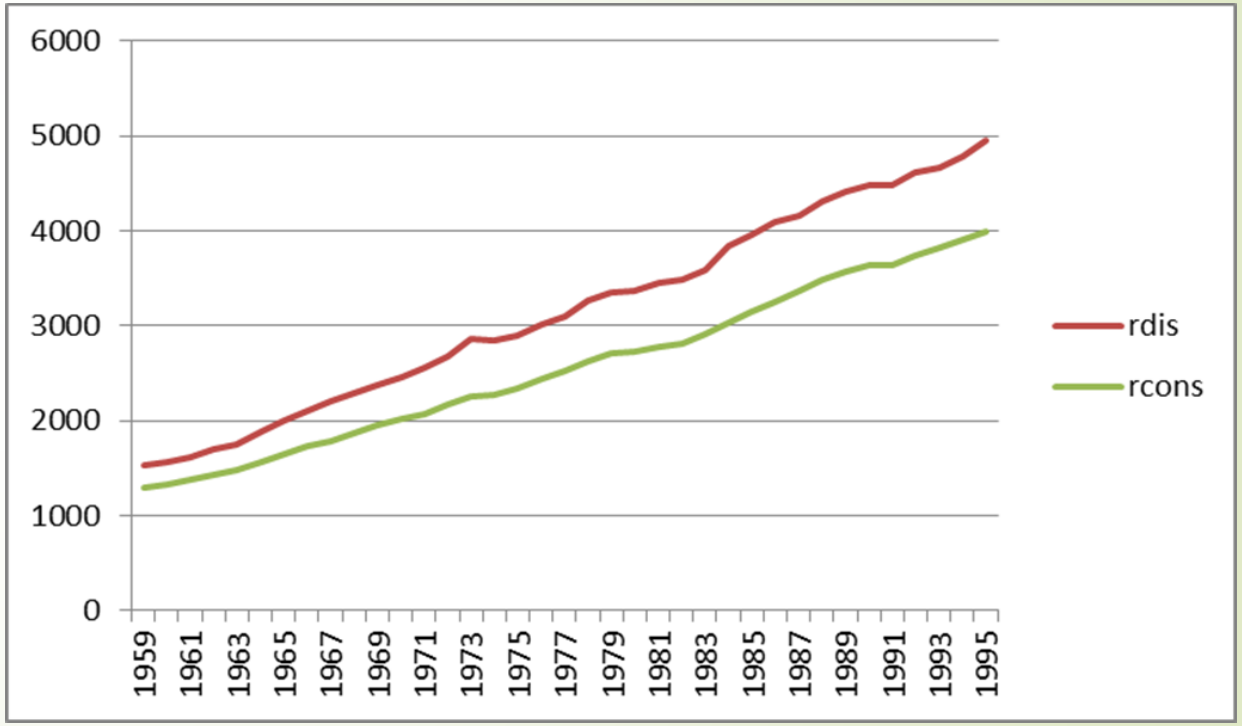
A modelação com séries temporais consiste em captar
A dependência de uma variável com o seu comportamento passado e/ou os valores presentes e passados de outras variáveis



7.1. A natureza dos dados de séries temporais



Tendências e sazonalidade são características comuns nas séries temporais que devem ser integradas na modelação



7.2. Modelos dinâmicos

- **Modelos estáticos**

  os valores de uma variável são função dos valores de variáveis explicativas no mesmo período de tempo

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 r_t + u_t \quad t = 1970, 1971, \ldots, 2015$$

$$inf_t = \beta_0 + \beta_1 Des_t + u_t \quad t = 1, 2, \ldots, T$$

- **Modelos dinâmicos**
  - os valores de uma variável são função dos valores de variáveis explicativas em períodos anteriores

    $\Downarrow$

    <u>modelos de desfasamentos distribuídos em número finito (FDL)</u>
  - os valores de uma variável são função dos seus valores passados e dos valores de variáveis explicativas no mesmo período de tempo e em períodos anteriores

    $\Downarrow$

    <u>modelos autorregressivos de desfasamentos distribuídos em número finito (AFDL)</u>

7.2. Modelos dinâmicos

- **Modelos de desfasamentos distribuídos em número finito (FDL)**
- Exemplo: $inv_t = \beta_0 + \beta_1 pop_t + \beta_2 pop_{t-1} + \beta_3 prec_t + \beta_4 prec_{t-1} + u_t$

Investimento real em habitação em milhões \$

População em milhares

Índice de preços da habitação

- **Estudo do modelo FDL(q)**

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_q z_{t-q} + u_t$$

- **Efeito de um choque transitório**

Efeito de uma variação transitória unitária de z no próprio período – ef. Contemporâneo ou **multiplicador de curto prazo**

$$\frac{\partial y_t}{\partial z_t} = \delta_0 \qquad \frac{\partial y_t}{\partial z_{t-s}} = \delta_s$$

Em geral, o efeito de uma variação transitória unitária de z no período *t-s* (esgotando-se nesse período) provoca um efeito temporário na v. dependente no momento t igual ao coeficiente de z no desfasamento *t-s*

- **Efeito de um choque permanente (multiplicador de longo prazo)**

$$\frac{\partial y_t}{\partial z_{t-q}} + \ldots + \frac{\partial y_t}{\partial z_{t-1}} + \frac{\partial y_t}{\partial z_t} = \delta_q + \ldots + \delta_1 + \delta_0$$

o efeito de uma variação sustentada unitária de z em todos os períodos provoca um efeito na v. dependente no longo prazo igual à soma dos coeficientes de todos os desfasamentos relevantes

7.2. Modelos dinâmicos

- **Exemplo 1: FDL(1,2)**

$$\hat{inv}_t = -15.52 + 1.51\, pop_t - 0.51\, pop_{t-1} + 511.43\, prec_t - 432.80\, prec_{t-1} + 33.06\, prec_{t-2}$$

- **Multiplicador de impacto ou de curto prazo:**
  - Da população: 1.51 ➡ se a população variar transitoriamente uma unidade (1 milhar) mantendo tudo o resto constante, então estima-se que o investimento do próprio período varie 1.51 unidades (milhões de dólares).
  - Do índice de preços da habitação: 511.43 ➡ se o índice de preços da habitação variar transitoriamente um ponto, mantendo tudo o resto constante, então estima-se que o investimento do próprio período varie 511.43 milhões de dólares.
- **Multiplicador de Longo prazo**
  - Da população: 1.51 – 0.51= 1 ➡ se a população tiver um variação sustentada de uma unidade (1 milhar) mantendo tudo o resto constante, então estima-se que o investimento no longo prazo varie 1 unidade (milhão de dólares)
  - Do índice de preços da habitação: 511.43-432.80+33.05 =111.68 ➡ se o índice de preços da habitação tiver um variação sustentada de um ponto, mantendo tudo o resto constante, então estima-se que o investimento no longo prazo varie 111.68 unidades (milhões de dólares)

7.2. Modelos dinâmicos

- **Exemplo 2: FDL(1,2)**

$$\log(\hat{inv})_t = -8.49 + 1.30\log(pop)_t + 0.38\log(pop)_{t-1} + 0.12\,prec_t - 0.07\,prec_{t-1} + 0.02\,prec_{t-2}$$

- **Multiplicador de impacto ou de curto prazo:**
  - Da população: 1.30 $\Rightarrow$ **elasticidade de curto prazo:** se a população variar transitoriamente 1% mantendo tudo o resto constante, então estima-se que o investimento do próprio período varie 1.30%
  - Do índice de preços da habitação: 0.12 $\Rightarrow$ se o índice de preços da habitação variar transitoriamente um ponto, mantendo tudo o resto constante, então estima-se que o investimento do próprio período varie aproximadamente 12%.
- **Multiplicador de Longo prazo**
  - Da população: 1.30+0.38= 1.68 $\Rightarrow$ **elasticidade de longo prazo** se a população tiver um variação sustentada de 1% mantendo tudo o resto constante, então estima-se que o investimento no longo prazo varie 1.68%
  - Do índice de preços da habitação: 0.12 - 0.07 + 0.02 = 0.07 $\Rightarrow$ se o índice de preços da habitação tiver um variação sustentada de um ponto, mantendo tudo o resto constante, então estima-se que o investimento no longo prazo varie aproximadamente 7%

7.2. Modelos dinâmicos

- **Modelos autorregressivos de desfasamento distribuídos**
- Exemplo: $invpc_t = \alpha_0 + \alpha_1 invpc_{t-1} + \delta_0 prec_t + \delta_1 prec_{t-1} + u_t$

Investimento real *per capita* em habitação em milhões \$

Índice de preços da habitação

- **Estudo do modelo AFDL(p,q)**

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_{t-p} y_{t-p} + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_{t-q} z_{t-q} + u_t$$

- **Efeito de um choque transitório**

Efeito de uma variação transitória unitária de z no próprio período – ef. Contemporâneo ou de **curto prazo**

$$\frac{\partial y_t}{\partial z_t} = \delta_0$$

$$\frac{\partial y_t}{\partial z_{t-s}} = \delta_s$$

Em geral, o efeito de uma variação transitória unitária de z no período t-s (esgotando-se nesse período) provoca um efeito temporário na v. dependente no momento t

- **Efeito de um choque permanente (efeito de longo prazo)**

$$\frac{\delta_0 + \delta_1 + \ldots + \delta_q}{1 - \alpha_1 - \alpha_2 - \ldots - \alpha_p}$$

o efeito de uma variação sustentada unitária de z em todos os períodos provoca um efeito na v. dependente no longo prazo igual à soma dos coeficientes de todos os lags relevantes dividida por 1 menos a soma dos coeficientes de todos os lags de y

7.2. Modelos dinâmicos

- **Exemplo**

$$\widehat{invpc}_t = 0.04 + 0.71 invpc_{t-1} + 1.79\, prec_t - 3.86\, prec_{t-1} + 2.20\, prec_{t-2}$$

- **Multiplicador ou efeito de curto prazo** do índice de preços: 1.79

  se o índice de preços da habitação variar transitoriamente um ponto, mantendo tudo o resto constante, então estima-se que o investimento *per capita* do próprio período varie 1.79 milhões de dólares.

- **Multiplicador total ou de longo prazo** do índice de preços:

$$\frac{1.79 - 3.86 + 2.20}{1 - 0.71} = \frac{0.13}{0.29} = 0.448$$

se o índice de preços da habitação tiver um variação sustentada de um ponto, mantendo tudo o resto constante, então estima-se que o investimento *per capita* no longo prazo varie 0.448 unidades (milhões de dólares)

7.3 Propriedades do OLS em amostras finitas

- **Hipóteses**

- **ST.1: Modelo linear nos parâmetros**

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t$$

os processos estocásticos $y_t, x_{t1}, \ldots, x_{tk}$ são observados, o erro u_t não é observado. As variáveis explicativas podem ser funções ou valores desfasados da mesma variável explicativa ou valores desafasados da variável dependente

- **ST.2: Não existe multicolinearidade perfeita**

Na amostra nenhuma variável explicativa pode ser constante (supondo que existe termo independente) ou ser uma função linear das outras variáveis explicativas

- **ST.3: média condicionada igual a zero**

$$E(u_t \mid \mathbf{X}) = 0 \quad \text{com} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

7.3 Propriedades do OLS em amostras finitas

- **Observações sobre TS.3: noções de exogeneidade das variáveis**
- **Exogeneidade contemporânea:**

  $E(u_t \mid \mathbf{x}_t) = 0$ ➡ As variáveis explicativas não estão correlacionadas com o erro do próprio período

- **Exogeneidade estricta:**

  $E(u_t \mid \mathbf{X}) = 0$ ➡ As variáveis explicativas em todos os períodos não estão correlacionadas com o erro de um dado período

- **Variáveis pré-determinadas - exogeneidade sequencial:**

  $E(u_t \mid \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-s}, \ldots, \mathbf{x}_1) = 0$ ➡ O erro não está correlacionado com as variáveis explicativas do mesmo período e dos períodos anteriores

- **Outras observações:**
  - A hipótese TS.3 não permite que exista o mecanismo de *feedback* entre as variáveis
  - Se o erro está relacionado com valores passados das variáveis explicativas então estas devem entrar na especificação da variável dependente para o momento *t*

7.3 Propriedades do OLS em amostras finitas

- **Teorema: estimador OLS centrado**

  Verificando-se as hipóteses TS.1 a TS.3 então o estimado OLS dos coeficientes é **centrado** $E(\hat{\beta}_j)=\beta_j \quad j=1,\dots,k$

- **Hipótese TS.4: homocedasticidade**

$$Var(u_t \mid \mathbf{X}) = Var(u_t) = \sigma^2$$

A variância dos erros não pode depender das variáveis explicativas em qualquer dos períodos e é constante no tempo

- **Hipótese TS.5: ausência de autocorrelação**

$$Cov(u_t, u_s \mid \mathbf{X}) = Cov(u_t, u_s) = 0 \quad t \neq s$$

Condicionado às variáveis explicativas em qualquer dos períodos os erros não podem ser correlacionados no tempo

- **Observações sobre TS.5**
  - Para os dados seccionais não era necessário pois as observações são independentes dado o processo de amostragem
  - Muitas vezes esta hipótese não se verifica, nomeadamente se existem variáveis omitidas que estão correlacionadas no tempo: **autocorrelação** $\Rightarrow$ **má especificação do modelo**
  - Geralmente a autocorrelação desaparece quando se introduzem mais desfasamentos no modelo das variáveis explicativas e, eventualmente, da variável dependente

7.3 Propriedades do OLS em amostras finitas

- **Teorema: variância do estimador OLS**

$$Var(\hat{\beta}_j \mid \mathbf{X}) = \frac{\sigma^2}{SST_j(1-R_j^2)}$$

fórmula igual à obtida para os dados seccionais

- **Estimador da variância**

$$V\hat{a}r(\hat{\beta}_j \mid \mathbf{X}) = \frac{\hat{\sigma}^2}{SST_j(1-R_j^2)} \quad \text{com} \quad \hat{\sigma}^2 = \frac{\sum_{t=1}^{n} \hat{u}_t^2}{n-k-1}$$

fórmula igual à obtida para os dados seccionais

- **Teorema:** sob as Hipóteses TS.1 a TS.5 então $E(\hat{\sigma}^2) = \sigma^2$

- **Teorema de Gauss-Markov:** Sob as hipóteses TS.1 a TS.5 o estimador OLS é o estimador mais eficiente na classe dos estimadores centrados e lineares em *y* (tal como nos dados seccionais)

- **Hipótese TS.6:**

$$u_t \sim N(0, \sigma^2) \text{ independente de } \mathbf{X}$$

Esta hipótese implica as hipóteses TS.3 a TS.5

7.3 Propriedades do OLS em amostras finitas

- **Distribuições para inferência sobre os parâmetros**

  Supondo que se verificam as hipóteses TS.1 a TS.6 as distribuições habituais t e F são válidas. Toda a inferência habitual (vista no capítulo 3) é válida.

- **Exemplos:** Curva de Phillips

$$\hat{inf}_t = \underset{(0.289)}{1.42} + 0.468 Desemp_t \qquad R^2 = 0,053 \qquad n = 49$$

Contrariamente ao estabelecido pela teoria económica não se verifica um *tradeoff* entre desemprego e inflação

Discussão das hipóteses

- TS1: $inf_t = \beta_0 + \beta_1 Desemp_t + u_t$

O erro contém fatores como choques na economia, como choques monetários, choques decorrentes de grandes variações do preço do petróleo, choques de alterações na procura/rendimento, da oferta; etc…

- TS2: verifica-se
- TS3: $E\left(u_t \mid Desemp_1, \ldots, Desemp_t, \ldots, Desemp_n\right) = 0$

  Esta hipótese é facilmente violada. Por exemplo: se o Desemprego em t-1 aumentou significativamente pode provocar um choque na procura em t que se repercute em u_t e provocando uma diminuição da inflação. Conclusão: u_t e $Desemp_{t-1}$ estão correlacionados levando a que TS3 não se verifique.

7.3 Propriedades do OLS em amostras finitas

- TS4: $Var\left(u_t \mid Desemp_1,\ldots,Desemp_t,\ldots,Desemp_n\right)=\sigma^2$

Pode ser violada se a política económica que influencia a inflação for mais instável quando o desemprego é elevado

- TS5: $corr\left(u_t,u_s \mid Desemp_1,\ldots,Desemp_t,\ldots,Desemp_n\right)=0 \; t\neq s$

Pode ser violada se, por exemplo, os choques económicos que influenciam a inflação (para além do desemprego do próprio período) forem persistentes no tempo

- TS6: $u_t \mid Desemp_1,\ldots,Desemp_t,\ldots,Desemp_n \;\sim\; N(0,\sigma^2)$

Se TS3, TS4 e/ou TS5 não se verificarem então TS6 é violada. Também pode ser discutível que a distribuição (não condicionada) dos erros seja normal

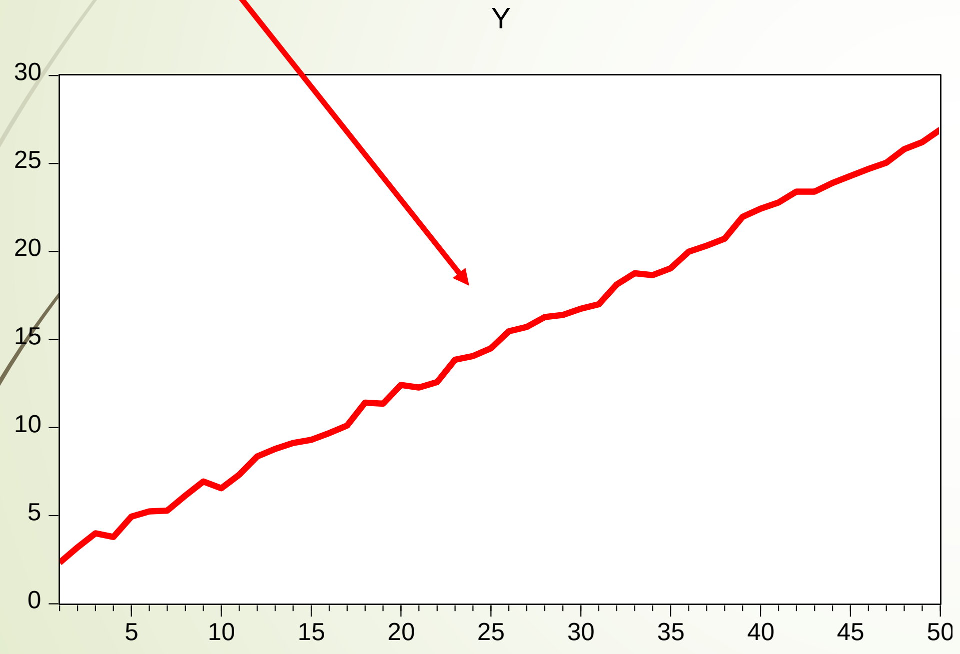
- Conclusão: Como e suspeita que TS3 não se verifica então o OLS pode ser enviesado o que explica a estimativa positiva do coeficiente da taxa de desemprego quando a teoria económica aponta para um coeficiente negativo. Por outro lado, a suspeita de que TS4, TS5 e TS6 também não se verificam invalida a estimativa do erro padrão e toda a inferência habitual sobre os coeficientes.

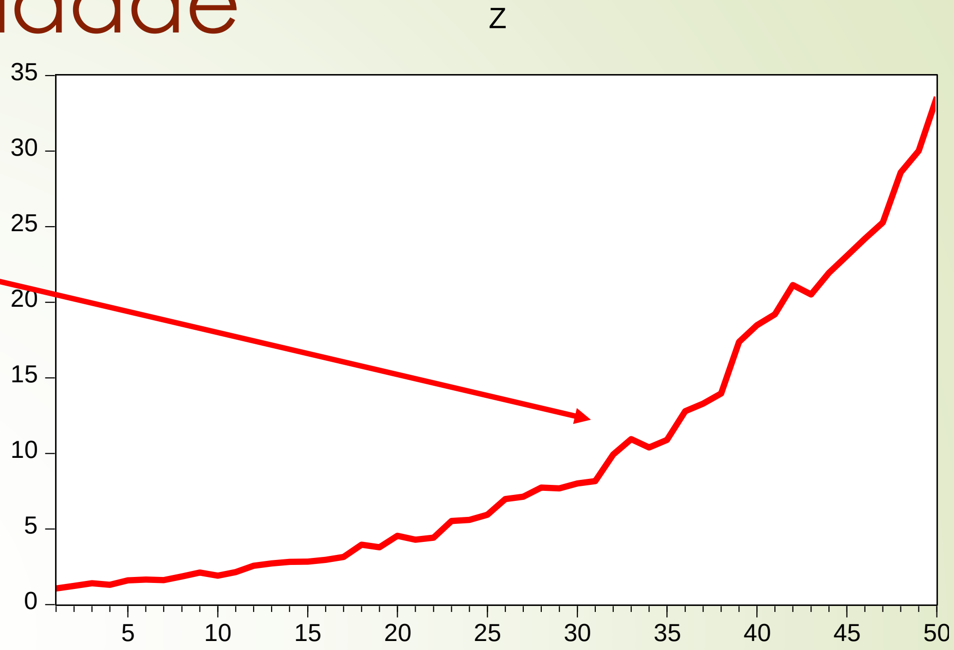
7.4 Tendência e Sazonalidade

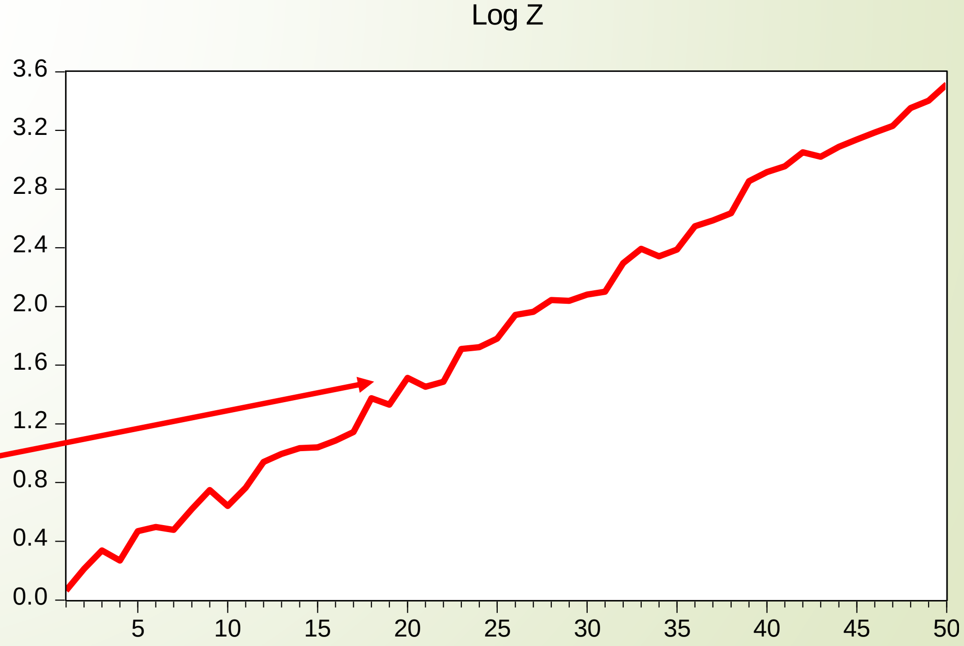
- **Tendências**

Tendência exponencial

Tendência linear







Tendência linear

7.4 Tendência e Sazonalidade

- **Tendência linear**

$$y_t = \alpha + \delta t + u_t \quad t = 1,2,\dots,n$$

δ é a variação média no período em unidades de y por unidade de tempo

- **Exemplo:**

$$\hat{y}_t = 1103.96 + 76.68t$$

Estima-se que em média consumo real varie 76.68 bilhões de $ por ano

Consumo real dos EUA em bilhões $

- **Tendência exponencial**

$$\log(y_t) = \alpha + \delta t + u_t \quad t = 1,2,\dots,n$$

$\delta \times 100\%$ é a taxa de crescimento média de y no período por unidade de tempo

- **Exemplo**

$$\hat{\log}(y_t) = 7.19 + 0.032t$$

Estima-se que o consumo real cresça à taxa média de 3.2% ao ano

7.4 Tendência e Sazonalidade

- **Regressão com variáveis com tendência**

- Se as variáveis explicativas e/ou a variável dependente têm tendência então <u>deve-se adicionar a tendência</u> na especificação do modelo a estimar se não corre-se o risco de ter uma **regressão espúria**.

- **Exemplo**

Investimento per capita em habitação

Índice de preços da habitação

$$\hat{\log}(invpc) = 0.550 + \underset{(0.382)}{1.241}\log(price) \quad n = 42 \quad R^2 = 0,208$$

Regressão espúria

$$\hat{\log}(invpc) = -0.913 + \underset{(0.679)}{0.381}\log(price) + \underset{(0.0035)}{0.0098}t \quad n = 42 \quad R^2 = 0.341$$

Incluindo a tendência deixa de existir uma relação estatisticamente significativa entre $\log(invpc)$ e $\log(price)$

7.4 Tendência e Sazonalidade

- **Sazonalidade**

- Dados trimestrais

v. *Dummy* igual a 1 se a observação t pertence ao trimestre 3

$$y_t = \beta_0 + \delta_1 trimestre1_t + \delta_2 trimestre2_t + \delta_3 trimestre3_t + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$$

δ_2 representa a variação de y no trimestre 2 relativamente ao 4º trimestre (mantendo tudo o resto constante)

- Dados mensais

v. *Dummy* igual a 1 se a observação t pertence ao mês de Novembro

$$y_t = \beta_0 + \delta_1 Jan_t + \delta_2 Fev_t + \ldots + \delta_{11} Nov_t + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$$

δ_{11} representa a variação de y em Novembro relativamente a Dezembro (mantendo tudo o resto constante)

- Teste de sazonalidade

$H_0 : \delta_1 = \delta_2 = \delta_3 = 0$ Dados trimestrais

$H_0 : \delta_1 = \delta_2 = \ldots = \delta_{11} = 0$ Dados mensais

Teste F

7.4 Tendência e Sazonalidade

- **Exemplo**

$n = 79,\ R^2 = 0.914 \quad \widehat{lvendas}_t = 36.11 + 0.014t + 0.21Q1_t + 0.015Q2_t - 0.048Q3_t$

Logaritmo das vendas de uma empresa no trimestre t

Dummies trimestrais

$n = 79,\ R^2 = 0.899 \quad \widehat{lvendas}_t = 21.51 + 0.03t$

- **Teste de Sazonalidade**

$H_0 : \delta_1 = \delta_2 = \delta_3 = 0 \qquad H_1 : H_0$ falsa

$F_{obs} = 4.302 \qquad p_{obs} = 0.007$ ➡ Evidência de sazonalidade

- **Interpretação dos parâmetros**
  - No 1º trimestre as vendas são maiores aproximadamente 21% relativamente às vendas do 4º trimestre, depois de removida a tendência
  - No 2º trimestre as vendas são maiores aproximadamente 1.5% relativamente às vendas do 4º trimestre, depois de removida a tendência
  - No 3º trimestre as vendas são menores aproximadamente 4.8% relativamente às vendas do 4º trimestre, depois de removida a tendência

7.5 Modelação de acontecimentos

- **Exemplo**
- Krupp and Pollard (1996) analyzed the effects of antidumping filings by U.S. chemical industries on imports of various chemicals.
- We focus here on barium chloride, a cleaning agent used in various chemical processes and in gasoline production.
- Important events:
- In the early 1980s, U.S. barium chloride producers believed that China was offering its U.S. imports at an unfairly low price (an action known as dumping), and the barium chloride industry filed a complaint with the U.S. International Trade Commission (ITC) in October 1983.
- The ITC ruled in favor of the U.S. barium chloride industry in October 1984.

7.5 Modelação de acontecimentos

- Questions
  - Were imports unusually high in the period immediately preceding the initial filing?
  - Did imports change noticeably after an antidumping filing?
  - What was the reduction in imports after a decision in favor of the U.S. industry?
  - Need of dummy variables (following Krupp and Pollard)
    - **befile6:** is equal to 1 during the six months before filing
    - **affile6:** is equal to 1 during the six months after filing
    - **afdec6:** is equal to 1 the six months after the positive decision

7.5 Modelação de acontecimentos

- The model:Using monthly data from February 1978 through December 1988

$$\hat{\log}(chnimp) = -17.80 + 13.12 \log(chempi) + 0.196 \log(gas)$$
$$(21.05) \quad (0.48) \qquad\qquad (0.907)$$
$$+0.983 \log(rtwex) + 0.060 befiled6 - 0.032 affile6 - 0.565 afdec6$$
$$(0.400) \qquad (0.261) \qquad (0.264) \qquad (0.286)$$
$$n = 131,\ R^2 = 305$$

- **chnimp** – the volume of imports of barium chloride from China,
- **chempi** – the index of chemical production (to control for overall demand for barium chloride), defined to be 100 in June 1977
- **gas** - the volume of gasoline production, (another demand variable)
- **rtwex** - the exchange rate index which measures the strength of the dollar against severalother currencies.

7.5 Modelação de acontecimentos

- **Conclusions**
  - befile6 is statistically insignificant, so there is no evidence that Chinese imports were unusually high during the six months before the suit was filed.
  - although the estimate on affile6 is negative, the coefficient is small (indicating about a 3.2% fall in Chinese imports), and it is statistically very insignificant.
  - The coefficient on afdec6 shows a substantial fall in Chinese imports of barium chloride after the decision in favor of the U.S. industry. The exact percentage change is -43.2%. The coefficient is statistically significant at the 5% level against a two-sided alternative.

7.5 Modelação de acontecimentos

- **Conclusions:**
  - coefficient signs on the control variables are what we expect:
  - an increase in overall chemical production increases the demand for the cleaning agent.
  - Gasoline production does not affect Chinese imports significantly.
  - The coefficient on log(rtwex) shows that an increase in the value of the dollar relative to other currencies increases the demand for Chinese imports, as is predicted by economic theory