
Séries Temporais para Finanças

2019/2020

João Nicolau

1 Introdução

1.1 Objecto e Método

Em muitos aspectos STF confunde-se com Econometria Financeira.

“Financial econometrics is simply the application of econometric tools to financial data” (Engle, 2001).

Vamos estudar alguns métodos econométricos para séries temporais financeiras. Estes métodos servem de suporte a variadíssimos estudos:

- Modelação dos retornos e previsão da volatilidade;
- Avaliação do risco (por exemplo, através do *Value at Risk*);
- Avaliação de opções, etc.;
- Selecção e gestão de *portfolios*;
- Análise da previsibilidade e eficiência dos mercados, etc.;

1.2 Séries Financeiras versus Séries Macroeconómicas

Principais diferenças entre as séries financeiras e séries macroeconómicas:

- dados de natureza macroeconómica (consumo, produto, taxa de desemprego) podem ser observados mensalmente, trimestralmente ou anualmente; dados financeiros, como por exemplo, retornos de acções ou taxas de câmbio podem ser observados com uma frequência muito superior; nalguns casos, com intervalos de minutos ou segundos entre duas observações consecutivas;
- o número de observações disponíveis é muito maior com séries financeiras;
- os dados macroeconómicos são menos fiáveis
- as séries financeiras exibem habitualmente fortes efeitos não lineares e distribuições não normais.

1.3 Preços e Retornos

Ponto de partida:

$$P_1, P_2, \dots, P_n$$

(por exemplo, a série das cotações de fecho de um ativo na bolsa).

Depois: calculam-se os retornos.

Duas formas de calcular os retornos:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$
$$r_t = \log P_t - \log P_{t-1}$$

Para obter estas expressões considere-se:

- P_0 (valor inicial) P_1 (valor final)
- Retorno Simples R . O capital valoriza-se apenas uma vez $P_1 = P_0(1 + R)$. Resolvendo em ordem a R

$$R = \frac{P_1 - P_0}{P_1}$$

- Retorno composto. O capital valoriza-se n vezes $P_1 = P_0 \left(1 + \frac{r}{n}\right)^n$. Admita-se $n \rightarrow \infty$ (valorização contínua). Tem-se

$$P_1 = \lim P_0 \left(1 + \frac{r}{n}\right)^n = P_0 e^r.$$

Resolvendo $P_1 = P_0 e^r$ em ordem a r obtém-se

$$r = \log \left(\frac{P_1}{P_0} \right) = \log(P_1) - \log(P_0).$$

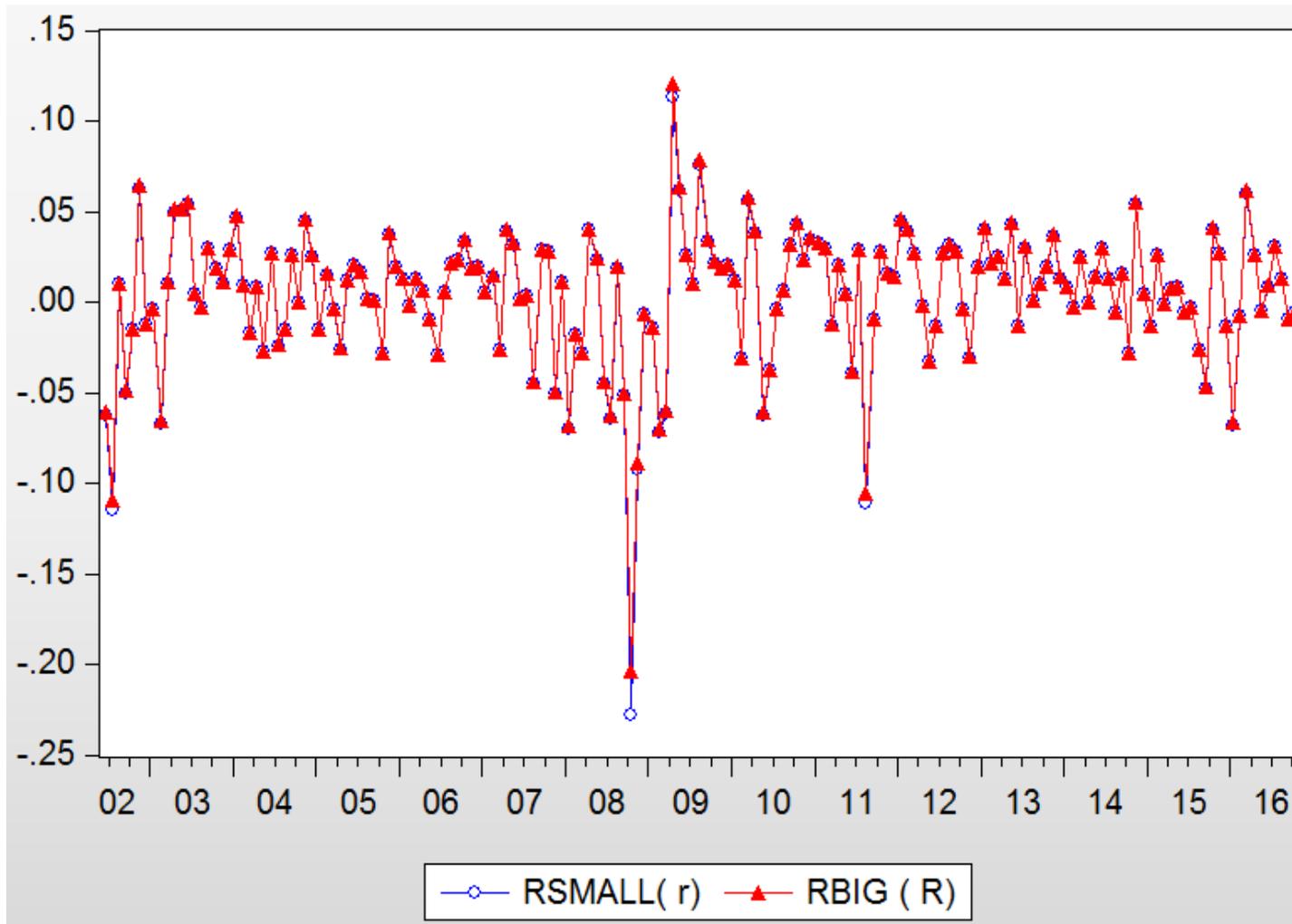
Relação entre R_t e r_t :

$$\begin{aligned}r_t &= \log P_t - \log P_{t-1} \\ &= \log \frac{P_t}{P_{t-1}} \\ &= \log \left(1 + \frac{P_t - P_{t-1}}{P_{t-1}} \right) \\ &= \log (1 + R_t) .\end{aligned}$$

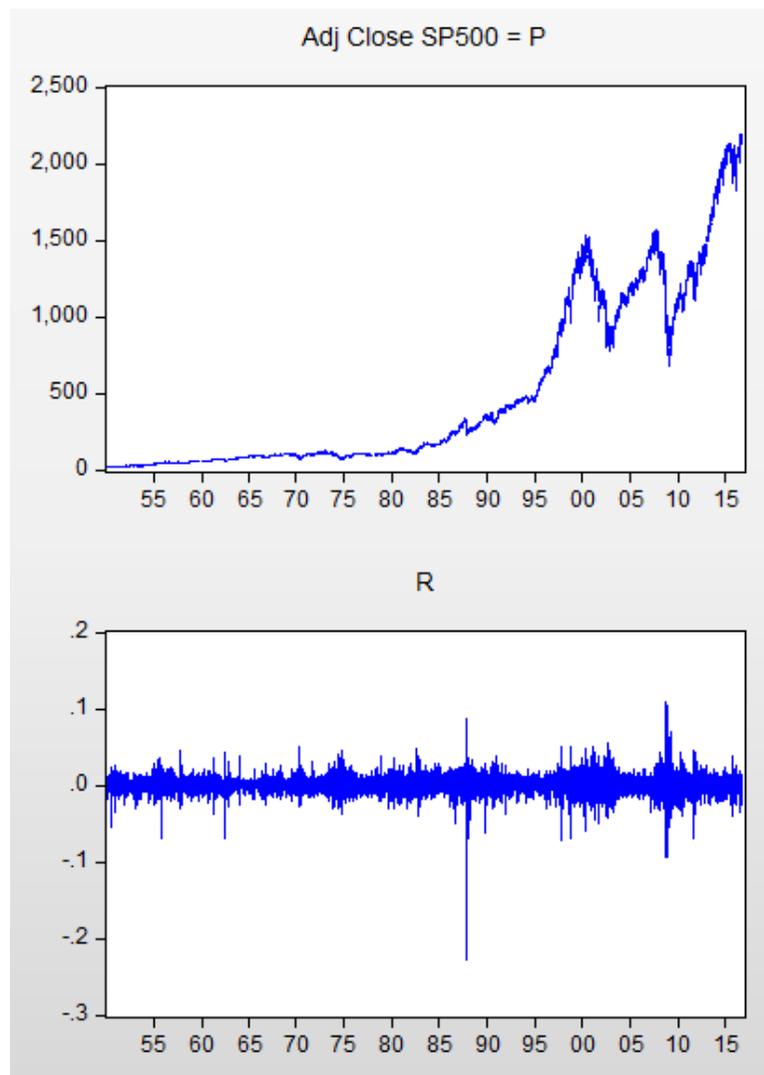
Mas, em geral,

$$r_t \simeq R_t$$

Para dados diários, semanais ou mensais $r_t \simeq R_t$. Retornos mensais do SP500:



Dados Diários SP500



Estudam-se os preços ou os retornos? Prefere-se os retornos porque:

- os retornos fornecem tanta informação quanto os preços;
- os retornos são mais fáceis de modelarem.

Estuda-se r_t ou R_t ? Prefere-se geralmente r_t (discutimos isto na secção 1.5)

Retornos Anualizados

Qual é a taxa de rendibilidade anual, r_A , tal que, aplicada a um investimento P_0 permite ao fim de T anos obter o investimento P_n ? Ou seja, qual é o valor r_A que resolve a equação

$$P_0 e^{r_A T} = P_n ? \Rightarrow r_A = \frac{1}{T} \log \left(\frac{P_n}{P_0} \right).$$

Se os preços P_0, P_1, \dots, P_n são diários e se admitirmos que num ano se observam 250 preços, então deduz-se a relação $T = n/250$. Nestas condições,

$$r_A = \frac{250}{n} \log \left(\frac{P_n}{P_0} \right).$$

Em termos gerais, se num ano se observam N preços (por exemplo, $N = 12$ se as observações são mensais) e dispomos de n observações sobre os preços, então $T = n/N$ e

$$r_A = \dots = N \bar{r}.$$

Volatilidade Anualizada

Retorno anual

$$X = \log P_N - \log P_0 = \sum_{t=1}^N r_t.$$

Supondo $E(r_1) = E(r_2) = \dots = E(r_N)$, tem-se que o retorno médio anual é dado por

$$E(X) = N E(r_t).$$

Estimativa de $E(X)$ é, precisamente, $r_A = N\bar{r}$.

Suponha-se que a sucessão $\{r_t\}$ é não autocorrelacionada e que $\text{Var}(r_t) = \sigma^2$. Nestas condições, a variância anualizada é dada por

$$\text{Var}(X) = \text{Var}\left(\sum_{t=1}^N r_t\right) = N\sigma^2.$$

Entre os investidores fala-se numa “volatilidade anualizada” dada por

$$\sqrt{N}\sigma$$

onde σ é o desvio padrão associado a um medida intra anual.

Dados diários, vol. anualizada $\sqrt{250}\sigma_d$ (σ_d é o dp associado aos dados diários);

Dados mensais, vol. anualizada $\sqrt{12}\sigma_m$ (σ_m é o dp associado aos dados mensais).

A informação anualizada pode ser dada em percentagem:

$$N\bar{r} \times 100\%, \quad \sqrt{N}\sigma \times 100\%.$$

1.4 Retorno Multi-Períodos

$\{P_1, P_2, \dots, P_n\} \rightarrow$ preços diários. Retorno semanal (supondo 1 semana = 5 dias)? Basta considerar

$$r_t(5) = \log \left(\frac{P_t}{P_{t-5}} \right)$$

De uma forma geral,

$$r_t(m) = \log \left(\frac{P_t}{P_{t-m}} \right) = \log(P_t) - \log(P_{t-m}).$$

Como calcular $r_t(m)$ a partir dos retornos em $t = 1, 2, \dots$? Para exemplificar, suponha-se que se têm retornos diários e procura-se o retorno semanal, i.e., admita-se o seguinte:

retorno 2 ^a feira	$r_1 = \log P_1 - \log P_0$
retorno 3 ^a feira	$r_2 = \log P_2 - \log P_1$
retorno 4 ^a feira	$r_3 = \log P_3 - \log P_2$
retorno 5 ^a feira	$r_4 = \log P_4 - \log P_3$
retorno 6 ^a feira	$r_5 = \log P_5 - \log P_4$
retorno da semana	
	$\log P_5 - \log P_0 = r_1 + r_2 + \dots + r_5$

A tabela anterior sugere que o retorno da semana é igual à soma dos retornos diários da semana. Com efeito,

$$\begin{aligned} \log P_5 - \log P_0 &= \underbrace{\log P_5 - \log P_4}_{r_5} + \underbrace{\log P_4 - \log P_3}_{r_4} + \underbrace{\log P_3 - \log P_2}_{r_3} \\ &\quad + \underbrace{\log P_2 - \log P_1}_{r_2} + \underbrace{\log P_1 - \log P_0}_{r_1} \end{aligned}$$

Conclui-se:

$$\log P_5 - \log P_0 = \sum_{t=1}^5 r_t$$

De uma forma geral

$$r_t(m) = r_t + r_{t-1} + \dots + r_{t-m+1}$$

1.5 Vantagens/Desvantagens de r_t versus R_t

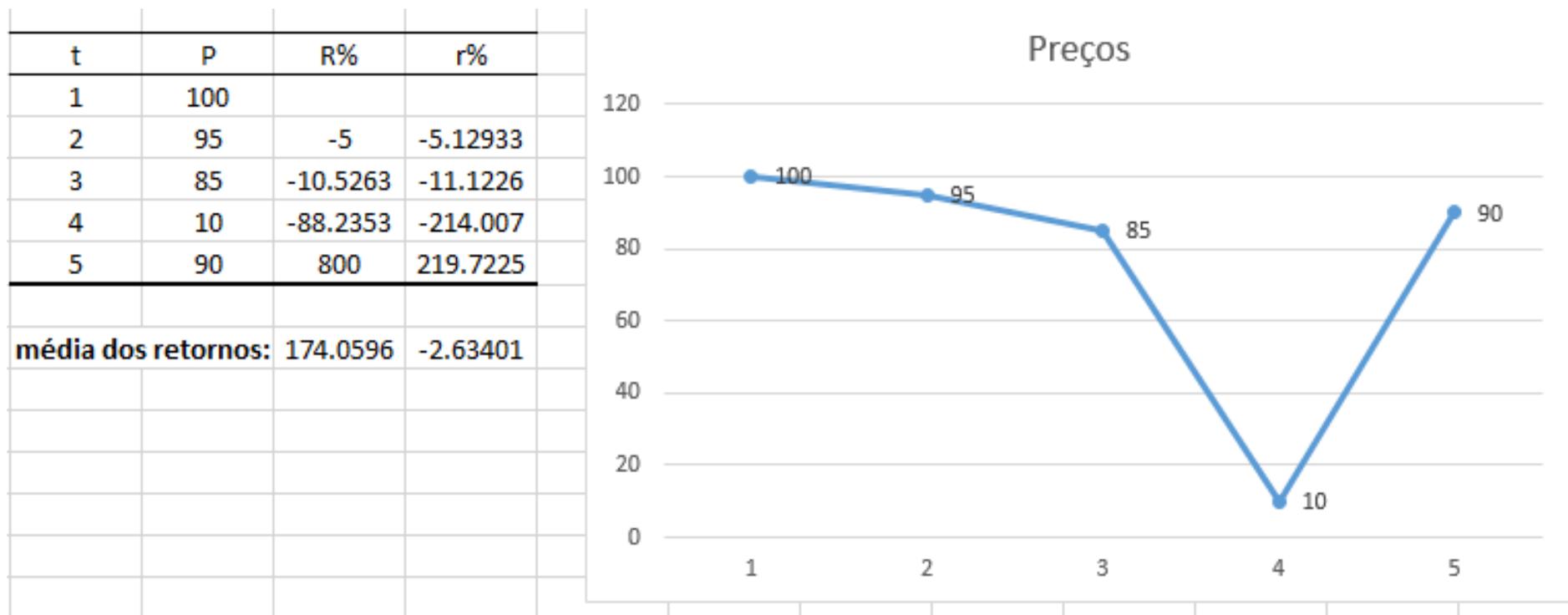
- Na maior parte dos estudos usa-se r_t
- Retornos contínuos multiperíodos são aditivos (*Time-additive*)

$$\log P_n - \log P_0 = \sum_{t=1}^n r_t.$$

- Os retornos discretos não são aditivos:

$$\frac{P_n - P_0}{P_0} = \prod_{t=1}^n (1 + R_t) - 1.$$

- Média dos retornos **simples** não fornece informação sobre a valorização do capital



- Retorno do Portfolio pode escrever-se como função dos retornos discretos. Por exemplo, com dois ativos:

$$R_{p,t} = \omega_1 R_{1,t} + \omega_2 R_{2,t}$$

mas

$$\begin{aligned} r_{p,t} &\neq \omega_1 r_{1,t} + \omega_2 r_{2,t} \\ r_{p,t} &= \log(1 + R_{p,t}) \end{aligned}$$

2 Factos Empíricos Estilizados das Séries Financeiras

2.1 Regularidade Empíricas relacionadas com a Distribuição Marginal

Comece-se por considerar a fdp marginal f de um certo retorno r_t . Estamos interessados em saber algo sobre f (que é geralmente desconhecida). Obtém-se alguma informação sobre f calculando vários momentos da amostra. Pelo método dos momentos, os parâmetros populacionais desconhecidos,

$$\mu = \mathbf{E}(r), \quad \sigma = \sqrt{\mathbf{Var}(r)}, \quad sk = \frac{\mathbf{E}\left((r - \mu)^3\right)}{\sigma^3}, \quad k = \frac{\mathbf{E}\left((r - \mu)^4\right)}{\sigma^4}$$

podem ser estimados de forma consistente, respectivamente, pelos estimadores

$$\begin{aligned} \bar{r} &= \frac{\sum_{t=1}^n r_t}{n}, \quad \hat{\sigma} = \sqrt{\frac{\sum_{t=1}^n (r_t - \bar{r})^2}{n}}, \\ \widehat{sk} &= \frac{n^{-1} \sum_{t=1}^n (r_t - \bar{r})^3}{\hat{\sigma}^3}, \quad \hat{k} = \frac{n^{-1} \sum_{t=1}^n (r_t - \bar{r})^4}{\hat{\sigma}^4}. \end{aligned}$$

Os factos empíricos estilizados que descreveremos a seguir envolvem explicitamente estes momentos. Concretamente, mostraremos a seguir que

- \bar{r} tende a ser maior do que o retorno do investimento sem risco;
- $\hat{\sigma}$ depende da natureza do activo;
- \widehat{sk} tende a ser negativo;
- \hat{k} tende a ser superior a 3.

2.1.1 Prémio de Risco Positivo

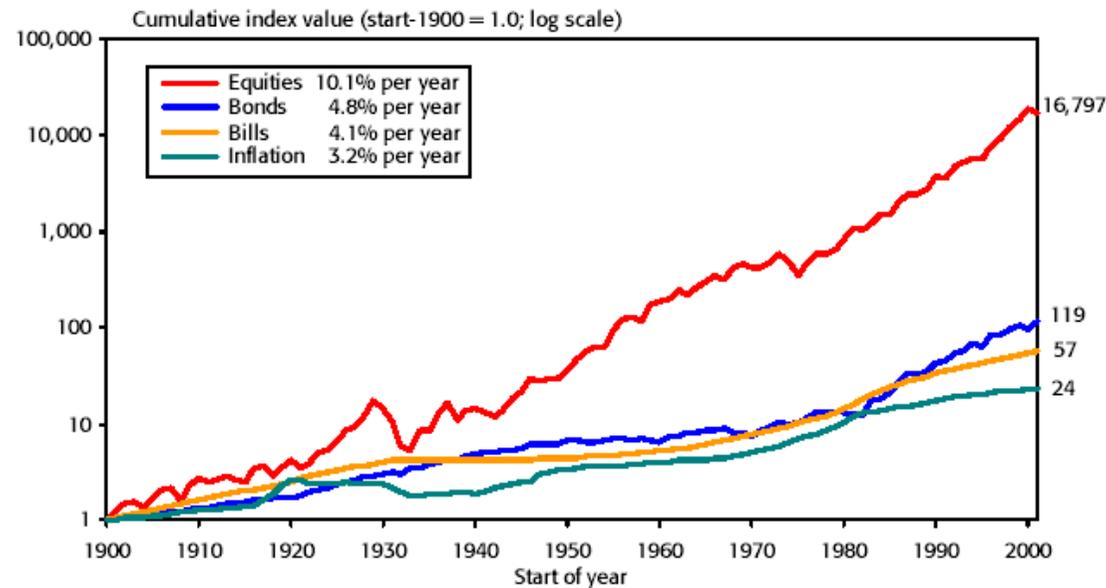
PR ou *Equity Risk Premia*

valor esperado do retorno de um investimento no mercado de capitais - o retorno do investimento sem risco.

De acordo com a teoria financeira este prémio deve ser positivo...

Dimson, Marsh e Staunton (2002) fizeram o seguinte exercício:

Figure 4-1: Cumulative return on US asset classes in nominal terms, 1900–2000



Um investimento de 1 dólar em acções norte americanas no início do ano 1900 (num índice representativo do mercado de acções norte-americano) seria vendido por 16797 dólares em 2000. Um investimento de 1 dólar em bilhetes do tesouro (investimento sem risco), seria vendido por 119 dólares. Os retornos anualizados são respectivamente

$$r_A 100\% = \frac{1}{101} \log \left(\frac{16796}{1} \right) 100\% = 9.6\%$$

$$r_A 100\% = \frac{1}{101} \log \left(\frac{119}{1} \right) 100\% = 4.7\%$$

2.1.2 Desvios Padrão Diferentes Consoante os Activos

	$r_A\%$	$\hat{\sigma}_A\%$	\widehat{sk}	\hat{k}	$\frac{\hat{P}(r_t - \bar{r} > 3\hat{\sigma})}{P(Z > 3)}$
<i>Cotações de Acções</i>					
Microsoft (01-88 a 7-05)	23.9%	36.3%	-0.097	6.8	3.85
Coca-Cola (11-86 a 12-05)	7.2%	33.2%	-0.97	18.1	4.53
PT (6-95 a 12-05)	12.3%	33.2%	-0.06	6.39	5.02
<i>Índices Bolsistas</i>					
Dax (11-90 a 11-05)	8.4%	22.1%	-0.196	6.66	5.79
CAC40 (03-90 a 11-05)	5.7%	20.5%	-0.101	5.77	5.61
Nikkei225 (01-84 a 11-05)	1.6%	20.5%	-0.116	10.77	3.71
FTSE100 (04-84 a 11-05)	7.3%	15.8%	-0.545	11.12	2.44
PSI20 (01-93 a 03-06)	8.6%	15.8%	-0.61	10.9	5.99
<i>Taxas de Câmbio</i>					
USD/EUR(12-98 a 11-05)	0.1%	7.9%	-0.05	4.86	4.55
YEN/USD (1-71 a 7-05)	-3.4%	9.4%	-0.78	14.2	1.67

Na última coluna assume-se que $Z \sim N(0, 1)$; $\hat{\sigma}_A\% = \sqrt{250}\hat{\sigma}100\%$

Conclui-se:

- Os activos com maior variabilidade (e, portanto com maior risco associado) são os títulos de empresas, seguidos dos índices bolsistas e taxas de câmbio (bilhetes do tesouro - resultados não apresentados - apresentam a menor variabilidade).
- Vários estudos indicam que a variabilidade dos retornos tende a diminuir à medida que a dimensão das empresas aumenta.

2.1.3 Retornos de Índices tendem a Apresentar Assimetria Negativa

- A distribuição de r é assimétrica negativa (positiva) se $sk < 0$ (> 0). Se $sk = 0$ a distribuição é simétrica.
- Podemos ter uma estimativa \widehat{sk} negativa se as variações negativas fortes forem mais acentuadas do que as variações positivas fortes. Ver a tabela anterior.
- As distribuições das rendibilidades de índice bolsistas tendem a ser assimétricas negativas. As fortes variações dos preços são maioritariamente de sinal negativo. Estas variações são obviamente *crashes* bolsistas.

- Sob certas hipóteses, incluindo $\{r_t\}$ é uma sequência de v.a. homocedásticas com distribuição normal, a estatística de teste

$$Z_1 = \sqrt{n} \frac{\widehat{sk}}{\sqrt{6}}$$

tem distribuição assintótica $N(0, 1)$. A hipótese nula $H_0: sk = 0$ pode ser testada a partir deste resultado. Mas as hipóteses de partida, normalidade e homocedasticidade, são relativamente severas.

- Para as taxas de câmbio não há razão especial para esperar $sk > 0$ ou $sk < 0$

2.1.4 Retornos Apresentam Distribuições Leptocúrticas

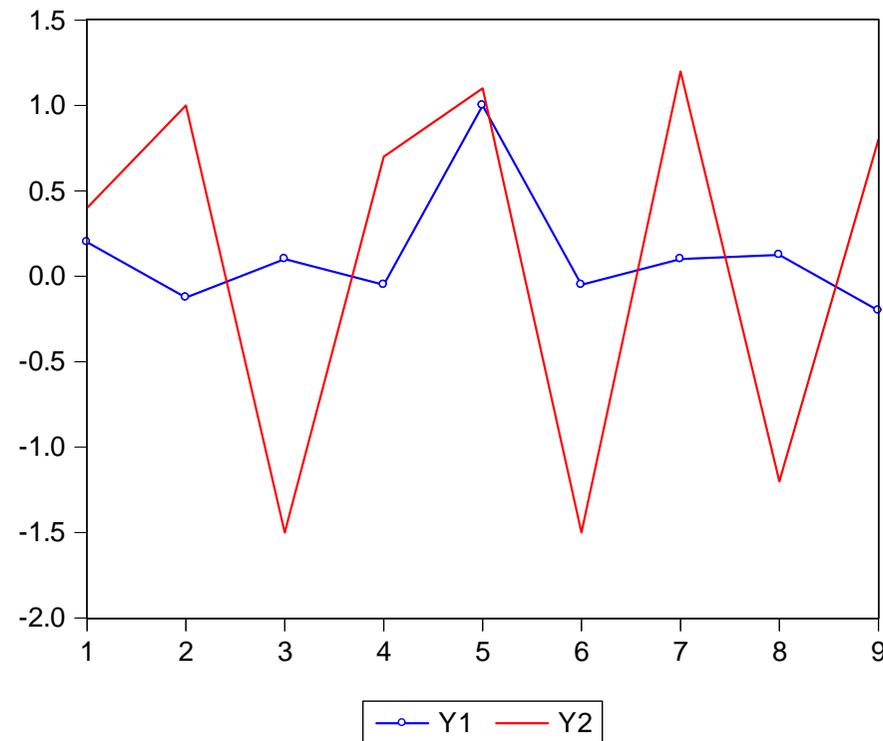
Recorde-se

$$k = \frac{\mathbb{E} \left((r - \mu)^4 \right)}{\sigma^4}$$

Diz-se que a distribuição f é:

- mesocúrtica se $k = 3$
- platicúrtica se $k < 3$
- leptocúrtica se $k > 3$.

Valores altos de \hat{k} (acima de 3) indicam que certos valores da amostra se encontram muito afastados da média, comparativamente aos demais valores da amostra. A existência de *outliers* faz aumentar a estatística \hat{k} .

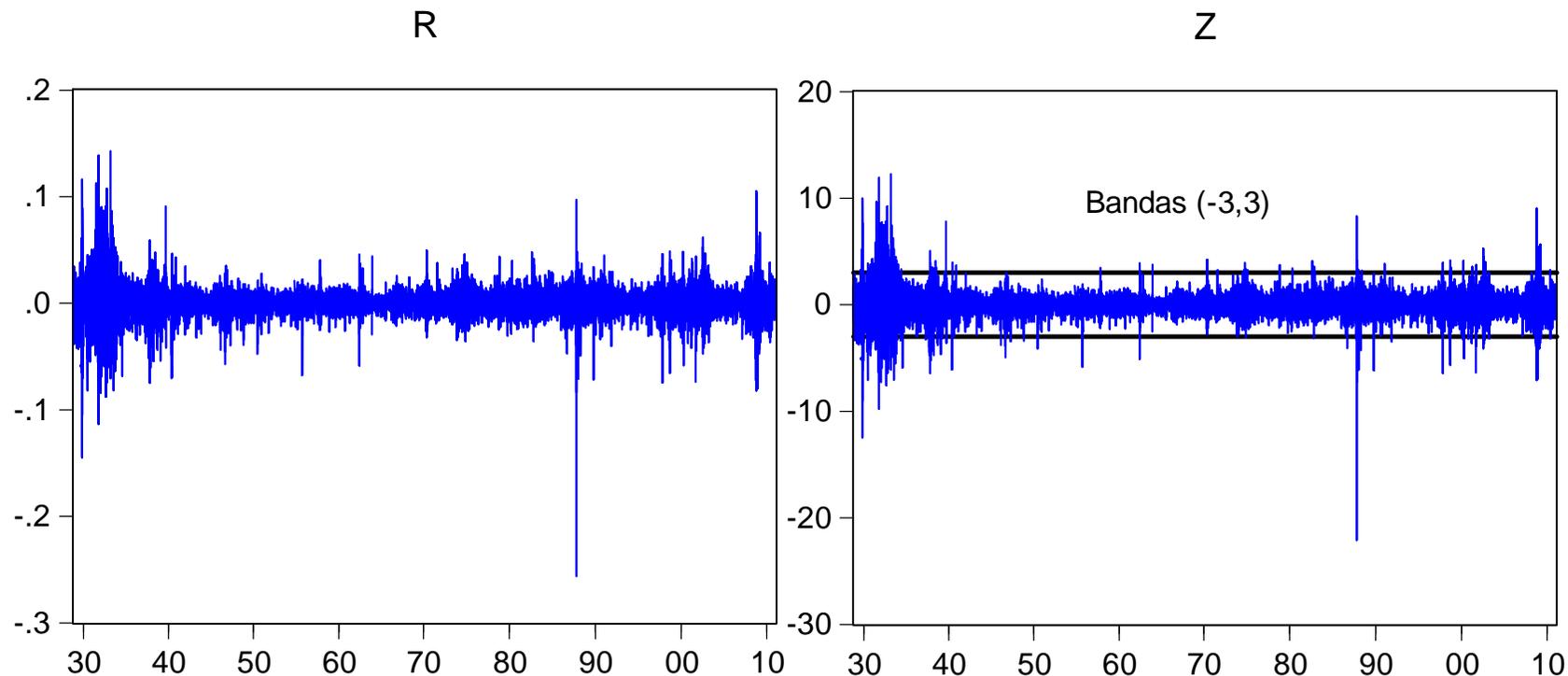


Qual das duas séries tem *kurtosis* estimada mais alta?

O coeficiente k estimado vem quase sempre (bastante) acima de 3, o que sugere que a distribuição dos retornos (de cotações, índice, taxas de câmbio e mesmo taxas de juro) é leptocúrtica.

	$r_A\%$	$\hat{\sigma}_A\%$	\widehat{sk}	\hat{k}	$\frac{\hat{P}(r_t - \bar{r} > 3\hat{\sigma})}{P(Z > 3)}$
<i>Cotações de Acções</i>					
Microsoft (01-88 a 7-05)	23.9%	36.3%	-0.097	6.8	3.85
Coca-Cola (11-86 a 12-05)	7.2%	33.2%	-0.97	18.1	4.53
PT (6-95 a 12-05)	12.3%	33.2%	-0.06	6.39	5.02
<i>Índices Bolsistas</i>					
Dax (11-90 a 11-05)	8.4%	22.1%	-0.196	6.66	5.79
CAC40 (03-90 a 11-05)	5.7%	20.5%	-0.101	5.77	5.61
Nikkei225 (01-84 a 11-05)	1.6%	20.5%	-0.116	10.77	3.71
FTSE100 (04-84 a 11-05)	7.3%	15.8%	-0.545	11.12	2.44
PSI20 (01-93 a 03-06)	8.6%	15.8%	-0.61	10.9	5.99
<i>Taxas de Câmbio</i>					
USD/EUR(12-98 a 11-05)	0.1%	7.9%	-0.05	4.86	4.55
YEN/USD (1-71 a 7-05)	-3.4%	9.4%	-0.78	14.2	1.67

Na última coluna assume-se que $Z \sim N(0, 1)$; $\hat{\sigma}_A\% = \sqrt{250\hat{\sigma}}100\%$



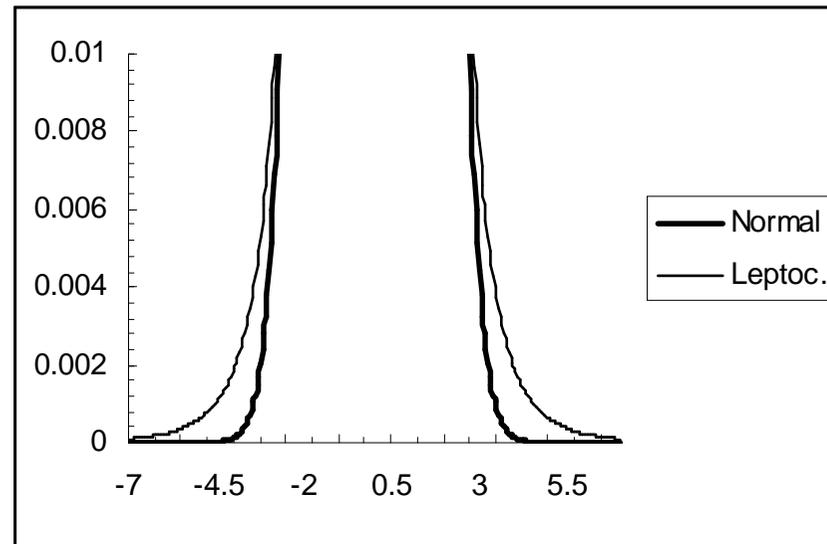
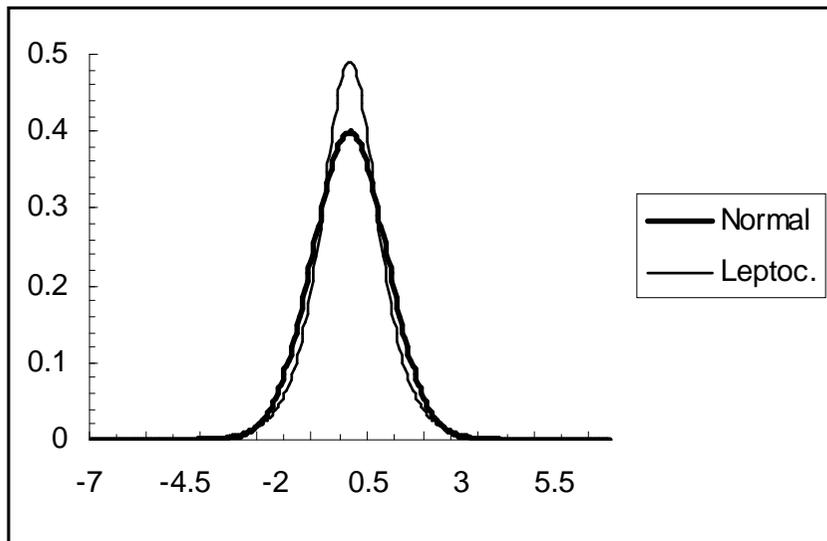
Painel Esquerdo: retornos diários do Dow Jones no período 02/10/1928 a 3/02/2011 (20678 observações). Painel direito retornos estandardizados, $z_t = (r_t - \bar{r}) / \hat{\sigma}$

Se os retornos seguissem uma distribuição normal seria de esperar que $z_t = (r_t - \bar{r}) / \hat{\sigma}$ excedesse os limiares 3 e -3 em cerca de 0.27% das vezes, tendo em conta, $P(|Z| > 3) = 0.0027$, supondo $Z \sim N(0, 1)$. Ora, na verdade z_t excede os limiares 3 e -3 em cerca de 1.73% das observações (z_t excede os limiares 359 vezes, em 20678 observações).

Tem-se assim

$$\frac{\hat{P}(|r_t - \bar{r}| > 3\hat{\sigma})}{P(|Z| > 3)} = \frac{0.017}{0.0027} = 6.29$$

Por que razão a distribuição leptocúrtica se designa habitualmente de “distribuição de caudas pesadas”?



O ensaio $H_0: k = 3$ [$y \sim Normal$ & y é *i.i.d*] pode ser conduzido pela estatística de teste

$$Z_1 = \sqrt{n} \frac{(\hat{k} - 3)}{\sqrt{24}} \xrightarrow{d} N(0, 1).$$

Por exemplo, para a Microsoft e sabendo que no período considerado se observaram 4415 dados diários ($n = 4415$) tem-se

$$z_1 = \sqrt{4415} \frac{(6.8 - 3)}{\sqrt{24}} = 51.54.$$

O valor-p é $P(|Z_1| > 51.54) \approx 0$. Existe forte evidência contra H_0 .

Podemos ainda testar a hipótese conjunta $H_0: k = 3$ & $sk = 0$ [assumindo $r \sim Normal$ & r é *i.i.d*] através da estatística de Bera-Jarque

$$Z_1^2 + Z_2^2 = n \left(\frac{(\hat{k} - 3)^2}{24} + \frac{\widehat{sk}^2}{6} \right) \xrightarrow{d} \chi^2_{(2)}$$

2.1.5 Aumento da Frequência das Observações Acentua a Não Normalidade das Distribuições

	\bar{r}	$\hat{\sigma}$	\widehat{sk}	\hat{k}
<i>Retornos Diários</i>				
Frankfurt	0.00035	0.0123	-0.946	15.0
Hong Kong	0.00057	0.0169	-5.0	119.24
Londres	0.00041	0.0092	-1.59	27.4
Nova York	0.00049	0.0099	-4.30	99.68
Paris	0.00026	0.0120	-0.53	10.56
Tóquio	0.00005	0.0136	-.213	14.798
<i>Retornos Semanais</i>				
Frankfurt	0.00169	0.0264	-1.06	8.09
Hong Kong	0.00283	0.0370	-2.19	18.25
Londres	0.00207	0.0215	-1.478	15.54
Nova York	0.00246	0.0206	-1.37	11.25
Paris	0.0028	0.0284	-.995	9.16
Tóquio	0.00025	0.0288	-.398	4.897

A diminuição da frequência das observações atenua o problema da não normalidade.

Explicação Teórica:

- P_0, P_1, P_2, \dots preços diários.

Retornos diários:

$$\{\log P_1 - \log P_0, \log P_2 - \log P_1, \dots\} = \{r_1, r_2, \dots\}$$

- P_0, P_5, P_{10}, \dots preços semanais (5 dias úteis).

Retornos semanais:

$$\left\{ \underbrace{\log P_5 - \log P_0}_{\text{retorno } 1^{\text{a}} \text{ semana}}, \underbrace{\log P_{10} - \log P_5}_{\text{retorno } 2^{\text{a}} \text{ semana}}, \dots \right\} = \{r_1 + \dots + r_5, r_6 + \dots + r_{10}, \dots\}$$

- P_0, P_h, P_{2h}, \dots Preços mensais, semestrais, etc. Retornos:

$$\log P_h - \log P_0, \log P_{2h} - \log P_h, \dots$$

Considerando, por exemplo, o primeiro retorno $\log P_h - \log P_0$, tem-se

$$\log P_h - \log P_0 = r_1 + r_2 + \dots + r_h = \sum_{i=1}^h r_i$$

(r_i são os retornos diários). Pelo TLC (sob certas condições)

$$\frac{\sum_{i=1}^h r_i - \mathbf{E} \left(\sum_{i=1}^h r_i \right)}{\sqrt{\mathbf{Var} \left(\sum_{i=1}^h r_i \right)}} = \frac{\sqrt{h} (\bar{r} - \mathbf{E}(r_i))}{\sigma} \xrightarrow{d} N(0, 1) \quad \text{quando } h \rightarrow \infty.$$

2.1.6 Efeitos de Calendário

“Efeitos de calendário”: quando a rendibilidade e/ou a volatilidade varia com o calendário.

Por exemplo, se certo título regista maior rendibilidade e/ou volatilidade às segundas-feiras, temos um efeito de calendário.

Serão estes efeitos de calendário anomalias?

Dia da Semana

“Efeito de segunda-feira”

- Rendibilidade mais alta?
- Volatilidade mais alta?

Testar este e outros efeitos:

$$r_t = \beta + \delta_1 ter_t + \delta_2 qua_t + \delta_3 qui_t + \delta_4 sex_t + u_t$$

onde *ter*, *qua*, etc. são variáveis *dummy* ... “Grupo base” = segunda-feira.

Ensaio: $H_0: \delta_1 = \delta_2 = \dots = \delta_4 = 0$ corresponde a testar a não existência de diferenças nas médias dos retornos dos vários dias da semana. A estatística habitual:

$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} \sim F(k - 1, n - k) \text{ sob } H_0.$$

Na presença de heterocedasticidade, os teste t e F são inválidos. Uma solução passa pela utilização de erros padrão robustos (ou da estatística F robusta) contra a presença de heterocedasticidade .

2.2 Regularidade Empíricas relacionadas com a Distribuição Condicional

Neste ponto discutimos regularidades que envolvem especificações dinâmicas (por exemplo, veremos como o retorno depende dos seus valores passados, ou como o quadrado dos retornos depende do quadrado dos retornos passados, entre outras especificações condicionais).

2.2.1 Autocorrelações Lineares Baixas entre os Retornos

Em geral os coeficientes de autocorrelação dos retornos são baixos. Por exemplo:.

Included observations: 5177

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.036	-0.036	6.6149	0.010
		2	-0.032	-0.034	11.993	0.002
		3	-0.043	-0.045	21.524	0.000
		4	-0.003	-0.007	21.564	0.000
		5	0.006	0.003	21.772	0.001
		6	0.001	-0.001	21.777	0.001
		7	0.004	0.004	21.878	0.003
		8	0.026	0.027	25.484	0.001
		9	-0.005	-0.002	25.601	0.002
		10	-0.014	-0.012	26.657	0.003
		11	-0.010	-0.009	27.168	0.004
		12	0.012	0.011	27.976	0.006
		13	0.016	0.015	29.370	0.006
		14	0.020	0.021	31.391	0.005
		15	-0.020	-0.016	33.404	0.004
		16	0.004	0.005	33.509	0.006
		17	0.003	0.005	33.569	0.010
		18	-0.017	-0.018	35.143	0.009
		19	-0.008	-0.009	35.506	0.012
		20	0.015	0.012	36.613	0.013

Funções de autocorrelação dos retornos diários (Microsoft 1986-2006)

Faria sentido existir correlações altas?

Teste Kendal e Stuart $H_0: \rho_k = 0$

$$\sqrt{n} (\hat{\rho}_k + 1/n) \xrightarrow{d} N(0, 1), \quad \hat{\rho}_k \stackrel{a}{\sim} N\left(-\frac{1}{n}, \frac{1}{\sqrt{n}}\right)$$

Teste Ljung-Box $H_0: \rho_1 = \dots = \rho_m = 0$

$$Q_m = n(n+2) \sum_{k=1}^m \frac{1}{n-k} \hat{\rho}_k^2 \xrightarrow{d} \chi_{(m)}^2$$

Analise-se $H_0: \rho_1 = \dots = \rho_{20} = 0$. Fixe-se por exemplo, $m = 20$. Tem-se $Q_{20} = 36.613$.
valor-p= 0.013. Existe evidência estatística contra a hipótese nula $H_0: \rho_1 = \dots = \rho_{20} = 0$.
Esta conclusão parece contraditória com a ideia de baixas autocorrelações dos retornos.

Há explicação?

Duas formas de mitigar a presença de heterocedasticidade:

1. Considerar erros padrão robustos.

2. Estandardizar os retornos:

$$r_t^* = \frac{r_t - \bar{r}}{\hat{\sigma}_t}$$

onde $\hat{\sigma}_t$ é uma estimativa do desvio padrão condicional de r_t (r_t^* pode ser encarado como os “retornos” centrados expurgados da heterocedasticidade).

Como obter $\hat{\sigma}_t$? Solução (“subóptima”):

$$\hat{\sigma}_t^2 = (1 - \lambda) r_{t-1}^2 + \lambda \hat{\sigma}_{t-1}^2, \quad \lambda = 0.96$$

Esta solução apenas se justifica para séries temporais financeiras observadas com frequência elevada (por exemplo, freq. diária).

Coeficientes de autocorrelação de r_t^* :

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.017	-0.017	1.4495	0.229
		2	-0.009	-0.010	1.8905	0.389
		3	-0.007	-0.007	2.1546	0.541
		4	-0.005	-0.006	2.3040	0.680
		5	-0.009	-0.010	2.7699	0.735
		6	-0.013	-0.013	3.6259	0.727
		7	0.000	-0.001	3.6259	0.822
		8	0.009	0.008	4.0296	0.854
		9	0.014	0.014	5.0949	0.826
		10	0.007	0.007	5.3206	0.869
		11	-0.008	-0.008	5.6878	0.893
		12	0.012	0.012	6.3912	0.895
		13	0.024	0.025	9.3486	0.746
		14	0.013	0.015	10.241	0.744
		15	-0.005	-0.004	10.396	0.794
		16	0.005	0.006	10.533	0.837
		17	0.009	0.010	10.967	0.858
		18	-0.015	-0.014	12.205	0.836
		19	-0.007	-0.006	12.446	0.866
		20	0.010	0.010	12.922	0.881

Funções de autocorrelação dos retornos diários estandardizados (Microsoft 1986-2006)

Para apreciarmos a diferença com uma série macroeconómica, veja-se:

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.461	0.461	48.583	0.000
		2	0.305	0.118	70.041	0.000
		3	0.069	-0.141	71.137	0.000
		4	-0.030	-0.062	71.340	0.000
		5	-0.162	-0.131	77.479	0.000
		6	-0.063	0.103	78.415	0.000
		7	0.023	0.112	78.541	0.000
		8	0.039	-0.036	78.909	0.000
		9	0.178	0.161	86.434	0.000
		10	0.212	0.078	97.115	0.000
		11	0.177	-0.000	104.62	0.000
		12	0.038	-0.092	104.97	0.000
		13	0.026	0.023	105.13	0.000
		14	0.038	0.136	105.47	0.000
		15	0.039	0.042	105.85	0.000

Funções de autocorrelação de $y_t = \log(GNP_t/GNP_{t-1})$ onde GNP é o PIB dos EUA (dados trimestrais de 1947 a 2003).

2.2.2 Volatility Clustering

Vimos: valores muito altos e muito baixos ocorrem frequentemente. Estes valores extremos não ocorrem isoladamente: tendem a ocorrer de forma seguida (*volatility clustering*).

Uma das propriedades mais importantes: fortes (baixas) variações são normalmente seguidas de fortes (baixas) variações em ambos os sentidos.

Este facto pode resultar do seguinte processo:

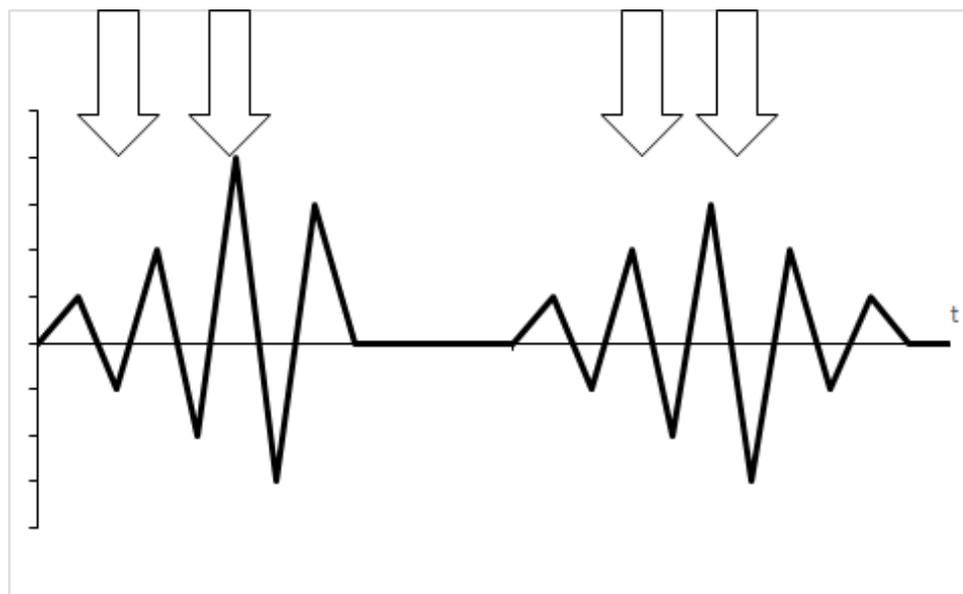
Equilíbrio (não há flutuação dos preços)

Chegada de informação ao mercado →

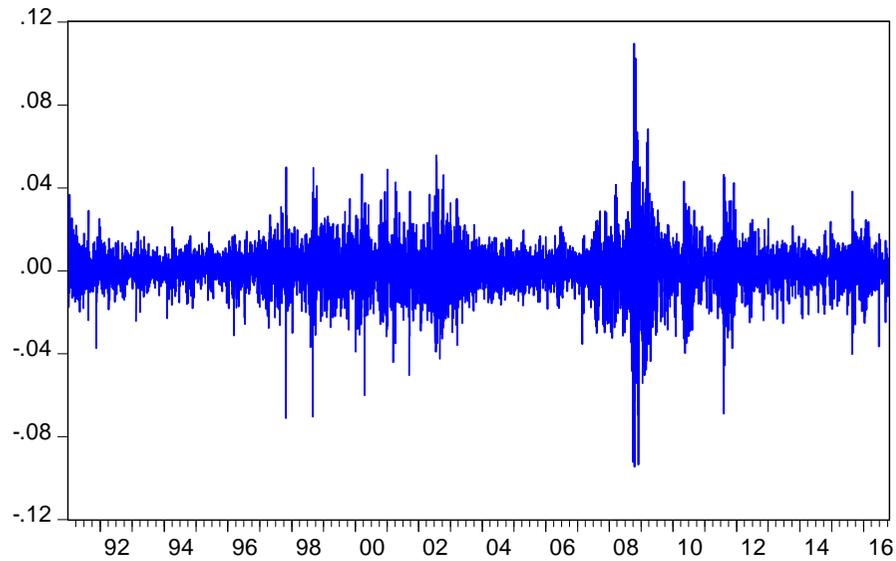
Agentes reavaliam as suas carteiras →

Aumento Vol. Transacções →

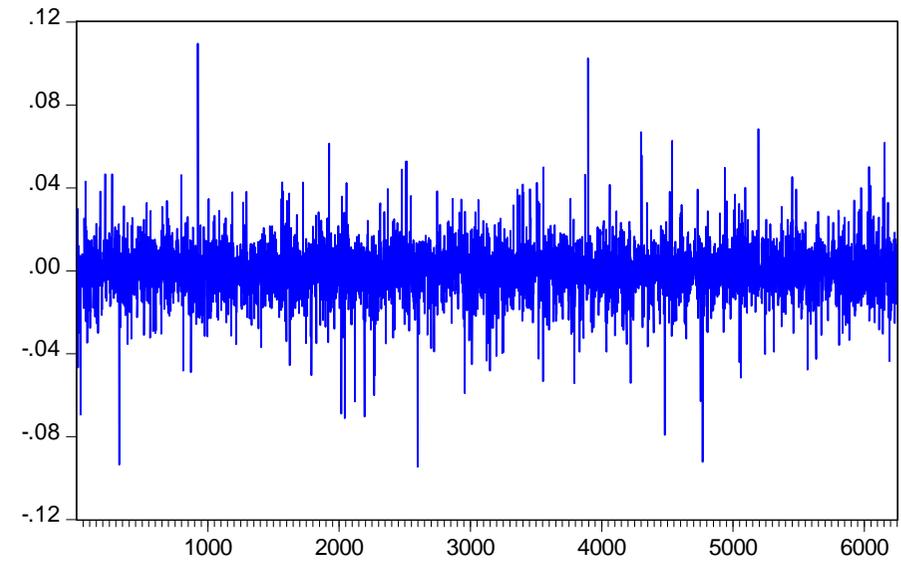
Flutuação dos preços (volatilidade)



Retornos diários do SP500



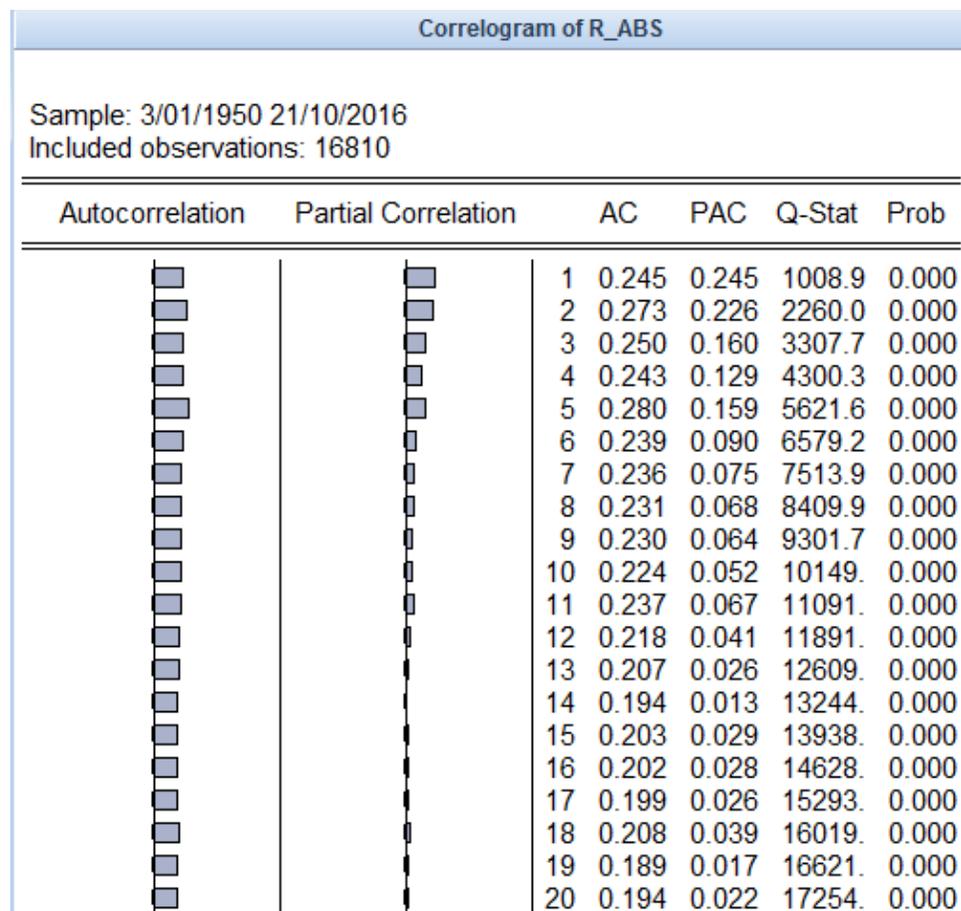
Retornos diários do SP500 dispostos por ordem aleatória



As séries possuem exatamente os mesmos momentos amostrais (média, variância, skewness, kurtosis, etc.)

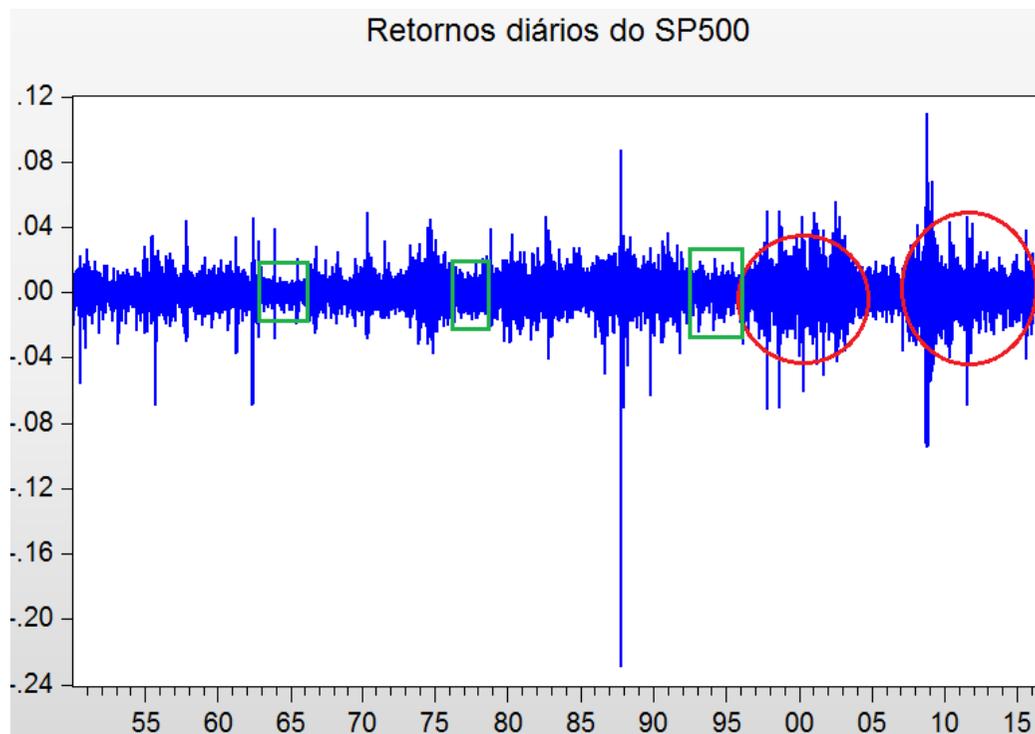
Se fortes (baixas) variações são normalmente seguidas de fortes (baixas) variações em ambos os sentidos, então a magnitude de r_t deve estar correlacionado com a magnitude de r_{t-i} .

Autocorrelação de $|r_t|$ (SP500):

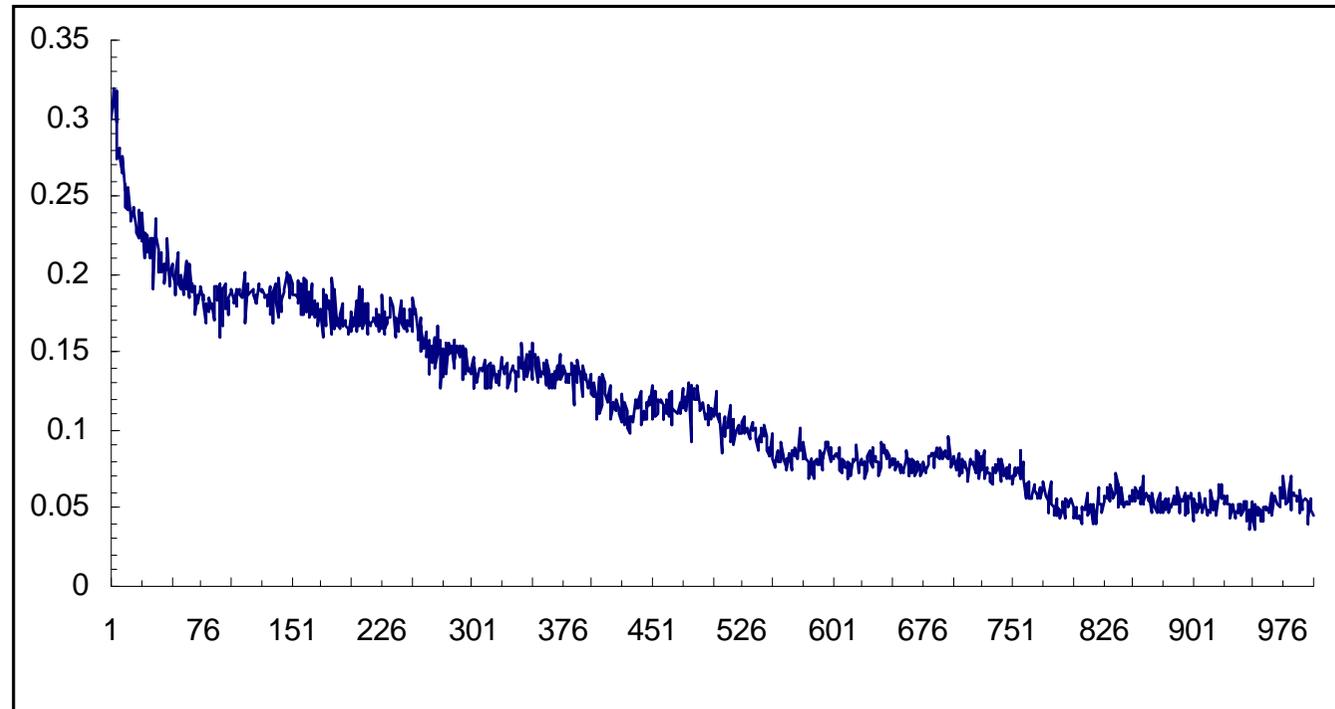


2.2.3 Forte Dependência Temporal da Volatilidade

Vimos: (1) valores muitos altos e muito baixos ocorrem frequentemente e (2) estes valores extremos aparecem de forma seguida (*volatility clustering*). Neste ponto reforça-se a ideia de *volatility clustering*: se a volatilidade é alta (baixa), então é razoável esperar que a volatilidade se mantenha alta (baixa) durante bastante tempo. Períodos de grande e baixa volatilidade prolongam-se por vários anos!

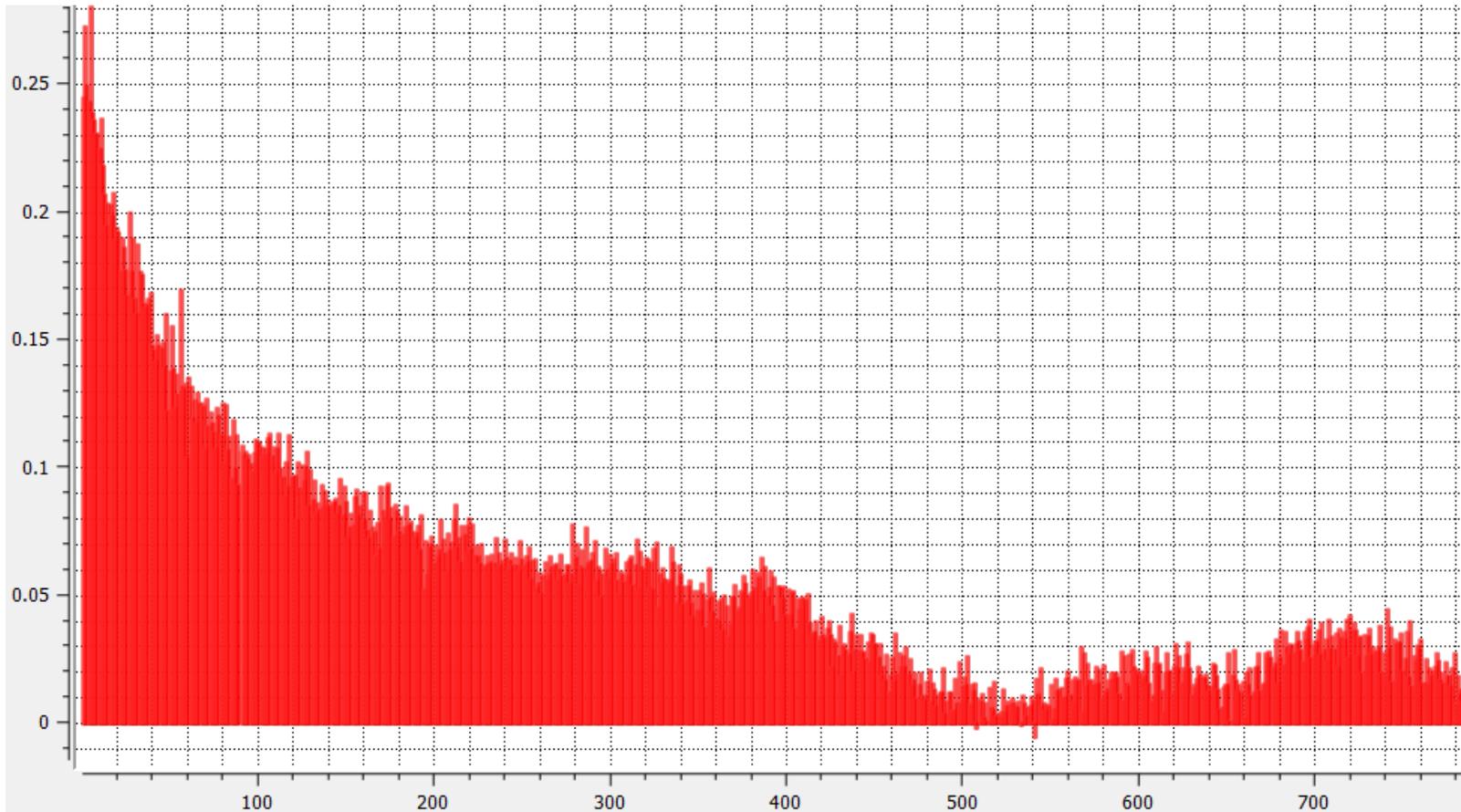


Para confirmarmos a ideia de forte dependência temporal da volatilidade calcule-se a FAC associado aos valores absolutos dos retornos do Dow Jones:



FAC de $|r_t|$ onde r_t é o retorno diário do Dow Jones (Jan. 1928 a Fev. 2006)

Outro exemplo:



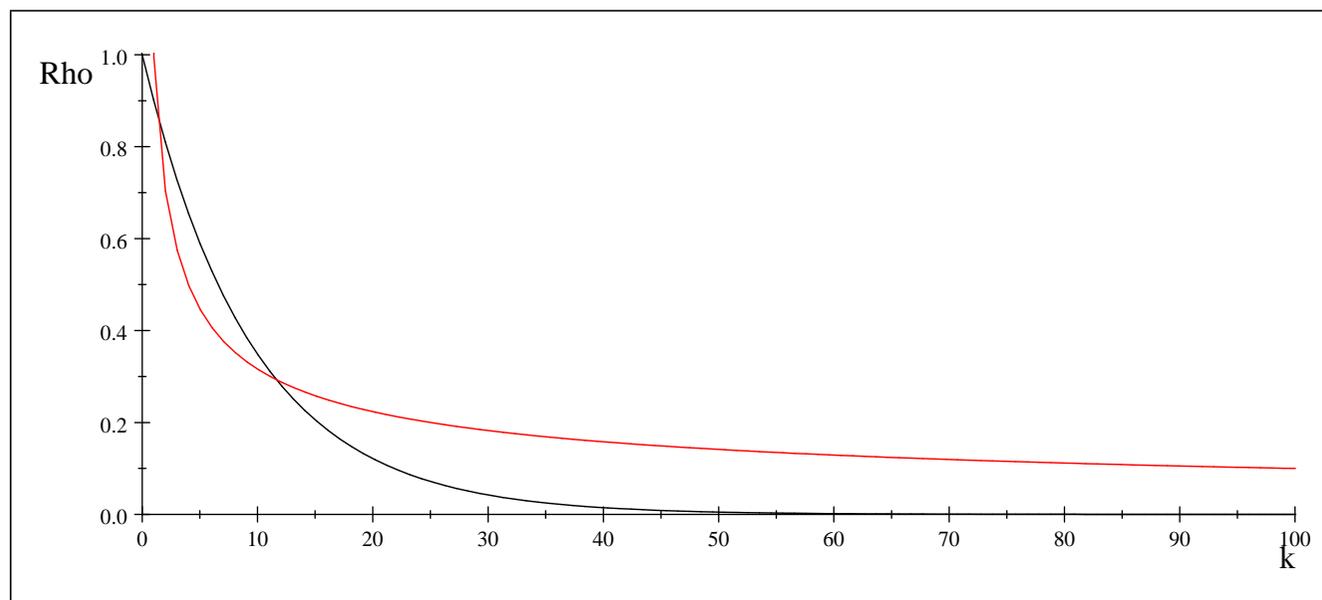
FAC de $|r_t|$ onde r_t é o retorno diário do SP500 (Jan. 1950 a Out 2016)

- Processos ARMA \Rightarrow decaimento exponencial da FAC:

$$|\rho_k| = C a^k, \quad 0 < C < \infty, \quad 0 < a < 1.$$

- Processos de memória longa \Rightarrow decaimento polinomial da FAC:

$$|\rho_k| = C |k|^{-\beta}, \quad \beta > 0$$



Decaimento Exponencial (0.9^k) vs. Polinomial ($k^{-0.5}$)

Discutir o decaimento das FAC dos retornos absolutos do Dow Jones e do SP500

2.2.4 Efeito Assimétrico

Correlação entre a volatilidade e a ocorrência de perdas significativas nos mercados de capitais.

<i>Índices Bolsistas</i>	$\widehat{\text{Corr}}(r_{t-1}, r_t^2)$
Amesterdão	-0.049
Frankfurt	-0.095
Hong Kong	-0.081
Nova York	-0.199
<i>Taxas de Câmbio</i>	
Libra Britânica	0.074
Dólar Canadano	0.041
Yen	-0.008
Franco Suíço	0.014

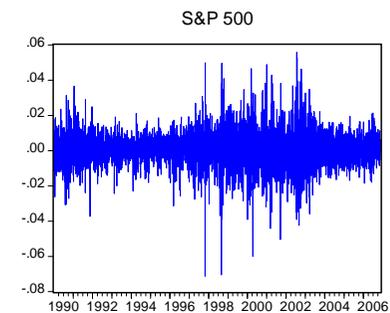
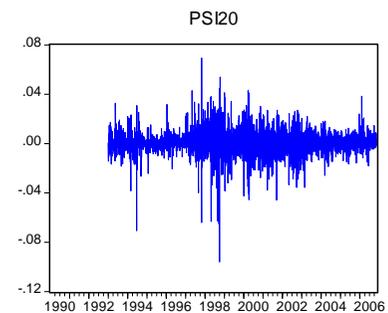
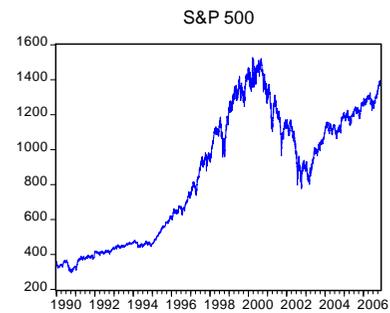
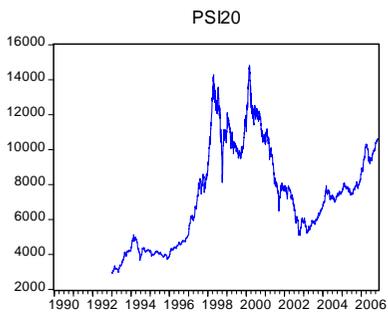
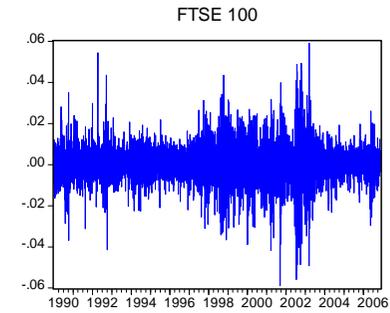
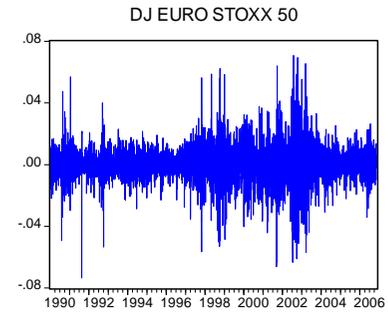
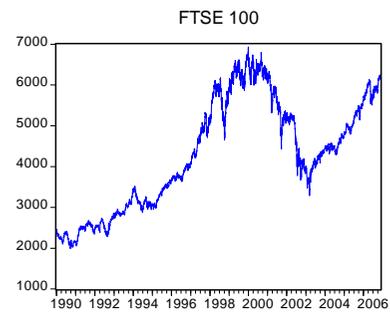
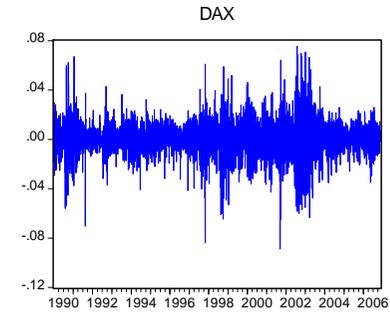
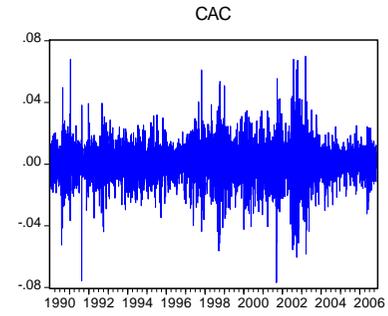
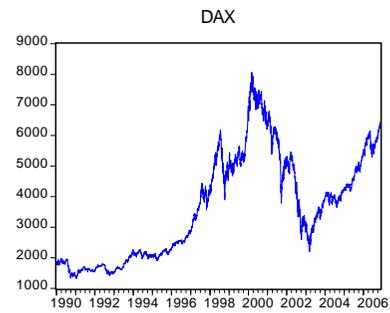
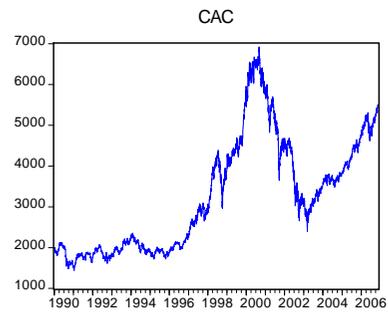
Como verificar se as estimativas são estatisticamente significativas?

2.2.5 Aumento da Frequência das Observações Acentua a Não Linearidade

Vários estudos indicam que os coeficientes de autocorrelações de r_t^2 e de $|r_t|$ tendem a aumentar com o aumento da frequência das observações.

2.2.6 Co-Movimentos de Rendibilidade e Volatilidade

Ao se analisarem duas ou mais séries financeiras de retornos ao longo do tempo, geralmente observam-se co-movimentos de rendibilidade e volatilidade, isto é, quando a rendibilidade e a volatilidade de uma série aumenta (diminui), a rendibilidade e a volatilidade das outras tende, em geral, a aumentar (diminuir).



Índices Bolsistas

Retornos de Índices

	CAC	DAX	DJ EURO 50	FTSE 100	PSI 20	S&P 500
CAC	1					
DAX	0.78	1				
DJ EURO 50	0.92	0.89	1			
FTSE 100	0.79	0.70	0.81	1		
PSI 20	0.52	0.50	0.57	0.47	1	
S&P 500	0.43	0.47	0.46	0.41	0.25	1

Matriz de correlações dos retornos diários (Jan 90-Nov 06)

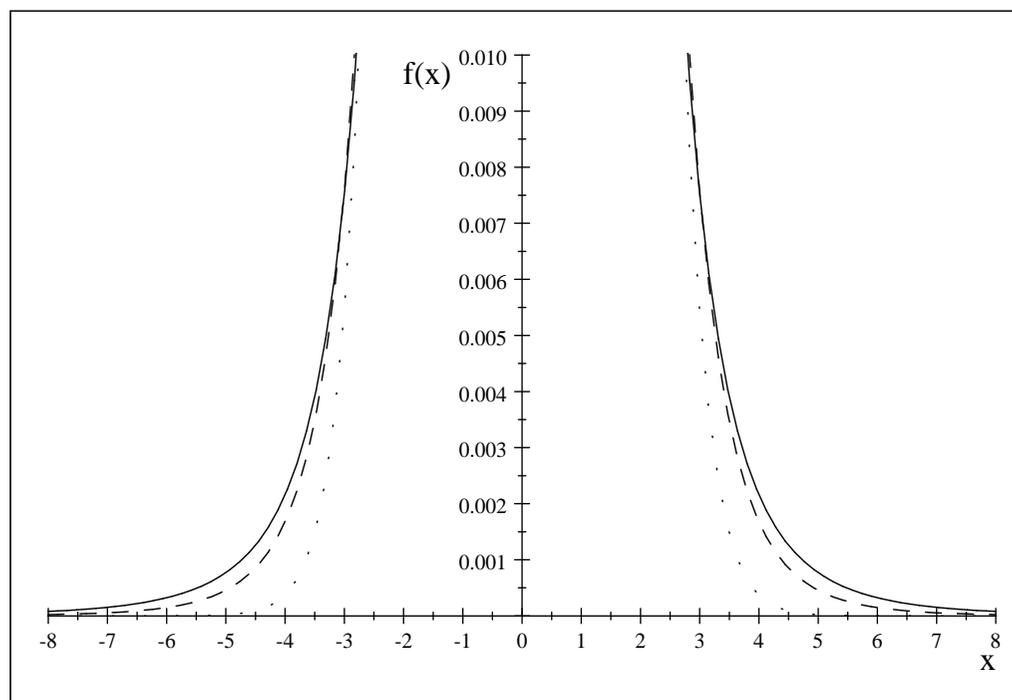
	CAC	DAX	DJ EURO 50	FTSE 100	PSI 20	S&P 500
CAC	1					
DAX	0.74	1				
DJ EURO 50	0.91	0.85	1			
FTSE 100	0.77	0.66	0.78	1		
PSI 20	0.39	0.39	0.43	0.30	1	
S&P 500	0.38	0.44	0.41	0.35	0.21	1

Matriz de correlações dos retornos diários ao quadrado (Jan 90-Nov 06)

2.3 Distribuições Leptocúrticas para Retornos

2.3.1 Distribuição t-Student

kurtosis $k = 3 + \frac{6}{v-4}$. Quando mais baixo for o número de graus de liberdade mais pesadas são as caudas. No caso $v = 4$ a kurtosis não existe.



Abas da t -Student ($- v = 4.1, - - v = 5, \dots v = 30$)

2.3.2 Mistura de Normais

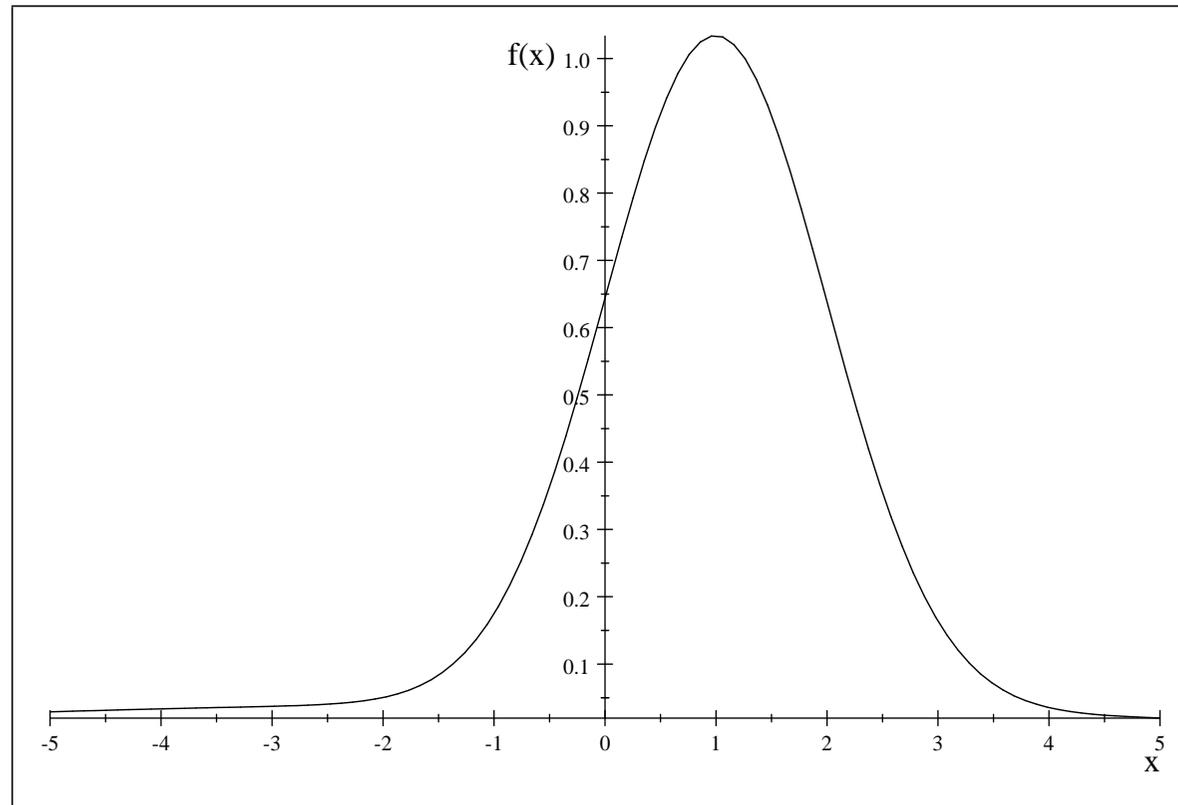
A variável X , que representa a mistura de normais, pode escrever-se da seguinte forma:

$$X = (1 - U) X_1 + U X_2,$$

$U \sim \text{Bernoulli}$, $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$. Pode-se mostrar que,

- $E(X) = \alpha\mu_1 + (1 - \alpha)\mu_2$;
- $\text{Var}(X) = \alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 + \alpha(1 - \alpha)(\mu_1 - \mu_2)^2$;
- $E\left((X - E(X))^3\right) = \alpha(1 - \alpha)(\mu_1 - \mu_2)\left((1 - 2\alpha)(\mu_1 - \mu_2)^2 + 3(\sigma_1^2 + \sigma_2^2)\right)$;
- $k = 3 + \frac{3\alpha(1-\alpha)(\sigma_1^2 - \sigma_2^2)^2}{(\alpha\sigma_1^2 + (1-\alpha)\sigma_2^2)^2} > 3$ supondo, para simplificar, que $\mu_1 = \mu_2 = 0$.

Representa-se $f(x)$ para $\alpha = 0.5$, $\mu_1 = -1$, $\mu_2 = 1$, $\sigma_1 = 1/5$, $\sigma_2 = 5$.



Mistura de Normais

Ver ficheiro *mistura normais.nb*

2.3.3 Distribuição com Caudas de Pareto

Seja $y \sim \text{Pareto}(c, \alpha)$. Então

$$(fdp) \ g(y) = \frac{\alpha c^\alpha}{y^{1+\alpha}}, \quad (fd) \ G(y) = 1 - \left(\frac{c}{y}\right)^\alpha \quad y > c.$$

Primeiros 4 momentos:

$$\mu = \int_c^\infty yg(y) dy = \frac{c\alpha}{\alpha - 1}, \quad \alpha > 1$$

$$\sigma^2 = \int_c^\infty (y - \mu)^2 g(y) dy = \frac{c^2\alpha}{(\alpha - 1)^2(\alpha - 2)}, \quad \alpha > 2$$

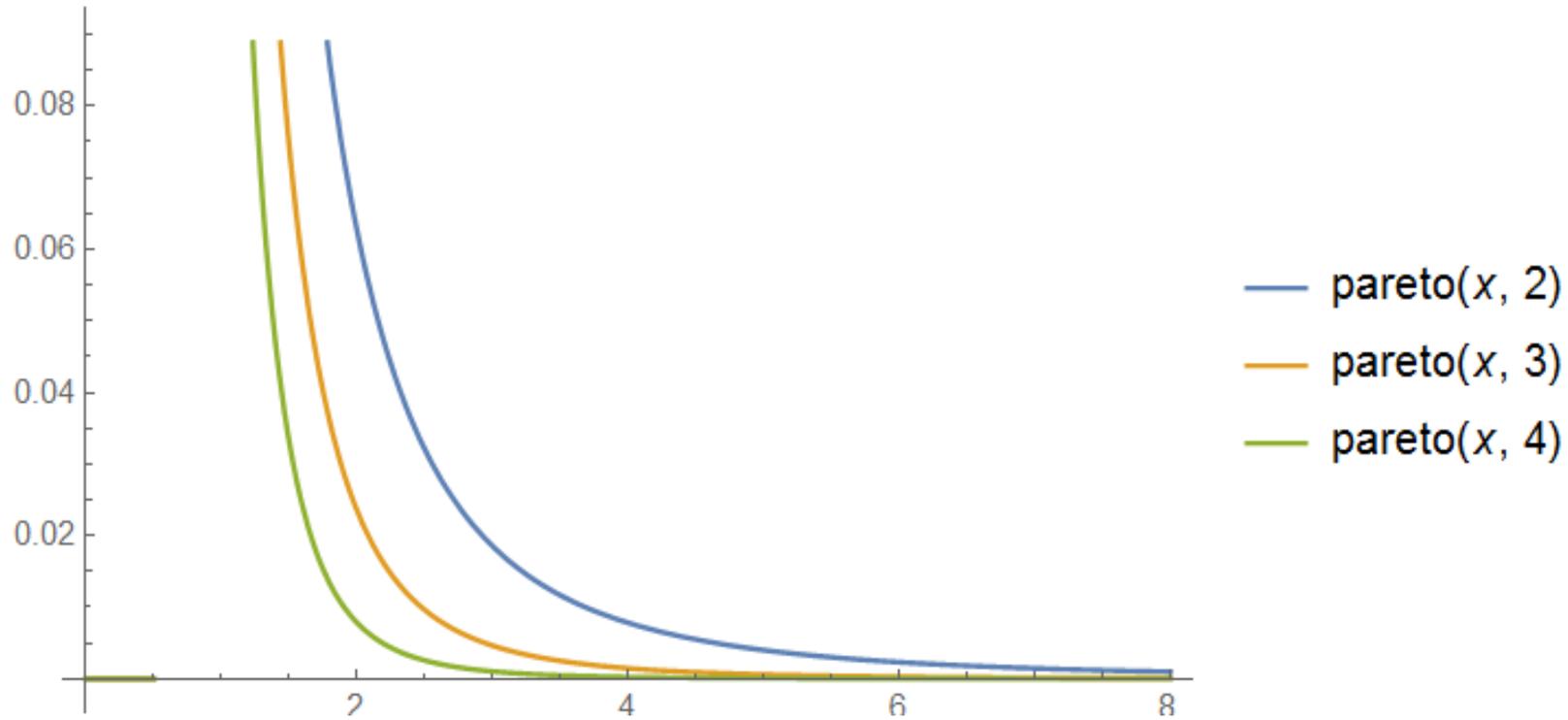
$$sk = \frac{\int_c^\infty (y - \mu)^3 g(y) dy}{\sigma^3} = \frac{2(1 + \alpha)}{\alpha - 3} \sqrt{\frac{\alpha - 2}{\alpha}}, \quad \alpha > 3$$

$$kurt = \frac{\int_c^\infty (y - \mu)^4 g(y) dy}{\sigma^4} = 3 + \frac{6(\alpha^3 + \alpha^2 - 6\alpha - 2)}{\alpha(\alpha - 3)(\alpha - 4)}, \quad \alpha > 4$$

Nota:

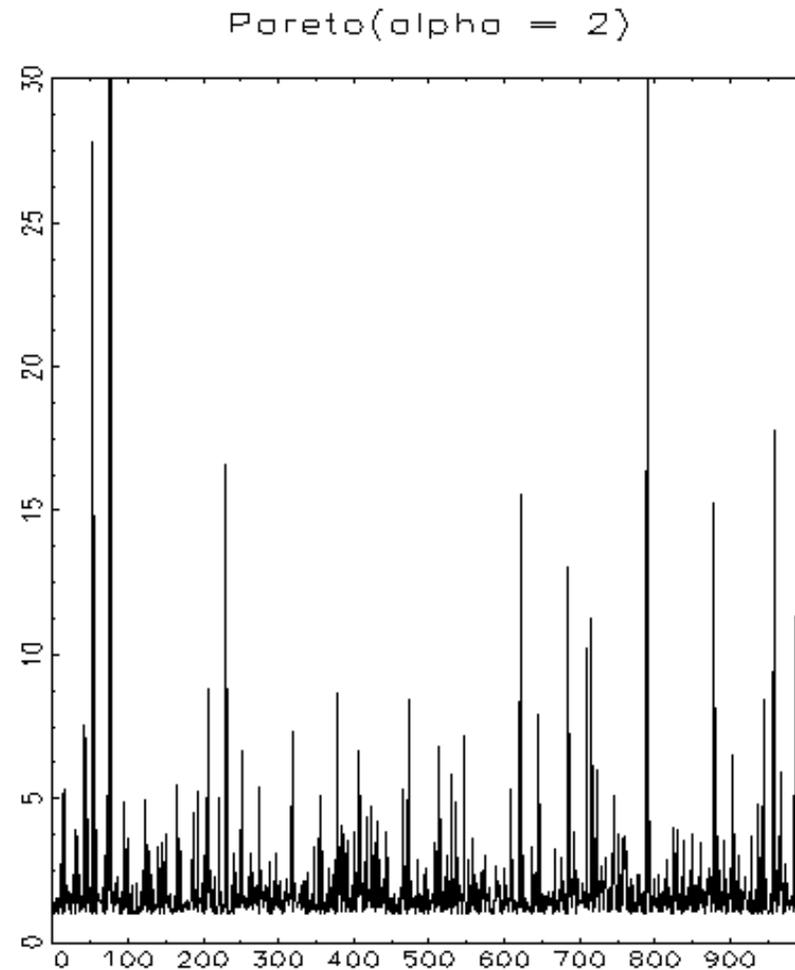
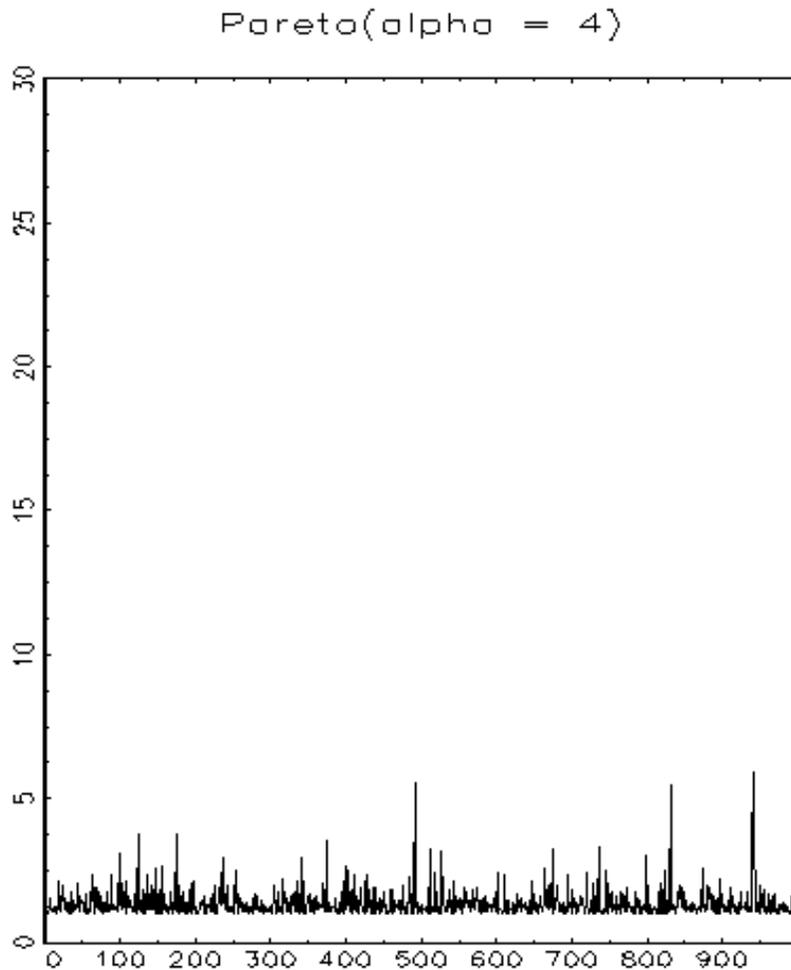
$$\mathbf{E}(|y|^k) = \int_c^\infty |y|^k \frac{\alpha c^\alpha}{y^{1+\alpha}} dy = \alpha c^\alpha \int_c^\infty y^{k-1-\alpha} dy = \begin{cases} \frac{\alpha c^k}{\alpha - k} & \alpha > k \\ \infty & \alpha \leq k \end{cases}$$

```
pareto[x_,  $\alpha$ _] = PDF[ParetoDistribution[1/2,  $\alpha$ ], x];  
Plot[{pareto[x, 2], pareto[x, 3], pareto[x, 4]}, {x, 0, 8}, PlotLegends  $\rightarrow$  "Expressions"]
```

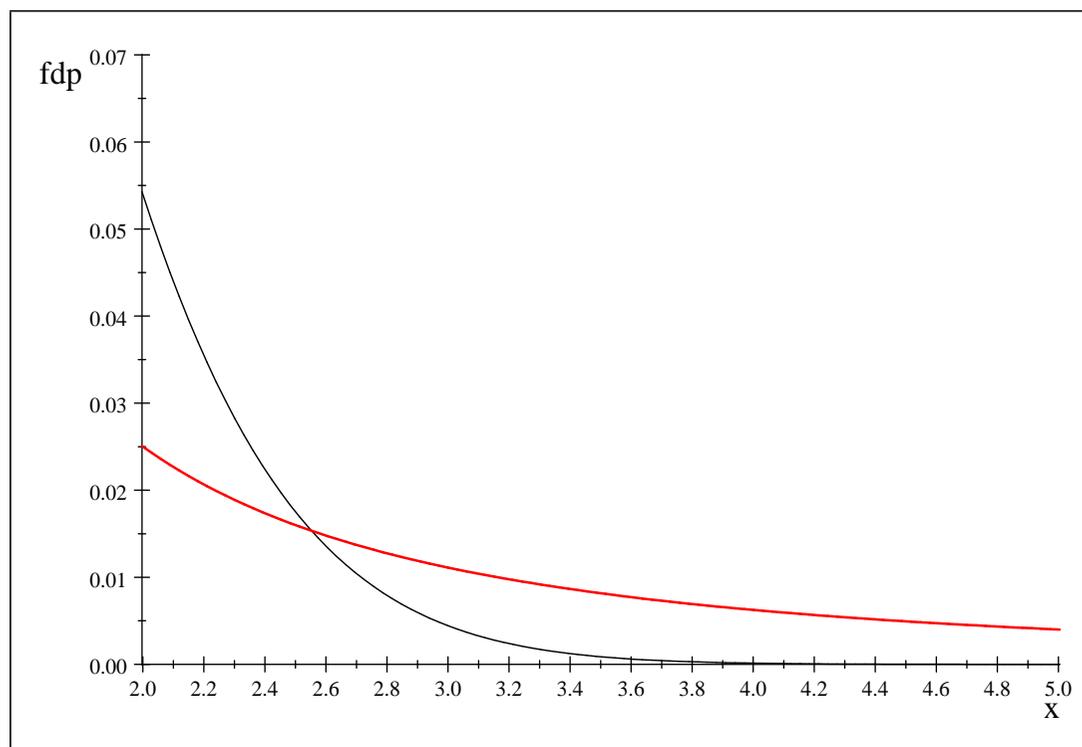


Exemplo Para simular $y \sim \text{Pareto}(c, \alpha)$ considerar o resultado $u \sim U(0, 1) \Rightarrow y = G^{-1}(u) \sim \text{Pareto}(c, \alpha)$. Dada uma sequência i.i.d. de v.a. $\{u_1, \dots, u_n\}$ tem-se:

$$y_i = c(1 - u_i)^{-\frac{1}{\alpha}} \sim \text{Pareto}(c, \alpha)$$



Uma distribuição com caudas de Pareto pode ser interessante para caracterizar o comportamento probabilístico dos retornos porque as caudas destas distribuições apresentam um decaimento polinomial para zero (decaimento lento para zero) e, portanto, caudas pesadas.



Cauda de Pareto vs. Cauda Gaussiana

Vimos que a fdp de Pareto é

$$g(y) = \alpha c^\alpha y^{-(\alpha+1)}, \quad y > c.$$

À primeira vista pode parecer que esta distribuição não serve ... Porquê?

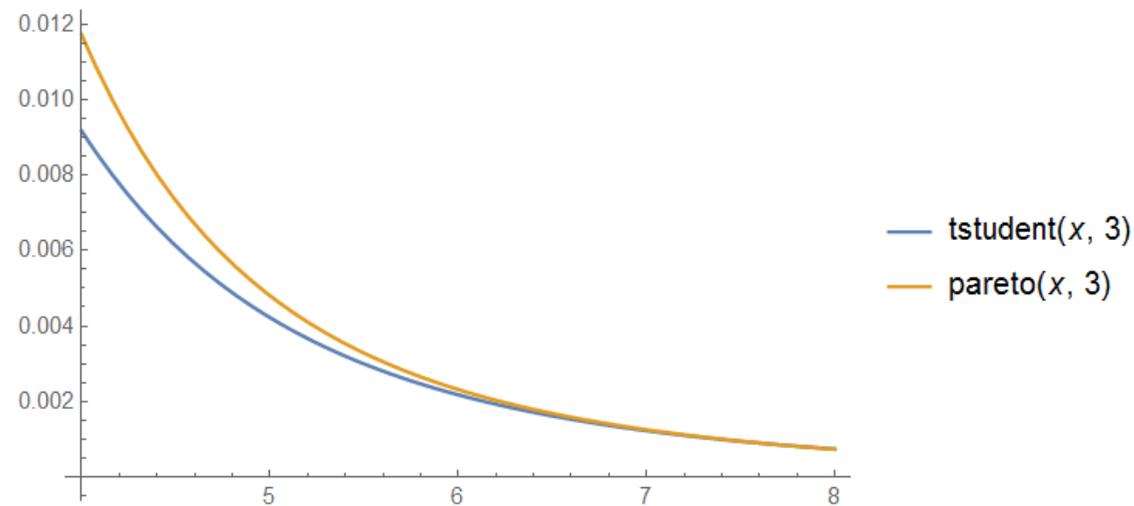
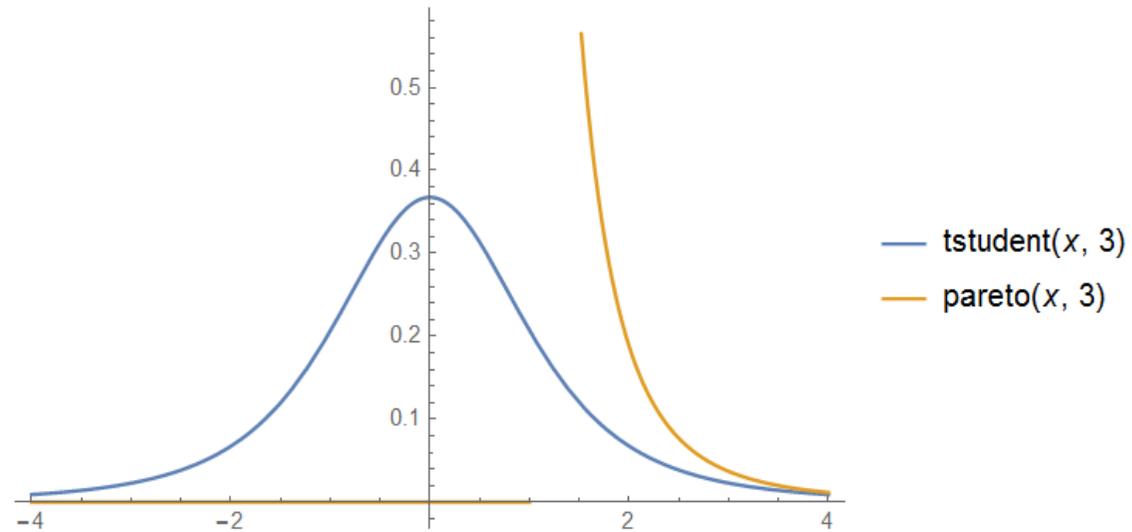
Diz-se que uma fdp $f(y)$ tem distribuição com caudas de Pareto (mesmo que não seja uma distribuição de Pareto) se

$$f(y) \sim Cy^{-(\alpha+1)}, \quad \alpha > 0$$

(C é uma constante). O sinal “ \sim ” significa aqui que

$$\lim_{y \rightarrow \infty} f(y) / Cy^{-(\alpha+1)} = 1.$$

Pode-se provar que a distribuição t-student (entre muitas outras) tem caudas de Pareto:



Estimador de Hill (aba direita)

Suponha-se $f(y) \sim Cy^{-(\alpha+1)}$. O parâmetro α é muito importante! Como estimar α ?

Passo 1: admita-se um cenário mais simples: $f(y)$ tem distribuição (exacta) de Pareto, $y \sim \text{Pareto}(c, \alpha)$, ou seja

$$f(y) = \frac{\alpha c^\alpha}{y^{\alpha+1}}, \quad y > c.$$

Resulta que o estimador de MV é

$$\hat{\alpha} = \frac{n}{\sum_{t=1}^n \log(y_t/c)}, \quad \sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{d} N(0, \alpha^2)$$

Passo 2: retome-se agora a hipótese $f(y) \sim Cy^{-(\alpha+1)}$. Como $f(y) \simeq Cy^{-(\alpha+1)}$ apenas se verifica para os valores grandes de y , então devemos considerar apenas os dados y_t tais que $y_t > q$ (onde q pode ser interpretado como um quantil de y , geralmente um quantil ordem igual ou superior a 0.95).

O estimador para α (designado por estimador de Hill) é agora

$$\hat{\alpha}(q) = \frac{n(q)}{\sum_{t=1}^n \log(y_t/q) \mathcal{I}_{\{y_t > q\}}}, \quad n(q) = \sum_{t=1}^n \mathcal{I}_{\{y_t > q\}}.$$

Pode-se mostrar

$$\sqrt{n(q)} (\hat{\alpha}(q) - \alpha(q)) \xrightarrow{d} N(0, \alpha^2)$$

quando $n \rightarrow \infty$, $n(q) \rightarrow \infty$ e $n(q)/n \rightarrow 0$. Observe-se

$$\text{Var}(\hat{\alpha}(q)) = \frac{\alpha^2}{n(q)}, \quad \widehat{\text{Var}}(\hat{\alpha}(q)) = \frac{\hat{\alpha}^2}{n(q)}.$$

Estimador de Hill (aba esquerda)

Considerar agora q um quantil de ordem igual ou inferior a 0.05 e

$$\hat{\alpha}(q) = \frac{n(q)}{\sum_{t=1}^n \log(y_t/q) \mathcal{I}_{\{y_t < q\}}}, \quad n(q) = \sum_{t=1}^n \mathcal{I}_{\{y_t < q\}}.$$

Exemplo (Aba direita) Resulta do quadro seguinte que $\hat{\alpha}(0.01) = 3/5.193 = 0.577$

y_t	$\mathcal{I}_{\{y_t > 0.01\}}$	$\log(y_t/0.01) \mathcal{I}_{\{y_t > 0.01\}}$
-0.110	0	0
0.090	1	2.197
0.100	1	2.303
-0.100	0	0
0.020	1	0.693
0.005	0	0
Σ	3	5.193

Ver ficheiros *Estimador de Hill.xlsx* e *hill_02.prg*

Qual é o valor do *threshold* q que devemos escolher? Temos um dilema de enviesamento *versus* variância:

- se q é alto a estimação de $\hat{\alpha}(q)$ é baseada em poucas observações, i.e., $n(q)$ é baixo, pelo que a variância de $\hat{\alpha}(q)$ é alta (observe-se $\text{Var}(\hat{\alpha}(q)) = \alpha^2/n(q)$);
- se q é baixo, perde-se a hipótese $f(y) \sim Cy^{-(\alpha+1)}$ e, como consequência, o estimador $\hat{\alpha}(q)$ é enviesado e mesmo inconsistente (recorde-se que $\hat{\alpha}(q)$ é baseado na hipótese $f(y) \sim Cy^{-(\alpha+1)}$).

Exemplo Estimação do índice de cauda (dentro de parênteses encontram-se o número de observações efectivamente utilizadas para estimar α).

Retornos	$\hat{\alpha} (q_{0.96})$	$\hat{\alpha} (q_{0.97})$	$\hat{\alpha} (q_{0.98})$	$\hat{\alpha} (q_{0.99})$
ORACLE	3.0 (153)	3.23 (115)	3.3 (73)	4.17 (39)
DOWJONES	2.67 (1114)	2.74 (835)	2.85 (557)	3.0 (279)
EURO/DÓLAR	4.70 (65)	5.12 (49)	5.33 (33)	5.14 (17)

- Em todos os casos é razoável admitir-se que a variância existe. Porquê?
- A distribuição dos retornos do DOWJONES parece ter as caudas mais pesadas; o quarto momento pode não existir.

2.4 Estimação Não Paramétrica da Função Densidade de Probabilidade

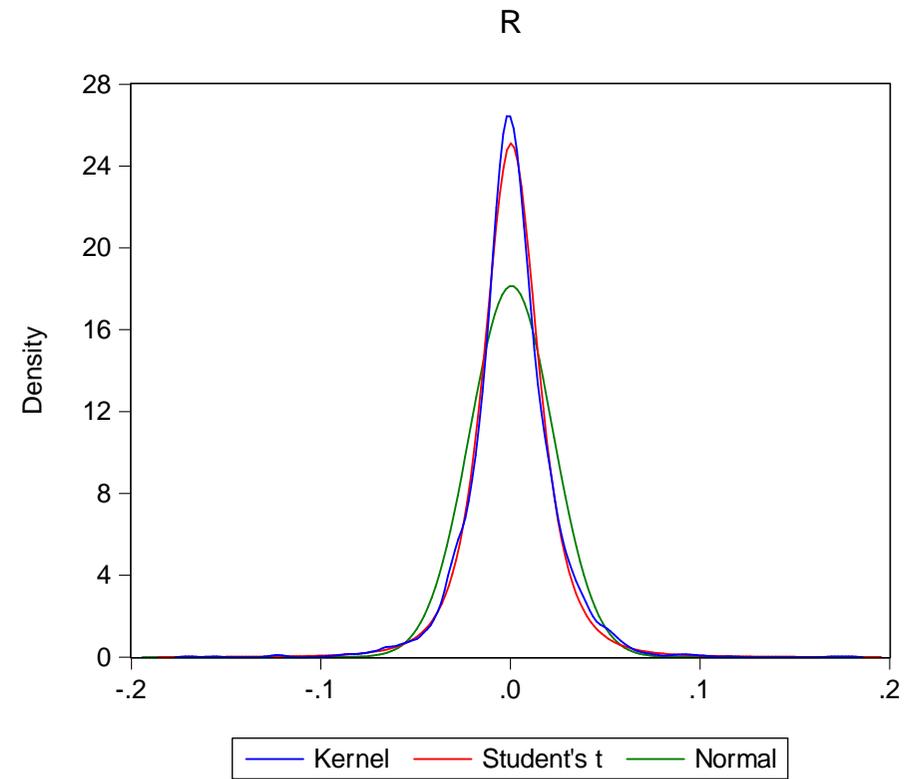
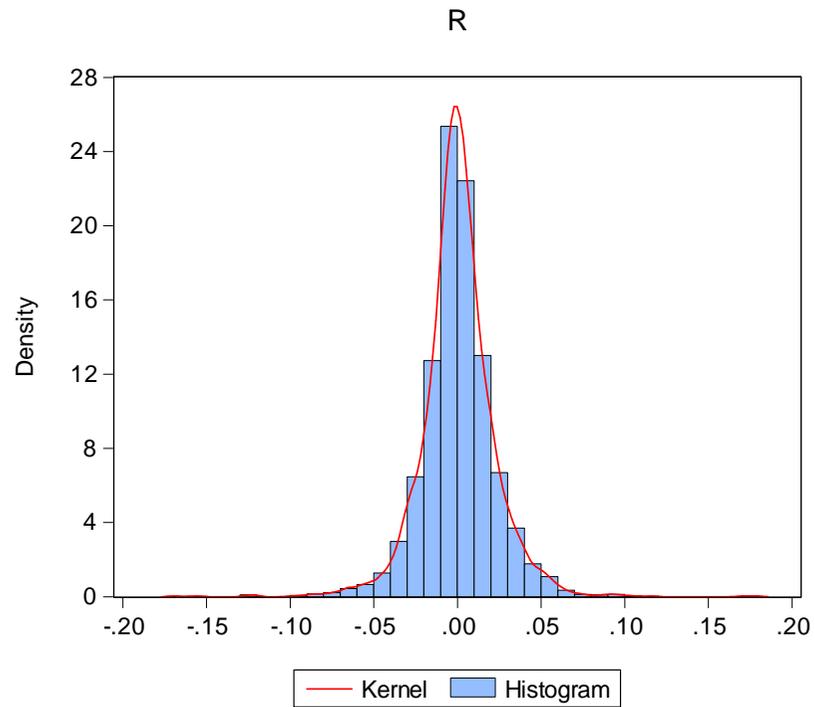
A forma mais simples de estimar $f(x)$ consiste em obter o histograma das frequências relativas. Existem, no entanto, estimadores preferíveis. Uma estimativa não paramétrica de $f(x)$ pode ser dada por

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

onde $K(u)$ é uma fdp. Sob certas condições, incluindo $h \rightarrow 0, n \rightarrow \infty, nh \rightarrow \infty$ pode-se provar

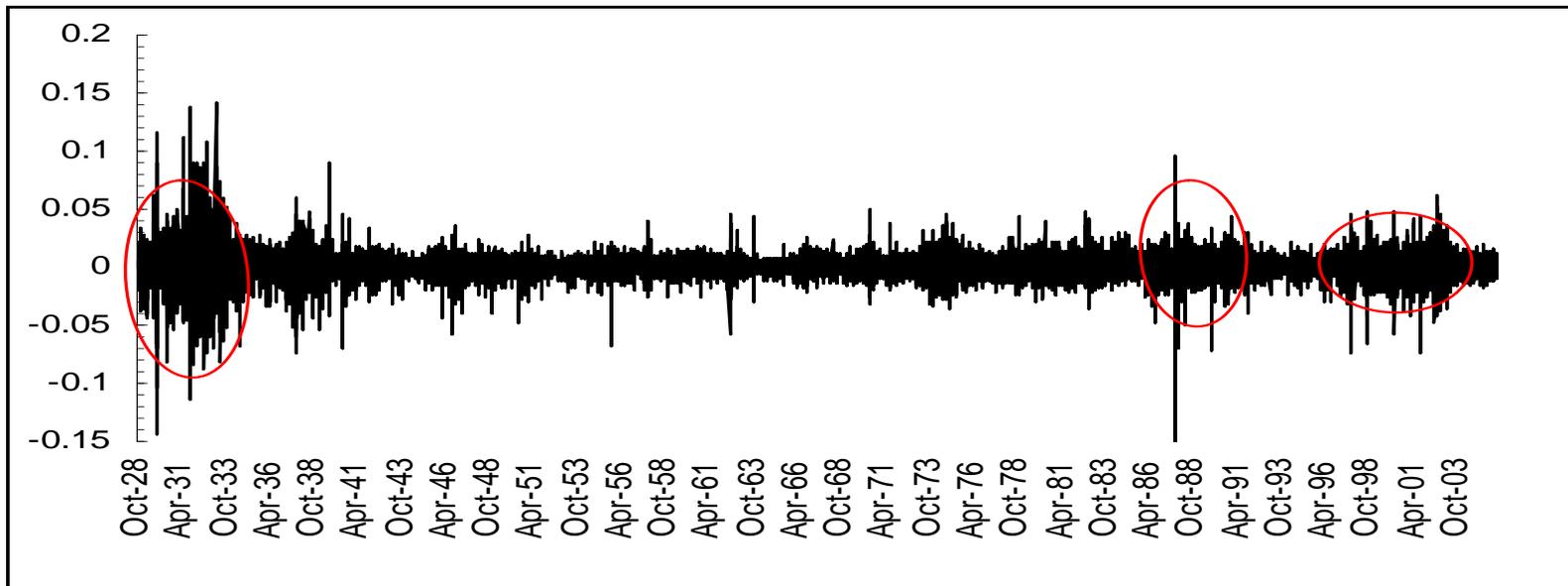
$$\hat{f}(x) \xrightarrow{p} f(x).$$

Exemplo Retornos Microsoft 3/1/1995 - 4/2/2010



3 Modelação da Heterocedasticidade Condicionada: Caso Univariado

3.1 Introdução



Por que razão a volatilidade não é constante?

- Uma parte da volatilidade pode ser relacionada com a especulação.
- Episódios de extrema volatilidade ocorrem quando uma “bolha especulativa” rebenta.
- Graves crises económicas e políticas também explicam momentos de alta volatilidade.
- Conceito "chegada de informação".

Replicar o fenómeno de volatilidade não constante a partir do conceito de chegada de informação.

Seja N_t o número de notícias no dia t .

$N_t = 1 \Rightarrow$ variação do preço numa quantidade aleatória dada por $\varepsilon_{1,t}$

$$\log P_t = \log P_{t-1} + \mu + \varepsilon_{1,t}$$

$N_t = 2 \Rightarrow$ variação do preço numa quantidade aleatória dada por $\varepsilon_{1,t} + \varepsilon_{2,t}$

$$\log P_t = \log P_{t-1} + \mu + \varepsilon_{1,t} + \varepsilon_{2,t}$$

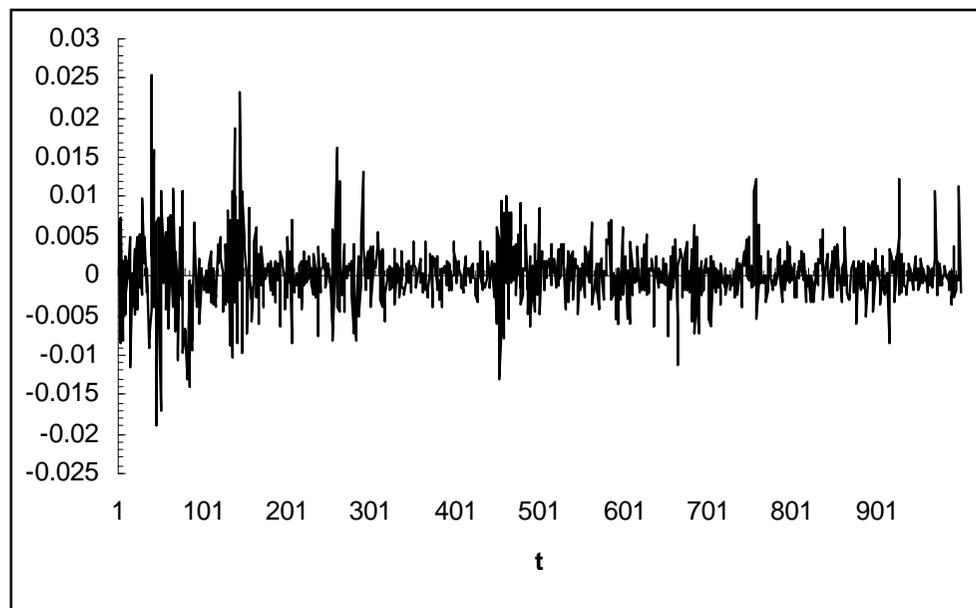
Retorno do activo:

$$r_t = \mu + \sum_{i=1}^{N_t} \varepsilon_{i,t}.$$

Resulta do modelo que a variância de r_t dado N_t é não constante, pois

$$\text{Var} [r_t | N_t] = N_t \sigma^2$$

(quanto maior é o número de notícias que chegam ao mercado, maior é a volatilidade).



O modelo anterior sugere a especificação

$$r_t = \mu + u_t, \quad \text{com } u_t = \sigma_t \varepsilon_t.$$

u_t tem variância condicional σ_t^2 não constante.

Processos Multiplicativos

Considere-se

$$u_t = \sigma_t \varepsilon_t$$

e as seguintes hipóteses:

H1 $\{\varepsilon_t\}$ é uma sequência de v.a. i.i.d. com $\mathbf{E}[\varepsilon_t] = 0$ e $\text{Var}[\varepsilon_t] = 1$;

H2 ε_t é independente de u_{t-k} , $k \in \mathbb{N}$;

H3 σ_t é \mathcal{F}_{t-1} mensurável ou \mathcal{F}_t mensurável (neste último caso se σ_t depender de variáveis exógenas no momento t).

Calcule-se $\mathbf{E}[u_t | \mathcal{F}_{t-1}]$ e $\text{Var}[u_t | \mathcal{F}_{t-1}]$ e conclua:

Processos multiplicativos do tipo $u_t = \sigma_t \varepsilon_t$ (H1-H3) são processos heterocedásticos (variância não constante) mas não autocorrelacionados.

Que função especificar para σ_t ou σ_t^2 ?

Volatility clustering: fortes variações são normalmente seguidas de fortes variações em ambos os sentidos $\Rightarrow \text{Corr}(r_t^2, r_{t-1}^2) > 0 \Rightarrow \text{Corr}(u_t^2, u_{t-1}^2) > 0$. Se u_{t-1}^2 é um valor alto (baixo), u_t^2 tenderá também a ser um valor alto (baixo).

Faz sentido escrever o seguinte modelo para σ_t^2 :

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2, \quad \omega > 0, \alpha_1 \geq 0.$$

Com efeito,

$$u_{t-1}^2 \text{ é "alto"} \Rightarrow \sigma_t^2 \text{ é "alto"} \Rightarrow u_t^2 = \sigma_t^2 \varepsilon_t^2 \text{ tende a ser "alto"}.$$

Distribuições de Caudas Pesada

Seja $u_t = \sigma_t \varepsilon_t$ sob as hipóteses H1-H3. Admita-se $\varepsilon_t \sim N(0, 1)$. Tem-se

$$\begin{aligned} E(u_t) &= 0 \\ \text{Var}(u_t) &= E(u_t^2) = E[\sigma_t^2] \\ E(u_t^3) &= 0 \Rightarrow \textit{skweness} = 0 \text{ (Verifique)} \\ k_u &= \frac{E(u_t^4)}{E(u_t^2)^2} > 3 \text{ [quadro]} \end{aligned}$$

Este resultado sugere que um modelo de HC pode ser adequado para modelar os retornos

Vantagens dos modelos de HC

Iremos ver que os modelos de HC permitem:

- modelar a volatilidade (e as covariâncias condicionais, no caso multivariado); como se sabe, a volatilidade é uma variável fundamental na análise do risco de mercado, na construção de *portfolios* dinâmicos, na valorização de opções, etc.;
- estimar de forma mais eficiente os parâmetros definidos na média condicional;
- estabelecer intervalos de confiança correctos para y . Isto é, se y exhibe HC e esta é negligenciada, os intervalos de previsão para y são incorrectos. Observe-se, com efeito, que os intervalos de confiança dependem da variância do erro de previsão e o erro de previsão depende (entre outros aspectos) da variância (condicional) da v.a. residual.

3.2 Modelo ARCH

Considere-se o seguinte modelo

$$y_t = \mu_t + u_t,$$

$$\mu_t = \mathbf{E}(y_t | \mathcal{F}_{t-1}) \text{ média condicional}$$

$$u_t = \sigma_t \varepsilon_t$$

Assumam-se as hipóteses H1-H3.

Definição u_t segue um modelo ARCH(q) (ou tem representação ARCH(q)) se

$$u_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \dots + \alpha_q u_{t-q}^2, \omega > 0, \alpha_i \geq 0$$

É importante constatar que $\sigma_t^2 \in \mathcal{F}_{t-1}$. website Robert Engle

Como a vol. exibe forte dependência temporal, raramente se considera $q = 1$.

Exercício de simulação:

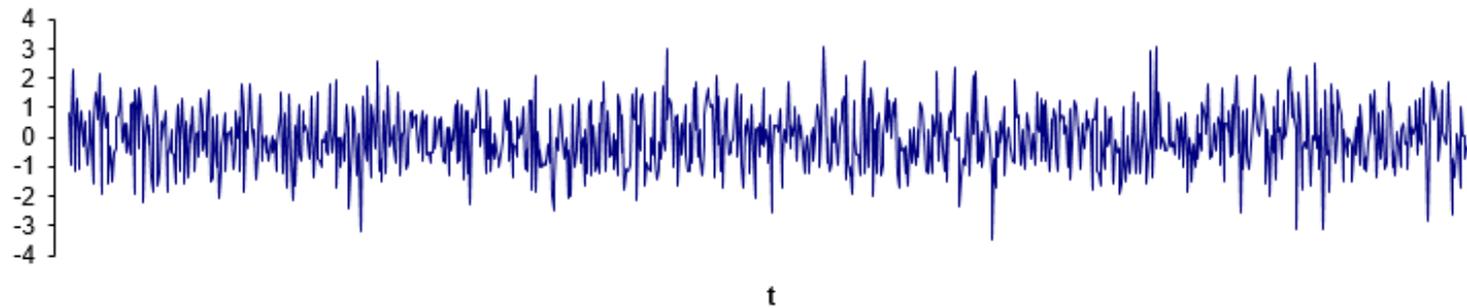
$$\begin{aligned}r_t &= u_t, & (\mu_t = 0) \\u_t &= \sigma_t \varepsilon_t, & \varepsilon_t \text{ RB Gaussiano com variância } 1 \\ \sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + \dots + \alpha_8 u_{t-8}^2.\end{aligned}$$

- ARCH(0) $\alpha_1 = \alpha_2 = \dots = \alpha_8 = 0$
- ARCH(1) $\alpha_1 = 0.8, \alpha_2 = \dots = \alpha_8 = 0$
- ARCH(3) $\alpha_1 = 0.3, \alpha_2 = 0.3, \alpha_3 = 0.2, \alpha_4 = \dots = \alpha_8 = 0$
- ARCH(8) $\alpha_1 = 0.2, \alpha_2 = \dots = \alpha_8 = 0.1$

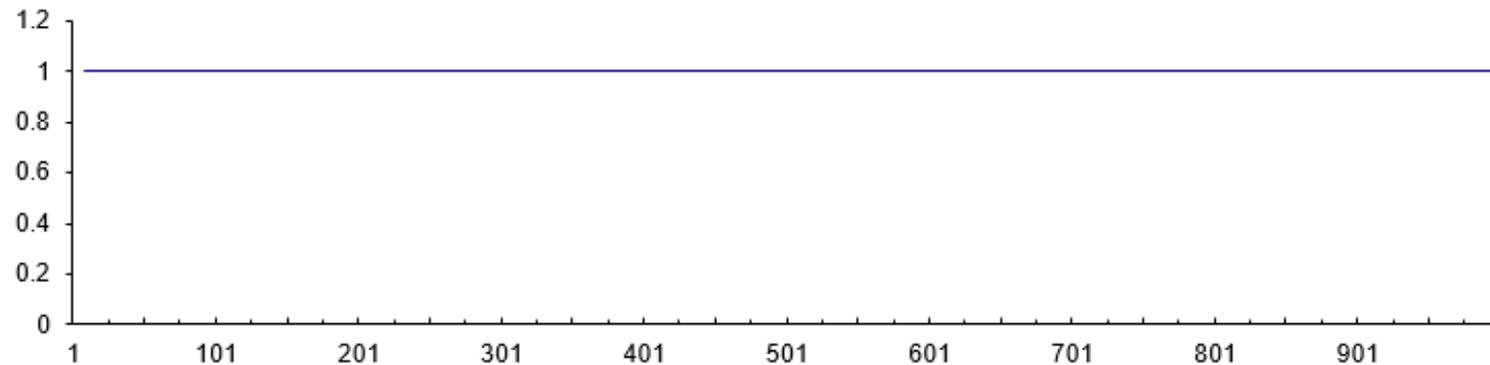
ARCH(0) $\alpha_1 = \alpha_2 = \dots = \alpha_8 = 0$

Parâmetros da Média		Parâmetros da Variância									kurt. Est.
constante	phi	w	alfa1	alfa2	alfa3	alfa4	alfa5	alfa6	alfa7	alfa8	
0	0	1	0	0	0	0	0	0	0	0	2.95

Retornos Simulados



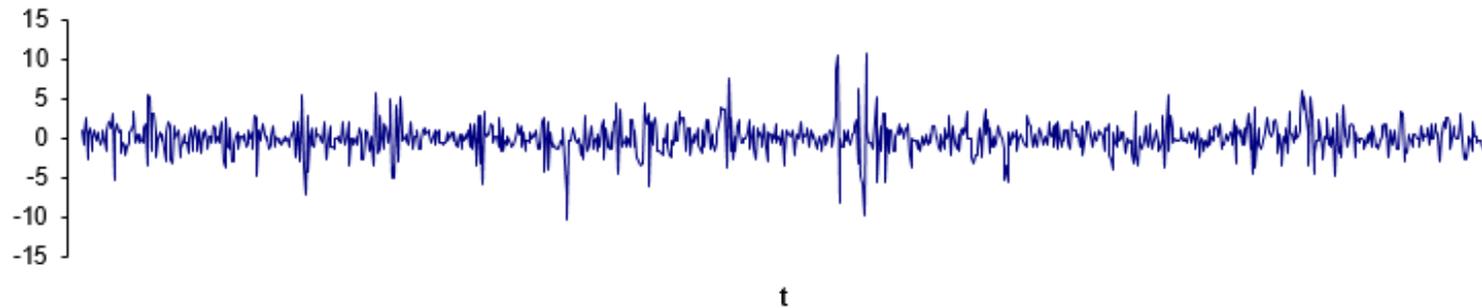
Volatilidade: sigma(t)



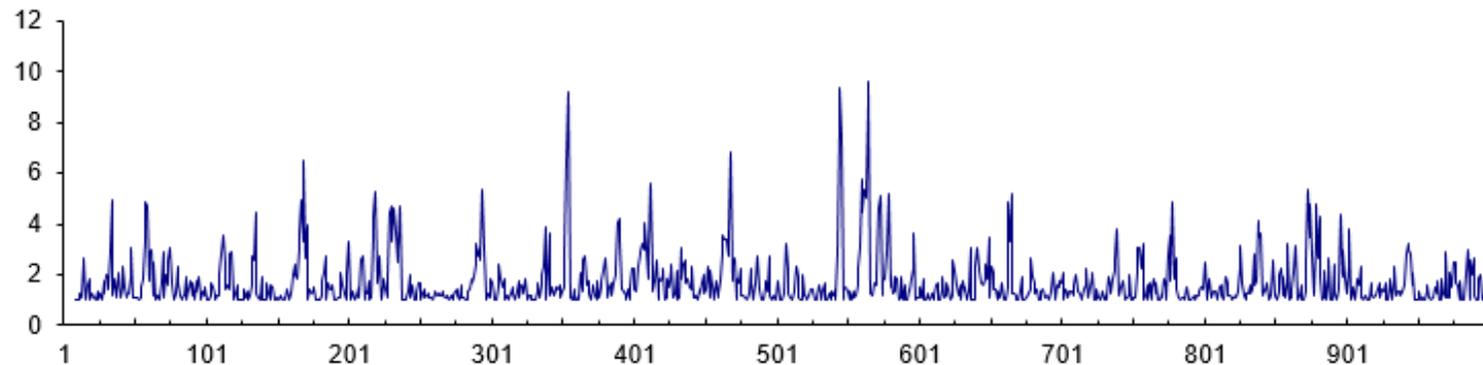
ARCH(1) $\alpha_1 = 0.8, \alpha_2 = \dots = \alpha_8 = 0$

Parâmetros da Média		Parâmetros da Variância								kurt. Est.	
constante	phi	w	alfa1	alfa2	alfa3	alfa4	alfa5	alfa6	alfa7		alfa8
0	0	1	0.8	0	0	0	0	0	0	0	7.23

Retornos Simulados



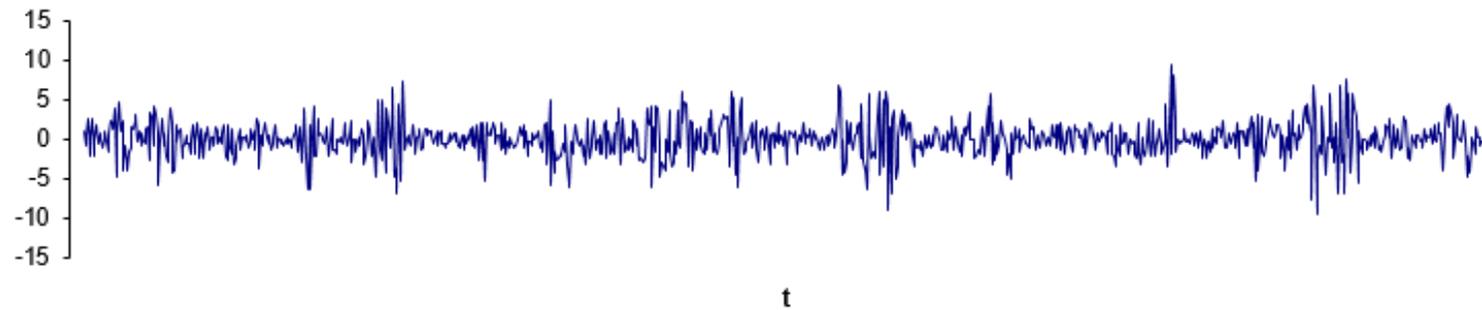
Volatilidade: sigma(t)



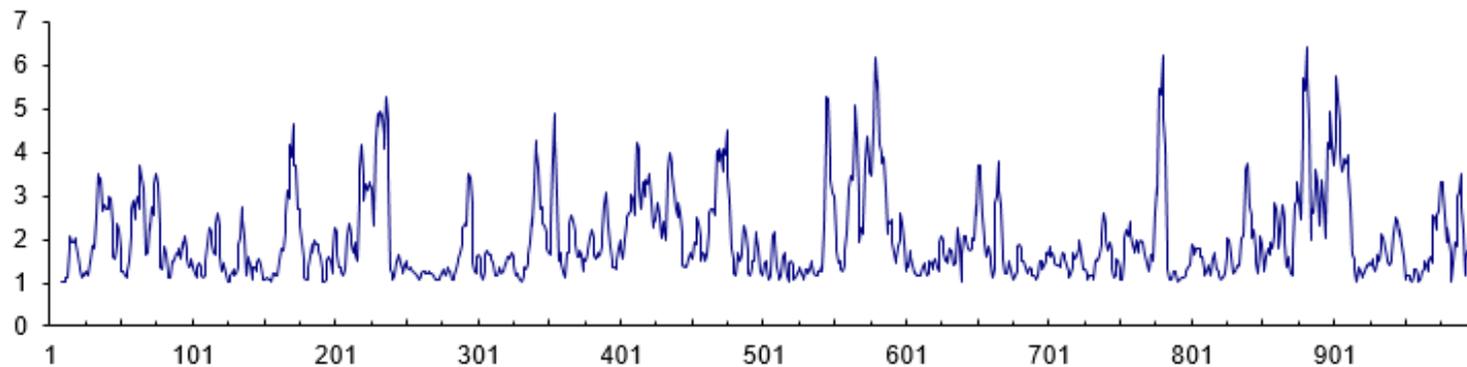
ARCH(3) $\alpha_1 = 0.3, \alpha_2 = 0.3, \alpha_3 = 0.2, \alpha_4 = \dots = \alpha_8 = 0$

Parâmetros da Média		Parâmetros da Variância									kurt. Est.
constante	phi	w	alfa1	alfa2	alfa3	alfa4	alfa5	alfa6	alfa7	alfa8	
0	0	1	0.3	0.3	0.2	0	0	0	0	0	4.78

Retornos Simulados



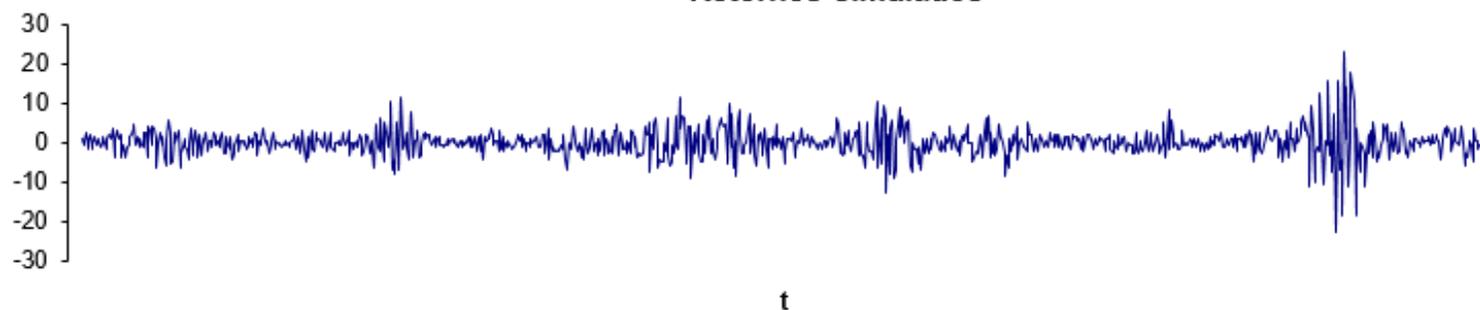
Volatilidade: sigma(t)



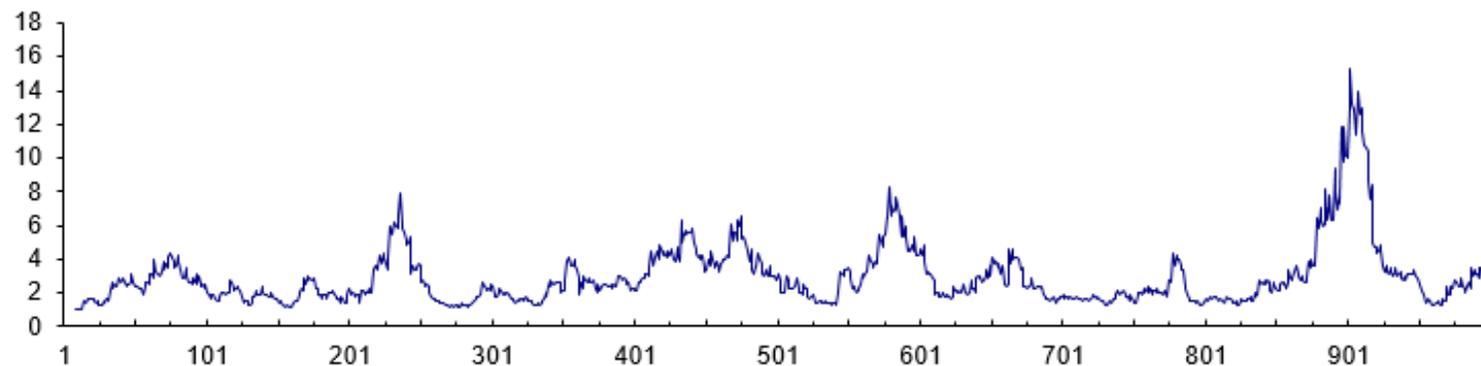
ARCH(8) $\alpha_1 = 0.2, \alpha_2 = \dots = \alpha_8 = 0.1$

Parâmetros da Média		Parâmetros da Variância									kurt. Est.
constante	phi	w	alfa1	alfa2	alfa3	alfa4	alfa5	alfa6	alfa7	alfa8	
0	0	1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	10.21

Retornos Simulados



Volatilidade: sigma(t)



3.2.1 Dois Primeiros Momentos de u_t

Como ε_t é independente de u_{t-k} , $k > 0$, segue-se que σ_t^2 (que é uma função de u_{t-k}) é independente de ε_t . Prove-se:

$$\begin{aligned} E(u_t) &= 0, \\ \text{Var}(u_t) &= E(\sigma_t^2) \\ \text{Cov}(u_t, u_{t-k}) &= 0, \quad k \in \mathbb{N}. \end{aligned}$$

3.2.2 Representação AR de um ARCH

Existem dependências no segundo momento do processo. A representação autoregressiva do processo ARCH mostra exactamente esse aspecto. Prove-se:

$$u_t^2 = \omega + \alpha_1 u_{t-1}^2 + v_t, \quad \alpha_1 \geq 0.$$

Como $E(v_t) = 0$ e $\text{Cov}(v_t, v_{t-k}) = 0$ conclui-se

$$u_t \sim \text{ARCH}(1) \Rightarrow u_t^2 \sim \text{AR}(1).$$

Logo o processo u_t^2 é autocorrelacionado (se $\alpha_1 > 0$) e apresenta as características básicas de um processo AR(1). De igual forma se conclui:

$$u_t \sim \text{ARCH}(q) \Rightarrow u_t^2 \sim \text{AR}(q).$$

3.2.3 Estacionaridade Segunda Ordem no ARCH(q)

Estude-se a ESO de u_t . Vimos que $E(u_t)$ e $\text{Cov}(u_t, u_{t-k})$ são finitos e não dependem de t ; só falta estudar $\text{Var}(u_t)$. O ARCH(1) na sua representação autoregressiva:

$$u_t^2 = \omega + \alpha_1 u_{t-1}^2 + v_t, \quad \alpha_1 \geq 0.$$

Da estrutura autoregressiva resulta que a condição de ESO é $0 \leq \alpha < 1$. Se u_t é ESO vem

$$E(u_t^2) = \frac{\omega}{1 - \alpha_1}$$

No caso geral, ARCH(q). Condição de ESO:

$$\alpha_1 + \alpha_2 + \dots + \alpha_q < 1, \quad (\alpha_i \geq 0).$$

Neste caso, depois de algumas contas, obtém-se

$$\text{Var}(u_t) = E(u_t^2) = \frac{\omega}{1 - (\alpha_1 + \alpha_2 + \dots + \alpha_q)}.$$

Observação: Embora a expressão $\text{Var} [u_t | \mathcal{F}_{t-1}]$ seja variável, $\text{Var} (u_t)$ é constante. Assim: u_t é condicionalmente heterocedástico (heterocedasticidade condicional) mas em termos não condicionais ou marginais, u_t é homocedástico. De forma análoga, também num processo estacionário, a média condicional é variável e a não condicional é constante. Por exemplo, num processo AR(1) estacionário, a média condicional é variável ao longo do tempo e dada por $\mu_t = c + \phi y_{t-1}$; no entanto, a média marginal $c / (1 - \phi)$ é constante.

3.2.4 FAC e FACP de um u_t^2 e Identificação do Processo ARCH(q)

Suponha-se que os quatro primeiros momentos de u são finitos não dependem de t . Vimos $u_t \sim \text{ARCH}(q) \Rightarrow u_t^2 \sim \text{AR}(q)$. Assim a FAC e a FACP de u^2 exibem o comportamento típico de um AR: [completar]

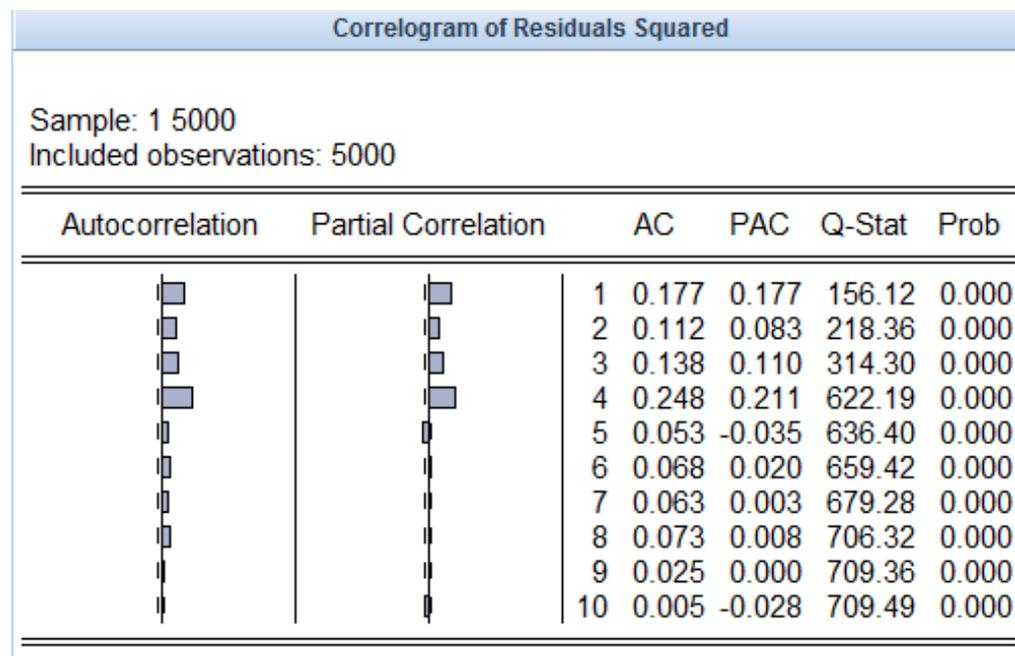
- ρ_k
- $\phi_{kk} = \begin{cases} \text{não se anula} & \text{se} \\ & \text{se} \end{cases}$
- Em particular, tem-se num ARCH(1)

$$\rho_k = \alpha_1^k, \quad k \geq 1, \quad \phi_{11} = \alpha \text{ e } \phi_{kk} = 0, \quad k \geq 2.$$

Passos para a identificação da ordem q de um processo ARCH(q)

1. Estima-se o modelo $y_t = \mu_t + u_t$ supondo σ_t^2 constante;
2. Obtêm-se os resíduos $\hat{u}_t = y_t - \hat{\mu}_t$, $t = 1, \dots, n$;
3. Calcula-se \hat{u}_t^2 , $t = 1, \dots, n$;
4. Calcula-se a FAC e a FACP de \hat{u}_t^2 e identifica-se a ordem q .

Na figura seguinte simulou-se um ARCH(q), $n = 5000$ observações. Qual a ordem de q ? Há um aspecto que confirma não se tratar de uma série financeira. Qual é?



3.2.5 Características da Distribuição Marginal de u_t

Suponha-se que $\varepsilon_t \sim N(0, 1)$. Então $u_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$. MAS:

- $E(u_t) = 0$
- $\text{Var}(u_t) = E(u_t^2) = \frac{\omega}{1 - (\alpha_1 + \dots + \alpha_q)}$
- $E(u_t^3) = 0 \Rightarrow \text{skewness} = 0$
- $k_u = \frac{E(u_t^4)}{E(u_t^2)^2} > k_\varepsilon = 3$ (já vimos).

- Pode-se demonstrar a distribuição marginal de u tem caudas de Pareto:

$$P(|u_t| > x) \sim cx^{-\kappa}$$

onde κ depende dos valores α_i .

Podemos obter uma expressão exacta para k_u . Suponha-se $u_t \sim \text{ARCH}(1)$, $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ e $3\alpha_1^2 < 1$. Pode-se mostrar:

$$k_u = 3 + \frac{6\alpha_1^2}{1 - 3\alpha_1^2} > 3 \quad [\text{fazer gráfico}]$$

Algumas conclusões:

- Embora $\{u_t\}$ seja um processo não autocorrelacionado, $\{u_t\}$ não é uma sequência de variáveis independentes (basta observar, por exemplo, $\text{Cov}(u_t^2, u_{t-1}^2) \neq 0$ ou que $E(u_t^2 | \mathcal{F}_{t-1})$ depende de u_{t-1}^2);
- Mesmo que u_t seja condicionalmente Gaussiano a distribuição marginal não é Gaussiana. Em particular, se u_t é condicionalmente Gaussiano então a distribuição marginal é leptocúrtica e tem caudas de Pareto.

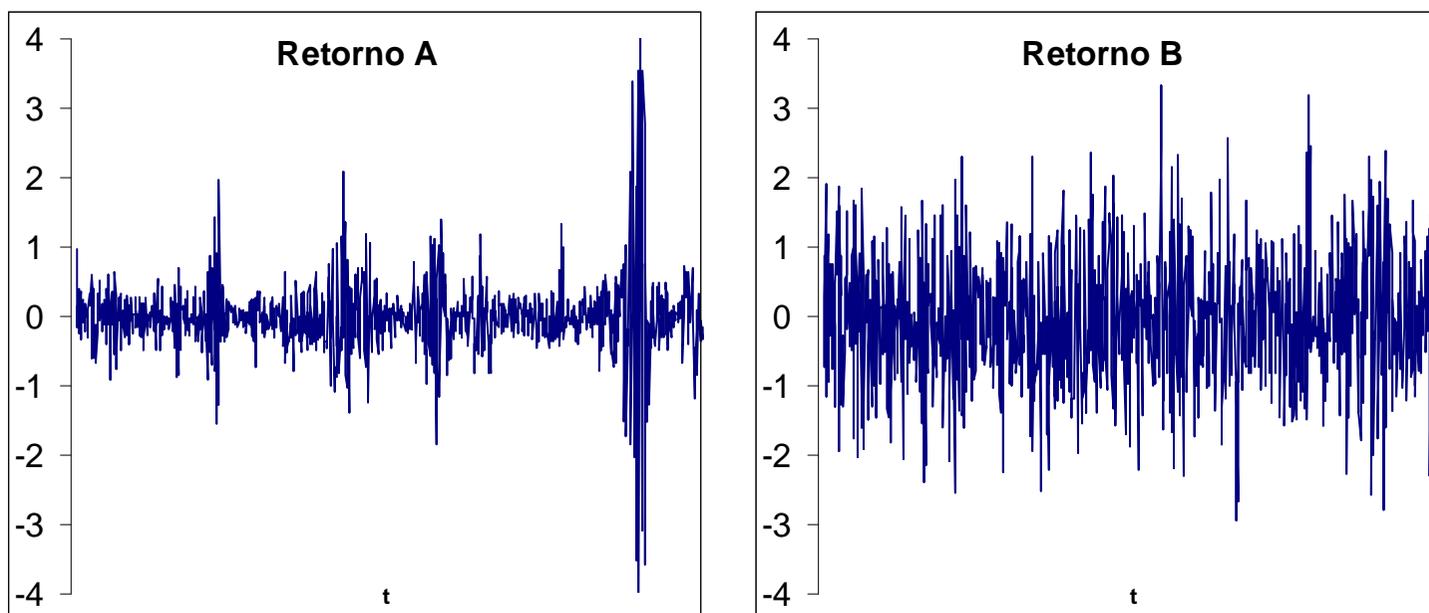
3.2.6 Momentos e Distribuição de y

Seja $y_t = \mu_t + u_t$, $u_t = \sigma_t \varepsilon_t$ (assumam-se as hipóteses habituais para ε_t). Então [completar]:

- $E(y_t | \mathcal{F}_{t-1}) =$
- $\text{Var}(y_t | \mathcal{F}_{t-1}) =$
- Se ε_t é Gaussiano então $y_t | \mathcal{F}_{t-1} \sim$
- $E(y_t) =$
- $\text{Var}(y_t) =$

3.2.7 Volatilidade: Definições

A **volatilidade condicional** no momento t ($= \sigma_t$) é uma medida da magnitude das variações (ou flutuações) não explicadas dos preços no momento t , $(y_t - \mu_t)$. No entanto, como $\mu_t \simeq 0$ podemos dizer que a volatilidade condicional é uma medida da magnitude das variações (ou flutuações) dos preços no momento t . É medida por σ_t . Um título A pode exibir, comparativamente a um outro título B, maior volatilidade condicional em certos momentos do tempo mas, globalmente A pode ser menos volátil do que B.



Qual é o retorno mais volátil?

Falamos neste caso de **volatilidade não condicional** que é uma medida da magnitude das variações (ou flutuações) dos preços num hiato de tempo (meses ou anos) (que não são explicadas). Pode ser medida através das estatísticas

$$\widehat{\text{Var}}(u_t) = \frac{\sum_{t=1}^n \hat{u}_t^2}{n} \text{ ou } \widehat{\text{Var}}(u_t) = \frac{\hat{\omega}}{1 - (\hat{\alpha}_1 + \dots + \hat{\alpha}_q)}.$$

No exemplo acima, o retorno B apresenta volatilidade constante.

3.3 Modelo GARCH

Tendo em conta a forte dependência temporal da volatilidade, era usual, nas primeiras aplicações (ainda na década de 80) considerar-se um ARCH de ordem elevada. Um ARCH de ordem elevada levanta problemas de estimação. A melhor solução apareceu com o modelo GARCH.

Definição u_t segue um modelo GARCH(p,q) se

$$\begin{aligned}u_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + \dots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2\end{aligned}$$

$\omega > 0, \alpha_i \geq 0, \beta \geq 0$.

Surpreendentemente, o modelo mais simples GARCH(1,1),

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

veio a revelar-se suficiente em muitas aplicações.

3.3.1 GARCH(p,q) representa um ARCH(∞)

Mostrar para o GARCH(1,1).

3.3.2 Representação ARMA de um GARCH

Mostrar para o GARCH(1,1).

Concluir: $u_t^2 \sim ARMA(1, 1)$. No caso geral pode-se mostrar

$$u_t \sim GARCH(p,q) \Rightarrow u_t^2 \sim ARMA(\max\{p, q\}, p).$$

Por exemplo,

$$u_t \sim GARCH(1,2) \Rightarrow u_t^2 \sim ARMA(2,1)$$

$$u_t \sim GARCH(2,1) \Rightarrow u_t^2 \sim ARMA(2,2)$$

$$u_t \sim GARCH(2,2) \Rightarrow u_t^2 \sim ARMA(2,2)$$

Em geral é problemático identificar o GARCH a partir das FAC e FACP de u_t^2 . Por duas razões:

1. O GARCH implica uma estrutura ARMA para u_t^2 e, como se sabe, no ARMA, nenhuma das funções de autocorrelação (FAC ou FACP) é nula a partir de certa ordem;
2. Não existe uma correspondência perfeita entre a estruturas ARMA e GARCH.

As funções de autocorrelação não são interessantes nesta fase? De forma alguma, por duas razões:

1. Se FAC e a FACP de u_t^2 não apresentarem coeficientes estatisticamente significativos então não existe efeito ARCH/GARCH;
2. A existência de vários coeficientes de autocorrelação e de autocorrelação parcial significativos é indício forte da presença de efeitos ARCH/GARCH.

Como regra geral, não devemos usar o ARCH em dados diários ou semanais; o GARCH é preferível. A identificação das ordens p e q do GARCH faz-se na fase da estimação (testes de significância/critérios de informação).

3.3.3 Estacionaridade de Segunda Ordem num GARCH(p,q)

Como se sabe $E(u_t) = \text{Cov}(u_t, u_{t-k}) = 0$, $\forall k \in \mathbb{N}$. Assim, para discutir a ESO do processo u , basta analisar $E(u_t^2)$.

Pode-se provar que a condição de ESO é

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1.$$

Note-se: $u_t^2 \sim \text{ARMA}(\max\{p, q\}, p)$.

No caso GARCH(1,1):

$$\alpha_1 + \beta_1 < 1.$$

3.4 Modelo IGARCH

Vamos analisar apenas o IGARCH(1,1):

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad \alpha_1 + \beta_1 = 1.$$

A designação *Integrated* GARCH resulta do facto de u_t^2 possuir uma raiz unitária:

$$\begin{aligned} u_t^2 &= \omega + \underbrace{(\alpha_1 + \beta_1)}_1 u_{t-1}^2 - \beta_1 v_{t-1} + v_t \\ u_t^2 &= \omega + u_{t-1}^2 - \beta_1 v_{t-1} + v_t \\ (1 - L) u_t^2 &= \omega - \beta_1 v_{t-1} + v_t \end{aligned}$$

Nestas condições, u_t não é ESO. Durante algum tempo pensou-se que u_t seria também não estacionário em sentido estrito. Isso é falso. Pode-se mostrar:

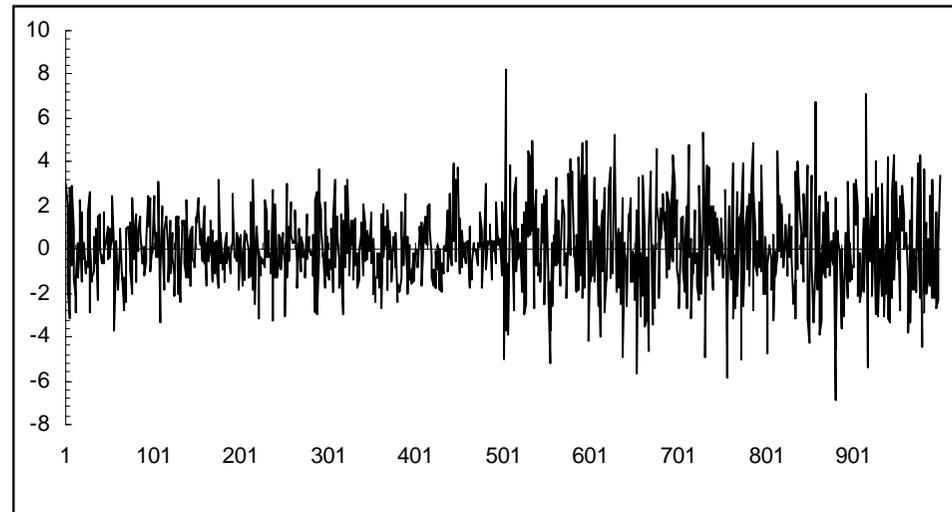
- 1) A condição necessária e suficiente para que u_t seja EE é $\mathbf{E} \left[\log \left(\beta_1 + \alpha_1 \varepsilon_t^2 \right) \right] < 0$;
- 2) Esta condição acaba por ser menos exigente que a condição de ESO, $\alpha_1 + \beta_1 < 1$.

3.4.1 Alterações de Estrutura e o IGARCH

Modelos aparentemente IGARCH podem também dever-se a alterações de estrutura (tal como processos aparentemente do tipo $y_t = y_{t-1} + u_t$ podem dever-se a alterações de estrutura). Por exemplo, considere-se a seguinte simulação de Monte Carlo:

$$\begin{aligned}
 y_t &= u_t, & t &= 1, 2, \dots, 1000 \\
 u_t &= \sigma_t \varepsilon_t \\
 \sigma_t^2 &= \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2, \\
 \alpha &= 0.1, \quad \beta = 0.6, \quad \omega &= \begin{cases} 0.5 & t = 1, 2, \dots, 500 \\ 1.5 & t = 501, 502, \dots, 1000 \end{cases}
 \end{aligned}$$

Este modelo foi simulado 500 vezes. Na figura seguinte representa-se uma das 500 trajetórias simuladas



Em cada simulação estimaram-se os parâmetros. Embora o modelo simulado não seja claramente um IGARCH, conclui-se que:

- em 83% das simulações a soma dos parâmetros $\hat{\alpha} + \hat{\beta}$ esteve acima de 0.99;
- em 99.6% das simulações a soma dos parâmetros $\hat{\alpha} + \hat{\beta}$ esteve acima de 0.95;
- em todos os casos $\hat{\alpha} + \hat{\beta}$ esteve acima de 0.9.

Assim, numa aplicação empírica, se $\hat{\alpha} + \hat{\beta}$ estiver próximo de um, convém verificar se o modelo subjacente é de facto um IGARCH ou, pelo contrário, se existem alterações de estrutura que causem um falso IGARCH. Neste último caso modelar a alteração de estrutura, por exemplo, através de uma dummy.

3.5 Modelo GJR-GARCH

O modelo GJR-GARCH é devido a Glosten, Jagannathan e Runkle.

Efeito assimétrico (i.e. momentos de maior volatilidade são despoletados por variações negativas nos preços). Vimos a medida (*naive*) $\text{Cov}(r_t^2, r_{t-1}) < 0$.

Veremos agora uma forma bastante mais eficiente de estimar esse efeito. O modelo ARCH/GARCH apenas detecta o chamado “efeito magnitude”. Isto é, a volatilidade só responde à magnitude do valor de u_t . Em esquema:

$$\uparrow u_{t-1}^2 \Rightarrow \uparrow \sigma_t^2 \quad (\text{efeito magnitude})$$

Para modelar o efeito assimétrico é necessário que a volatilidade responda assimetricamente ao sinal de u_t . Mais precisamente, a volatilidade deve aumentar mais quando $u_t < 0$ (“má notícia”) do que quando $u_t > 0$ (“boa notícia”).

Modelo GJR-GARCH \rightarrow modela não só o efeito magnitude como também o efeito assimétrico. A especificação mais simples:

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma_1 u_{t-1}^2 \mathcal{I}_{\{u_{t-1} < 0\}},$$

$$\mathcal{I}_{\{u_{t-1} < 0\}} = \begin{cases} 1 & \text{se } u_{t-1} < 0 \\ 0 & \text{se } u_{t-1} \geq 0. \end{cases}$$

De acordo com o efeito assimétrico devemos esperar $\gamma_1 > 0$. Para ensaiar o efeito assimétrico podemos considerar o ensaio $H_0: \gamma_1 = 0$.

Pode-se provar que a condição de ESO (no caso em que a distribuição de ε é simétrica) é $\alpha_1 + \gamma_1/2 + \beta_1 < 1$. Nesta circunstâncias,

$$\text{Var}(u_t) = \mathbf{E}[\sigma_t^2] = \frac{\omega}{1 - (\alpha_1 + \gamma_1/2 + \beta_1)}.$$

3.6 Modelo de Het. Cond. com Variáveis Explicativas

Considere-se um modelo GARCH(1,1) (poderia ser outro modelo qualquer) com variáveis explicativas:

$$\begin{aligned}y_t &= \mu_t + u_t \\u_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + g(x_t)\end{aligned}$$

onde $\omega + g : \mathbb{R} \rightarrow \mathbb{R}^+$ (exigimos que $g(x_t) + \omega > 0$). Que variáveis poderemos considerar para x_t ? Vejamos alguns exemplos:

- Dias da semana (ou qualquer outro efeito de calendário):

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \delta_1 S_t + \delta_2 T_t + \delta_3 Q_t^a + \delta_5 Q_t^u$$

onde $S_t = 1$ se t é uma segunda-feira, etc. (deverá ter-se $\omega + \min \{\delta_i\} > 0 \Rightarrow \sigma_t^2 > 0$).

- Ocorrência de factos, notícias significativas. Por exemplo,

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \delta_1 good_t + \delta_2 bad_t$$

$$good_t = \begin{cases} 1 & t = \text{são divulgados resultados da empresa ABC acima do esperado} \\ 0 & 0 \end{cases}$$

$$bad_t = \begin{cases} 1 & t = \text{são divulgados resultados da empresa ABC abaixo do esperado} \\ 0 & 0 \end{cases}$$

- Variação do preço do crude.
- Medida de volatilidade de outro activo/mercado.
- Volume de transacções:

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \delta_1 vol_{t-1}^*$$

onde $vol_{t-1}^* = \frac{vol_{t-1}}{\sigma_{vol}}$ ou $vol_{t-1}^* = \log(vol_{t-1})$ ou $vol_{t-1}^* = vol_{t-1}/vol_{t-2}$. Observe-se que o volume pode ser considerado como uma variável *proxy* da variável não observada chegada de informação.

- Qualquer outra variável (estacionária) que supostamente afecte a volatilidade.

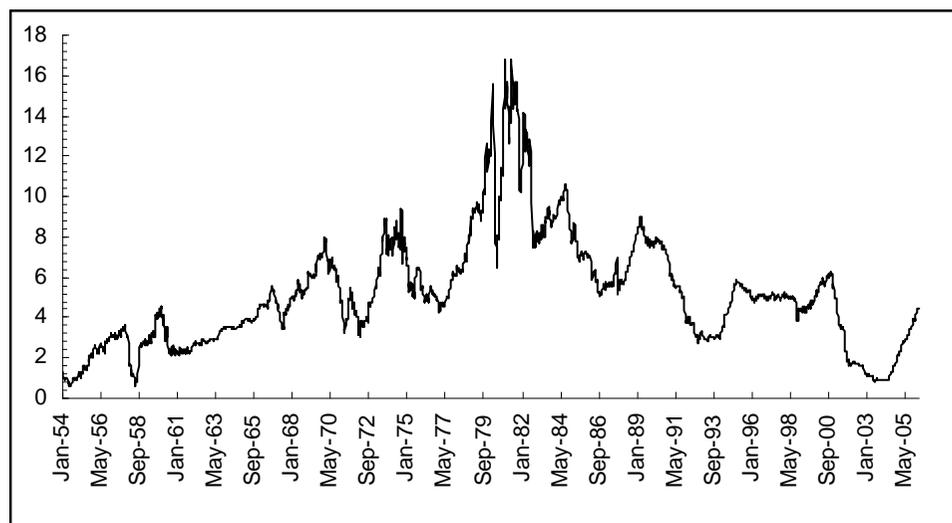
Exemplo A literatura dos modelos de taxas de juro (a um factor) sugere que a volatilidade da taxa de juro depende do nível da taxa de juro:

$$r_t = c + \phi r_{t-1} + u_t$$

$$u_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma r_{t-1}.$$

O ensaio $H_0: \gamma = 0$ vs. $H_1: \gamma > 0$ permite analisar se a nível da taxa de juro influencia positivamente a volatilidade. Geralmente conclui-se $\gamma > 0$. A figura sugere (claramente) $\gamma > 0$.



Taxa de Juro (Bilhetes do Tesouro a 3 meses -EUA)

3.7 Estimação

Considere-se

$$\begin{aligned}y_t &= \mu_t(\boldsymbol{\theta}) + u_t \\u_t &= \sigma_t(\boldsymbol{\theta}) \varepsilon_t, \quad \varepsilon_t \sim D(0, 1)\end{aligned}$$

A v.a. ε_t tem distribuição conhecida D (normal, t-Student ou outra) de média zero e variância um.

Por exemplo,

$$\begin{aligned}y_t &= \gamma_0 + \gamma_1 x_t + \phi_1 y_{t-1} + u_t \\u_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1) \\\sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \delta \text{seg}_t\end{aligned}$$

$$\boldsymbol{\theta} = (\gamma_0, \gamma_1, \phi_1, \omega, \alpha_1, \beta_1, \delta)'$$

3.7.1 Estimador de Máxima Verosimilhança

O estimador de máxima verosimilhança é

$$\hat{\theta}_n = \arg \max_{\theta} \sum_{t=1}^n \log f(y_t | \mathcal{I}_t; \theta), \quad \mathcal{I}_t = \{y_{t-1}, y_{t-2}, \dots; \mathbf{x}_t, \mathbf{x}_{t-1}, \dots\}.$$

Escreva-se $\log f(y_t | \mathcal{I}_t; \theta)$ no caso $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, 1)$.

Pode-se mostrar:

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

onde $I(\theta_0) = A(\theta_0) = B(\theta_0)$ é a matriz de informação de Fisher e

$$A(\theta) = -\mathbf{E} \left(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right), \quad B(\theta) = \mathbf{E} \left(\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta'} \right)$$

Ver EVIEWS

3.7.2 Estimador de Pseudo (ou Quase) Máxima Verosimilhança

Na prática, a distribuição de ε_t não é conhecida. Podemos ainda assim supor, por exemplo, $\varepsilon_t \sim N(0, 1)$ ou $\varepsilon_t \sim t(n)$? A resposta é afirmativa.

Suponhamos que a verdadeira mas desconhecida fdp condicional de y é f . O estimador de máxima verosimilhança

$$\hat{\theta}_n = \arg \max_{\theta} \sum_{t=1}^n \log f(y_t | \mathcal{I}_t; \theta)$$

não pode ser implementado, pois a função f é desconhecida.

O estimador de pseudo máxima verosimilhança usa como pseudo verdadeira fdp a função h (que na generalidade dos casos é diferente de f),

$$\hat{\theta}_n^{pmv} = \arg \max_{\theta} \sum_{t=1}^n \log h(y_t | \mathcal{I}_t; \theta).$$

Sob certas condições, mesmo que $h \neq f$, o estimador de pseudo máxima verosimilhança apresenta boas propriedades. As condições são:

- h pertence à família das densidades exponenciais quadráticas (a normal e a t-Student, entre muitas outras distribuições, pertencem a esta família);
- a média condicional está bem especificada;
- a variância condicional está bem especificada.

Pode-se provar, sob estas condições:

$$\hat{\boldsymbol{\theta}}_n^{pmv} \xrightarrow{p} \boldsymbol{\theta}_0$$

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_n^{pmv} - \boldsymbol{\theta}_0 \right) \xrightarrow{d} N \left(0, A(\boldsymbol{\theta}_0)^{-1} B(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0)^{-1} \right)$$

Se, por acaso, a função h é a própria função f , i.e., $f = h$, então o estimador de pseudo máxima verosimilhança é o estimador de máxima verosimilhança e $A(\boldsymbol{\theta}_0) = B(\boldsymbol{\theta}_0)$.

Em suma, mesmo que a distribuição de ε_t não seja conhecida podemos supor, por exemplo, $\varepsilon_t \sim N(0, 1)$ (ou $\varepsilon_t \sim D$ tal que a densidade h satisfaça as condições estabelecidas), porque $\hat{\theta}_n^{pmv}$ é ainda assim um estimador consistente (embora não assintoticamente eficiente) e tem distribuição assintótica normal. O único cuidado adicional é tomar como matriz de variâncias-covariâncias (assintótica) a expressão $A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1}$ e não $I(\theta_0)^{-1}$. Como fazer isto no EVIEWS? Basta escolher a opção “heteroskedasticity consistent covariance (Bollerslev-Wooldrige)” no menu “options” da estimação. Esta opção convém estar sempre activa.

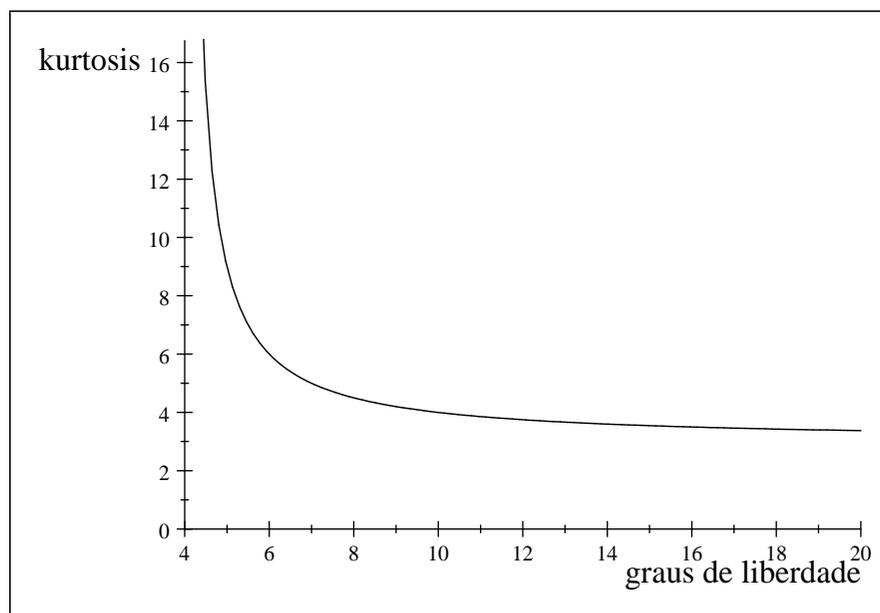
3.7.3 Método da Máxima Verosimilhança Revisitado: Distribuições Não Normais

No âmbito do método da máxima (ou da pseudo máxima) verosimilhança, normalmente assume-se $\varepsilon_t \sim N(0, 1)$. Contudo, verifica-se habitualmente que os resíduos estandardizados, $\hat{\varepsilon} = \hat{u}_t / \hat{\sigma}_t$ apresentam um valor de *kurtosis* quase sempre acima do valor 3, i.e., $k_{\hat{\varepsilon}} > 3$. Este resultado é, até certo ponto, inesperado. Porquê? Também a distribuição condicional $u_t | \mathcal{F}_{t-1}$ (e não só a marginal) é leptocúrtica.

Já vimos uma forma de lidar com este problema: basta tomar o estimador de pseudo máxima verosimilhança. Uma alternativa consiste em formular uma distribuição leptocúrtica para ε_t tal que $E[\varepsilon_t] = 0$ e $\text{Var}[\varepsilon_t] = 1$.

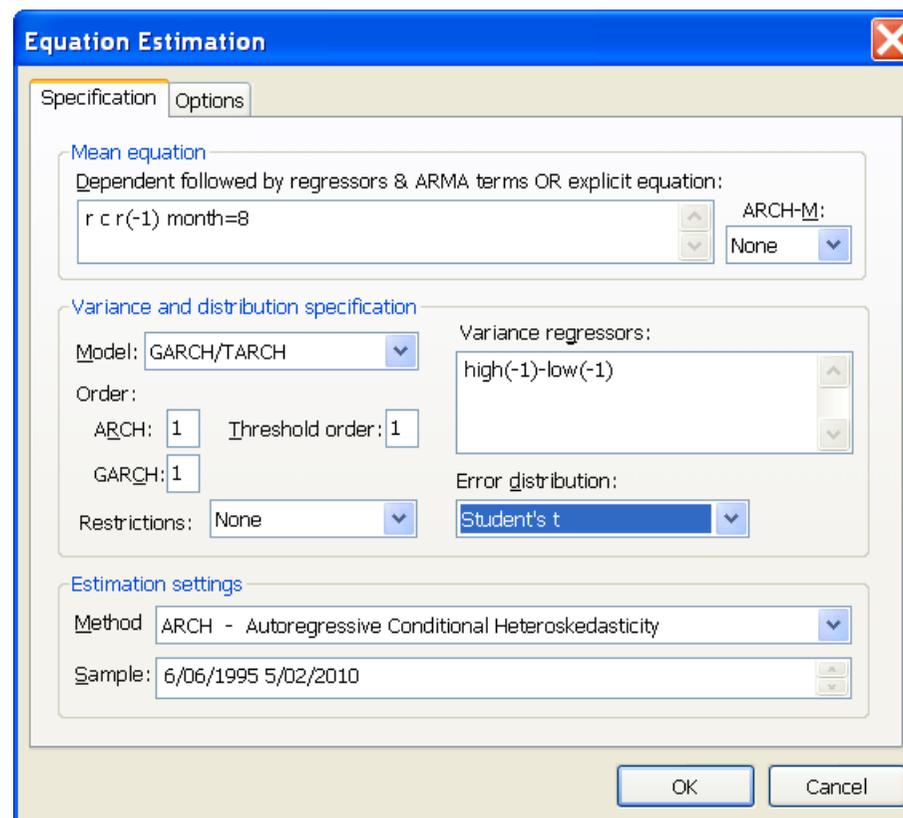
Hipótese: $\varepsilon_t \sim \mathbf{t}\text{-Student}(v)$

Estima-se também v , n^o de graus de liberdade. Se \hat{v} é baixo (abaixo de 10) \Rightarrow evidência de caudas pesadas para a dist. condicional



$$\text{Kurosis } k_{\varepsilon} = 3 + \frac{6}{v-4}$$

EIEWS:



The screenshot shows the "Equation Estimation" dialog box in EViews. It is divided into three main sections: "Mean equation", "Variance and distribution specification", and "Estimation settings".

- Mean equation:** The "Dependent followed by regressors & ARMA terms OR explicit equation:" field contains the text "r c r(-1) month=8". To the right, the "ARCH-M:" dropdown menu is set to "None".
- Variance and distribution specification:**
 - The "Model:" dropdown menu is set to "GARCH/TARCH".
 - The "Order:" section has "ARCH:" set to 1 and "Threshold order:" set to 1.
 - The "GARCH:" field is set to 1.
 - The "Restrictions:" dropdown menu is set to "None".
 - The "Variance regressors:" field contains the text "high(-1)-low(-1)".
 - The "Error distribution:" dropdown menu is set to "Student's t".
- Estimation settings:**
 - The "Method:" dropdown menu is set to "ARCH - Autoregressive Conditional Heteroskedasticity".
 - The "Sample:" field shows the date range "6/06/1995 5/02/2010".

At the bottom right of the dialog box, there are "OK" and "Cancel" buttons.

Escreva as equações que definem o modelo.

Output:

Dependent Variable: R
 Method: ML - ARCH (Marquardt) - Student's t distribution
 Sample: 6/06/1995 5/02/2010
 Included observations: 3689

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.058357	0.023806	2.451391	0.0142
R(-1)	-0.044791	0.015821	-2.831121	0.0046
MONTH=8	-0.129500	0.072937	-1.775506	0.0758
Variance Equation				
C	0.006052	0.008292	0.729870	0.4655
RESID(-1)^2	0.048468	0.009039	5.361915	0.0000
RESID(-1)^2*(RESID(-1)<0)	0.061235	0.015967	3.835009	0.0001
GARCH(-1)	0.920114	0.008649	106.3830	0.0000
HIGH(-1)-LOW(-1)	0.092678	0.043538	2.128673	0.0333
T-DIST. DOF	5.841394	0.540589	10.80561	0.0000
R-squared	0.000912	Mean dependent var	0.047209	
Adjusted R-squared	0.000370	S.D. dependent var	2.133478	
S.E. of regression	2.133083	Akaike info criterion	3.947494	
Sum squared resid	16771.46	Schwarz criterion	3.962652	
Log likelihood	-7272.152	Hannan-Quinn criter.	3.952889	
F-statistic	0.420581	Durbin-Watson stat	1.961961	
Prob(F-statistic)	0.909348			

3.8 Ensaios Estatísticos

Interessa:

- verificar se existe evidência do efeito ARCH;
- analisar a adequabilidade do modelo estimado;
- testar certa suposição sobre os α_i, β_i etc. a partir do modelo final selecionado.

3.8.1 Ensaios Pré-Estimação ARCH

Teste ARCH (teste multiplicador de Lagrange)

Considere-se

$$y_t = \mu_t + u_t$$

$$u_t = \sigma_t \varepsilon$$

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \dots + \alpha_q u_{t-q}^2.$$

Existe efeito ARCH se pelo menos um parâmetro α_i for diferente de zero. Se todos forem zero, não existe efeito ARCH.

Pode-se provar, sob a hipótese $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_q = 0$ que

$$nR^2 \xrightarrow{d} \chi_{(q)}^2$$

onde R^2 é o coeficiente de determinação da regressão de \hat{u}_t^2 sobre as variáveis

$$\left[1 \quad \hat{u}_{t-1}^2 \quad \dots \quad \hat{u}_{t-q}^2 \right].$$

Para a realização do teste os passos são:

1. Estimar o modelo $y_t = \mu_t + u_t$ supondo $\sigma_t^2 = \sigma^2$ constante;
2. Obter os resíduos $\hat{u}_t = y_t - \hat{\mu}_t$, $t = 1, \dots, n$;
3. Fazer a regressão OLS de \hat{u}_t^2 sobre $\left[\mathbf{1} \quad \hat{u}_{t-1}^2 \quad \dots \quad \hat{u}_{t-q}^2 \right]$;
4. Obter R^2 , nR^2 , etc.

FAC de \hat{u}_t^2

Como se viu, a existência de um processo GARCH implica a correlação das variáveis u_t^2 e u_{t-k}^2 . O teste Ljung-Box é assintoticamente equivalente ao teste ARCH. A sua hipótese nula é $H_0: \rho_1(\hat{u}_t^2) = \dots = \rho_m(\hat{u}_t^2) = 0$, sendo $\rho_i(\hat{u}_t^2)$ o coeficiente de autocorrelação entre \hat{u}_t^2 e \hat{u}_{t-i}^2 .

$$Q = n(n+2) \sum_{i=1}^m \frac{1}{n-i} \hat{\rho}_i^2(\hat{u}_t^2) \xrightarrow{d} \chi_{(m-k)}^2$$

onde k é o número de parâmetros estimados. Evidência contra a hipótese nula sugere a existência de um efeito ARCH.

EViews

Dependent Variable: R
 Method: **Least Squares**
 Sample: 6/06/1995 5/02/2010
 Included observations: 3689

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.069015	0.036806	1.875110	0.0609
R(-1)	-0.024767	0.016468	-1.503933	0.1327
MONTH=8	-0.231417	0.123256	-1.877532	0.0605

R-squared	0.001518	Mean dependent var	0.047209
Adjusted R-squared	0.000976	S.D. dependent var	2.133478
S.E. of regression	2.132436	Akaike info criterion	4.353220
Sum squared resid	16761.28	Schwarz criterion	4.358273
Log likelihood	-8026.514	Hannan-Quinn criter.	4.355018
F-statistic	2.801919	Durbin-Watson stat	2.001065
Prob(F-statistic)	0.060823		

Heteroskedasticity Test: ARCH

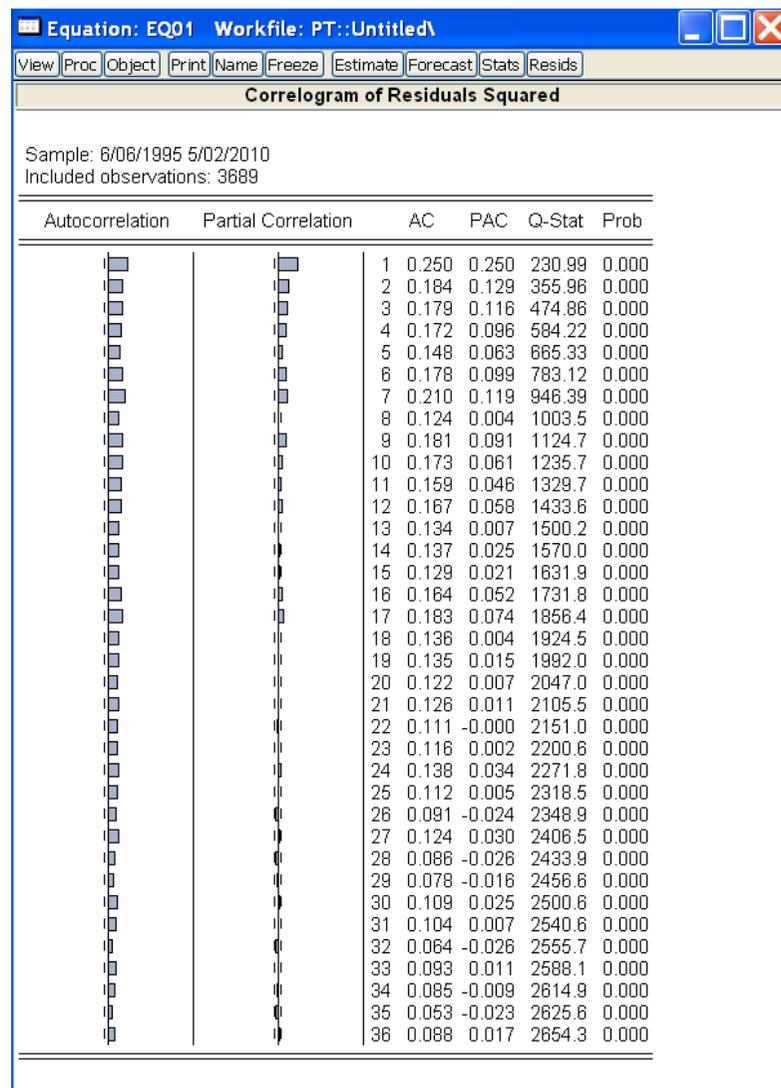
F-statistic	56.94060	Prob. F(10,3668)	0.0000
Obs*R-squared	494.3696	Prob. Chi-Square(10)	0.0000

Test Equation:
 Dependent Variable: RESID^2
 Method: Least Squares
 Date: 21/02/10 Time: 23:59
 Sample (adjusted): 20/06/1995 5/02/2010
 Included observations: 3679 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.509621	0.227111	6.647061	0.0000
RESID^2(-1)	0.161080	0.016481	9.773582	0.0000
RESID^2(-2)	0.064220	0.016649	3.857279	0.0001
RESID^2(-3)	0.058670	0.016681	3.517210	0.0004
RESID^2(-4)	0.058059	0.016614	3.494605	0.0005
RESID^2(-5)	0.026900	0.016605	1.620025	0.1053
RESID^2(-6)	0.066473	0.016605	4.003208	0.0001
RESID^2(-7)	0.107843	0.016616	6.490505	0.0000
RESID^2(-8)	-0.015260	0.016683	-0.914662	0.3604
RESID^2(-9)	0.080552	0.016652	4.837491	0.0000
RESID^2(-10)	0.061323	0.016490	3.718813	0.0002

R-squared	0.134376	Mean dependent var	4.553870
Adjusted R-squared	0.132016	S.D. dependent var	11.92536
S.E. of regression	11.11034	Akaike info criterion	7.656615
Sum squared resid	452776.6	Schwarz criterion	7.675184
Log likelihood	-14073.34	Hannan-Quinn criter.	7.663225
F-statistic	56.94060	Durbin-Watson stat	2.005539
Prob(F-statistic)	0.000000		

O teste Ljung-Box corrobora as conclusões do teste ARCH:



3.8.2 Ensaios Pós-Estimação

Depois do modelo estimado (pelo método da máxima (ou da pseudo) verosimilhança) há interesse em testar determinada suposição envolvendo os parâmetros ou em analisar a adequabilidade do modelo.

Tem-se

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, V_0). \quad (*)$$

$$W = n (\mathbf{R}\hat{\boldsymbol{\theta}}_n - \mathbf{r})' (\mathbf{R}\hat{V}_0\mathbf{R}')^{-1} (\mathbf{R}\hat{\boldsymbol{\theta}}_n - \mathbf{r}) \xrightarrow{d} \chi_{(p)}^2 \quad (**)$$

(*) permite analisar a significância individual das estimativas;

(**) permite analisar a significância conjunta das estimativas, por exemplo, no contexto do GARCH(1,1), poderemos estar interessados em testar

$$H_0 : \alpha + \beta = 1$$

Testes de Diagnóstico

O modelo em análise é $y_t = \mu_t + u_t$, $u_t = \sigma_t \varepsilon_t$ e as hipóteses são: $\{\varepsilon_t\}$ é um processo diferença de martingala ou ruído branco e $\{\varepsilon_t\}$ é um processo homocedástico. Nestas circunstância, se o modelo está bem especificado, deve ter-se: $\{\varepsilon_t\}$ **deve ser não autocorrelacionado** e $\{\varepsilon_t\}$ **deve ser condicionalmente homocedástico**. Assim, se

- (a) y é, por exemplo, um ARMA e a média condicional não captar esta estrutura, os processos $\{u_t\}$ e $\{\varepsilon_t\}$ exibirão autocorrelação;
- (b) de igual forma, se y segue um GARCH e a variância condicional não captar esta estrutura $\varepsilon_t^2 = u_t^2 / \sigma_t^2$ exibirá autocorrelação;
- (c) finalmente, se ε segue uma distribuição leptocúrtica então $k_{\hat{\varepsilon}} > 3$.

Para analisar (a), (b) e (c) devemos:

1. estimar um modelo ARMAX+GARCH;
2. obter os resíduos \hat{u}_t ;
3. obter os resíduos estandardizados $\hat{\varepsilon}_t = \hat{u}_t / \hat{\sigma}_t$;

Análise da questão (a). Efectuar o teste Ljung-Box tomando como hipótese nula, H_0 : $\rho_1(\hat{\varepsilon}_t) = \dots = \rho_m(\hat{\varepsilon}_t) = 0$ ($\rho_i(\hat{\varepsilon}_t)$ é o coeficiente de autocorrelação entre $\hat{\varepsilon}_t$ e $\hat{\varepsilon}_{t-i}$) e estatística de teste

$$Q = n(n+2) \sum_{i=1}^m \frac{1}{n-i} \hat{\rho}_i^2(\hat{\varepsilon}_t) \xrightarrow{d} \chi_{(m-k)}^2$$

onde k é o número de parâmetros estimados. Evidência contra a hipótese nula sugere que $\hat{\varepsilon}_t$ é autocorrelacionado. Neste caso é necessário rever a especificação da média condicional.

Análise da questão (b). Efectuar o teste Ljung-Box tomando como hipótese nula, H_0 : $\rho_1(\hat{\varepsilon}_t^2) = \dots = \rho_m(\hat{\varepsilon}_t^2) = 0$ ($\rho_i(\hat{\varepsilon}_t^2)$ é o coeficiente de autocorrelação entre $\hat{\varepsilon}_t^2$ e $\hat{\varepsilon}_{t-i}^2$) e estatística de teste

$$Q = n(n+2) \sum_{i=1}^m \frac{1}{n-i} \hat{\rho}_i^2(\hat{\varepsilon}_t^2) \xrightarrow{d} \chi_{(m-k)}^2$$

onde k é o número de parâmetros estimados. Evidência contra a hipótese nula sugere que $\hat{\varepsilon}_t^2$ é autocorrelacionado. Neste caso é necessário rever a especificação da variância condicional.

3.9 Previsão

Suponha-se que y segue um modelo do tipo ARMA+GARCH. Temos $\{y_1, y_2, \dots, y_n\}$ e procura-se:

- prever y_{n+1}, y_{n+2}, \dots ;
- estabelecer ICs para y_{n+1}, y_{n+2}, \dots ;
- prever $\sigma_{n+1}^2, \sigma_{n+2}^2, \dots$;
- estabelecer ICs para $\sigma_{n+1}^2, \sigma_{n+2}^2, \dots$;

3.9.1 Previsão da Variância Condicional

Previsor com EQM mínimo para σ_{n+h}^2 (dada a informação em \mathcal{F}_n):

$$\mathbf{E} \left(\sigma_{n+h}^2 \mid \mathcal{F}_n \right).$$

Note-se:

$$\begin{aligned} \mathbf{E} \left(u_{n+h}^2 \mid \mathcal{F}_n \right) &= \mathbf{E} \left(\sigma_{n+h}^2 \varepsilon_{n+h}^2 \mid \mathcal{F}_n \right) = \mathbf{E} \left(\sigma_{n+h}^2 \mid \mathcal{F}_n \right), \quad h \geq 1 \\ \mathbf{E} \left(u_{n-h}^2 \mid \mathcal{F}_n \right) &= u_{n-h}^2, \quad h \geq 0 \end{aligned}$$

Para facilitar a notação considere-se $\sigma_{n+h,n}^2 := \mathbf{E} \left(\sigma_{n+h}^2 \mid \mathcal{F}_n \right)$. Vejam-se os exemplos seguintes.

Modelo GARCH(1,1) $\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

$$h = 1 \quad \sigma_{n+1,n}^2 = \dots = \omega + \alpha_1 u_n^2 + \beta_1 \sigma_n^2$$

$$h = 2 \quad \sigma_{n+2,n}^2 = \dots = \omega + (\alpha_1 + \beta_1) \sigma_{n+1,n}^2$$

$$h \text{ geral} \quad \sigma_{n+h,n}^2 = \dots = \omega + (\alpha_1 + \beta_1) \sigma_{n+h-1,n}^2$$

Pode-se provar, sob a condição $0 \leq \alpha_1 + \beta_1 < 1$:

$$\sigma_{n+h,n}^2 = \frac{\omega (1 - (\alpha_1 + \beta_1)^h)}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{h-1} (\alpha_1 u_n^2 + \beta_1 \sigma_n^2).$$

Assim, no caso $\alpha_1 + \beta_1 < 1$, tem-se

$$\sigma_{n+h,n}^2 \rightarrow \frac{\omega}{1 - \alpha_1 - \beta_1} = \text{Var}(u_t) \quad (\text{quando } h \rightarrow \infty).$$

No caso $\alpha_1 + \beta_1 = 1$ (IGARCH(1,1)) vem

$$\begin{aligned} \sigma_{n+h,n}^2 &= \sigma_{n+1,n}^2, & \text{se } \omega &= 0 \\ \sigma_{n+h,n}^2 &= (h-1)\omega + \sigma_{n+1,n}^2, & \text{se } \omega &> 0 \\ \sigma_{n+h,n}^2 &\rightarrow \infty, & \text{se } \omega &> 0 \text{ (quando } h \rightarrow \infty). \end{aligned}$$

3.9.2 A Previsão da Variável Dependente y

A previsão pontual de y_{n+h} não envolve qualquer novidade ...

Todavia, a estimação por intervalos deve agora reflectir a presença de HC.

IC a $(1 - \alpha) 100\%$ para y_{n+h} :

$$E(y_{n+h} | \mathcal{F}_n) \pm q_{1-\alpha/2} \sqrt{\text{Var}(y_{n+h} | \mathcal{F}_n)}$$

Supondo normalidade, o IC a 95% para y_{n+h} é

$$E(y_{n+h} | \mathcal{F}_n) \pm 1.96 \sqrt{\text{Var}(y_{n+h} | \mathcal{F}_n)}.$$

Infelizmente esta expressão só pode ser usada para $h = 1$. Veremos um procedimento de *bootstrap* que permite obter IC correctos para y_{n+h} , com h finito.

Exemplo Considere-se o modelo AR(1)+GARCH(1,1)

$$\begin{aligned} y_t &= c + \phi y_{t-1} + u_t, \\ \sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

Sabendo $u_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$, obtenha então um IC a 95% para y_{n+1} .

Exemplo Obtenha um IC a 95% para r_{n+1} tendo em conta os seguintes resultados: $\hat{\sigma}_n^2 = 0.1$

Dependent Variable: R
 Method: ML - ARCH (Marquardt) - Normal distribution
 Date: 05/03/10 Time: 17:46
 Sample: 6/06/1995 5/02/2010
 Included observations: 3689

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.052406	0.025915	2.022227	0.0432
R(-1)	-0.028276	0.016513	-1.712268	0.0868
@MONTH=1	0.314734	0.068589	4.588681	0.0000
Variance Equation				
C	0.002167	0.005510	0.393239	0.6941
RESID(-1)^2	0.049796	0.005988	8.315851	0.0000
RESID(-1)^2*(RESID(-1)<0)	0.053276	0.008594	6.199160	0.0000
GARCH(-1)	0.918731	0.005656	162.4223	0.0000
HIGH(-1)-LOW(-1)	0.136121	0.027452	4.958580	0.0000

obs	R	RES	HIGH	LOW
29/12/2009	0.57166	0.47560	12.35	12.17
30/12/2009	-0.24460	-0.28084	12.34	12.22
31/12/2009	-0.90202	-0.96134	12.27	12.06
4/01/2010	1.71503	1.32238	12.41	12.31
5/01/2010	-0.73141	-1.05006	12.35	12.19
6/01/2010	-0.65467	-1.04249	12.28	12.16
7/01/2010	-0.74166	-1.12731	12.17	12.06
8/01/2010	1.06953	0.68142	12.25	12.09
11/01/2010	0.48980	0.15290	12.34	12.20
12/01/2010	-0.81767	-1.17096	12.26	12.12
13/01/2010	0.89907	0.50881	12.31	12.10
14/01/2010	-1.39289	-1.73461	12.19	12.05
15/01/2010	-3.35602	-3.76254	11.90	11.58
19/01/2010	-0.25630	-0.71833	11.75	11.55
20/01/2010	-2.95160	-3.32599	11.53	11.28
21/01/2010	-2.31739	-2.76799	11.27	10.95
22/01/2010	-2.92793	-3.36060	11.05	10.73
25/01/2010	2.47605	2.02613	11.18	10.95
26/01/2010	-1.82820	-2.12533	10.97	10.83
27/01/2010	-0.09229	-0.51113	10.88	10.69
28/01/2010	-3.76338	-4.13313	10.78	10.36
29/01/2010	-2.13186	-2.60542	10.46	10.15
1/02/2010	1.74762	1.63493	10.48	10.29
2/02/2010	2.75290	2.74991	10.75	10.58
3/02/2010	-2.27283	-2.24739	10.62	10.37
4/02/2010	-4.80720	-4.92387	10.31	9.95
5/02/2010	1.49629	1.30795	10.12	9.90

3.9.3 Intervalos de Confiança para y e para a Volatilidade baseados em Bootstrap

Vimos até agora as seguintes questões:

- intervalos de confiança para y_{n+1} ;
- previsão de σ_t^2 para os períodos $n + 1, n + 2, \dots$;

Estas questões são relativamente simples tratar. Já a obtenção de intervalos de confiança para y_{n+h} , $h > 1$ e para σ_{n+h}^2 , $h \geq 1$ é problemática, pois não são conhecidas as distribuições de interesse. Estas questões resolvem-se de forma muito satisfatória recorrendo ao *bootstrap*.

Para exemplificar considere-se o modelo

$$\begin{cases} y_t = c + \phi y_{t-1} + u_t \\ u_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2. \end{cases} \quad (1)$$

onde ε tem distribuição desconhecida de média nula e variância um. O algoritmo é o seguinte:

1. Estimar o modelo (1) e obter

$$\{\hat{\varepsilon}_t, t = 1, \dots, n\}, \text{ onde } \hat{\varepsilon}_t = \frac{\hat{u}_t}{\hat{\sigma}_t}$$

$$\hat{\sigma}^2 = \frac{\hat{\omega}}{1 - \hat{\alpha} - \hat{\beta}}, \quad \hat{\mu} = \frac{\hat{c}}{1 - \hat{\phi}}$$

$$\hat{\theta} = (\hat{c}, \hat{\phi}, \hat{\omega}, \hat{\alpha}, \hat{\beta})'$$

2. Simular o modelo

$$\begin{cases} y_t^* = \hat{c} + \hat{\phi}y_{t-1}^* + u_t^* \\ u_t^* = \sigma_t^* \varepsilon_t^* \\ \sigma_t^{*2} = \hat{\omega} + \hat{\alpha}u_{t-1}^{*2} + \hat{\beta}\sigma_{t-1}^{*2} \end{cases} \quad (2)$$

com os seguintes valores iniciais: $\sigma_0^{*2} = \hat{\sigma}^2$ e $y_0^* = \hat{\mu}$. Os valores de ε_t^* são retirados aleatoriamente com reposição do conjunto $\{\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n\}$.

3. Estimar o modelo (2) e obter as seguintes previsões:

$$\begin{cases} y_{n+h}^* = \hat{c}^* + \hat{\phi}^* y_{n+h-1}^* \\ \hat{\sigma}_{n+h}^{*2} = \hat{\omega}^* + \hat{\alpha}^* u_{n+h-1}^{*2} + \hat{\beta}^* \hat{\sigma}_{n+h-1}^{*2} \end{cases}$$

Note-se que $\hat{\theta}^* = (\hat{c}^*, \hat{\phi}^*, \hat{\omega}^*, \hat{\alpha}^*, \hat{\beta}^*)'$ é o vector das estimativas obtidas no contexto do modelo simulado (2).

4. Repetir os passos 2 e 3 B vezes. Com este procedimento obtêm-se as seguintes séries:

$$\begin{aligned} & \left\{ y_{n+j}^{*(1)}, y_{n+j}^{*(2)}, \dots, y_{n+j}^{*(B)} \right\}, \quad j = 1, \dots, h, \\ & \left\{ \hat{\sigma}_{n+j}^{*2(1)}, \hat{\sigma}_{n+j}^{*2(2)}, \dots, \hat{\sigma}_{n+j}^{*2(B)} \right\}, \quad j = 1, \dots, h. \end{aligned}$$

5. Um intervalo de previsão a $(1 - \alpha) 100\%$ para y_{n+j} é

$$\left[q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}} \right]$$

onde $q_{\frac{\alpha}{2}}$ e $q_{1-\frac{\alpha}{2}}$ são os quantis empíricos da amostra $\left\{ y_{n+j}^{*(1)}, y_{n+j}^{*(2)}, \dots, y_{n+j}^{*(B)} \right\}$.

6. Um intervalo de previsão a $(1 - \alpha) 100\%$ para σ_{n+j}^2 é

$$\left[q_{\frac{\alpha}{2}}^{\sigma^{*2}}, q_{1-\frac{\alpha}{2}}^{\sigma^{*2}} \right]$$

onde agora $q_{\frac{\alpha}{2}}$ e $q_{1-\frac{\alpha}{2}}$ são os quantis empíricos da amostra $\left\{ \hat{\sigma}_{n+j}^{*2(1)}, \hat{\sigma}_{n+j}^{*2(2)}, \dots, \hat{\sigma}_{n+j}^{*2(B)} \right\}$.

4 Modelação da Heterocedasticidade Condicionada: Caso Multivariado

4.1 Introdução

Estimação multivariada:

- relevante na selecção de *portfolios*, da gestão do risco, etc.
- estimação dos co-movimentos de rendibilidade e volatilidade. Por exemplo, como se transmite a volatilidade de um mercado aos demais mercados? qual a magnitude do impacto da volatilidade de um mercado sobre outro?

Para tratar estas questões vai considerar-se um modelo genérico, envolvendo m equações:

$$\begin{aligned}y_{1t} &= \mu_{1t} + u_{1t}, \\ &\dots \\ y_{mt} &= \mu_{mt} + u_{mt}\end{aligned}$$

Notação mais compacta:

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{pmatrix}, \quad \boldsymbol{\mu}_t = \begin{pmatrix} \mu_{1t} \\ \vdots \\ \mu_{mt} \end{pmatrix}, \quad \mathbf{u}_t = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{mt} \end{pmatrix}.$$

Assim

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{u}_t.$$

Modelo para μ_t . Por exemplo VAR(1), $m = 2$:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \underbrace{\begin{pmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix}}_{\boldsymbol{\mu}_t} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

Modelo de heterocedasticidade condicional multivariado:

$$\mathbf{u}_t = \mathbf{H}_t^{1/2} \boldsymbol{\varepsilon}_t$$

- $\boldsymbol{\varepsilon}_t$ vector de v.a. i.i.d. (condicionalmente homocedástico) tal que

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}, \quad \text{Var}(\boldsymbol{\varepsilon}_t) = \mathbf{I}_m.$$

- \mathbf{H}_t é uma matriz quadrada de ordem m , simétrica, definida positiva e F_{t-1} mensurável.

$\mathbf{H}_t^{1/2}$ é uma matriz quadrada ordem m tal que $\mathbf{H}_t^{1/2} \left(\mathbf{H}_t^{1/2} \right)' = \mathbf{H}_t$.

Dadas as hipóteses, tem-se

$$\text{Var}(\mathbf{y}_t | F_{t-1}) = \dots = \mathbf{H}_t$$

4.2 Densidade e Verosimilhança

Assuma-se normalidade dos erros: $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}_m)$. Tem-se:

$$\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}_m) \Rightarrow \mathbf{u}_t | F_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t) \Rightarrow \mathbf{y}_t | F_{t-1} \sim N(\boldsymbol{\mu}_t, \mathbf{H}_t).$$

Assim, a densidade conjunta condicional de \mathbf{y}_t é

$$f(\mathbf{y}_t | F_{t-1}) = (2\pi)^{-m/2} |\mathbf{H}_t|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_t - \boldsymbol{\mu}_t)' \mathbf{H}_t^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t) \right\}.$$

A função log-verosimilhança é então

$$\begin{aligned} \log L_n(\boldsymbol{\theta}) &= \sum_{t=1}^n \log f(\mathbf{y}_t | F_{t-1}) \\ &= \dots = -\frac{nm}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log |\mathbf{H}_t(\boldsymbol{\theta})| \\ &\quad - \frac{1}{2} \sum_{t=1}^n (\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta}))' \mathbf{H}_t^{-1}(\boldsymbol{\theta}) (\mathbf{y}_t - \boldsymbol{\mu}_t(\boldsymbol{\theta})). \end{aligned}$$

É necessário definir uma hipótese sobre a estrutura de $\boldsymbol{\mu}_t$ e de \mathbf{H}_t .

4.3 Modelo VECH (ou VEC)

Modelo VECH (ou VEC) GARCH(1,1):

$$\text{vech}(\mathbf{H}_t) = \mathbf{w} + \mathbf{A} \text{vech}(\mathbf{u}_{t-1} \mathbf{u}'_{t-1}) + \mathbf{B} \text{vech}(\mathbf{H}_{t-1}).$$

No caso $m = 2$ (processo bivariado) e GARCH(1,1):

$$\begin{aligned} \text{vech}(\mathbf{H}_t) &= \begin{pmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} w_{11} \\ w_{12} \\ w_{22} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \begin{pmatrix} u_{1,t-1}^2 \\ u_{1,t-1} u_{2,t-1} \\ u_{2,t-1}^2 \end{pmatrix} \\ &+ \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix}. \end{aligned}$$

Por exemplo,

$$\begin{aligned} h_{12,t} &= \mathbf{E}(u_{1t} u_{2t} | F_{t-1}) = w_{12} + \alpha_{21} u_{1,t-1}^2 + \alpha_{23} u_{2,t-1}^2 \\ &+ \alpha_{22} u_{1,t-1} u_{2,t-1} + \beta_{21} h_{11,t-1} + \beta_{22} h_{12,t-1} + \beta_{23} h_{22,t-1} \end{aligned}$$

Correlações condicionais:

$$\rho_{ij,t} = \frac{h_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t}}}, \quad i, j = 1, \dots, m, \quad t = 1, 2, \dots, n$$

$\{\mathbf{u}_t\}$ é ESO se todos os valores próprios de $\mathbf{A} + \mathbf{B}$ forem em módulo menores do que um.

Nestas condições:

$$E(\text{vech}(\mathbf{u}_t \mathbf{u}_t')) = E(\text{vech}(\mathbf{H}_t)) = (\mathbf{I} - \mathbf{A} - \mathbf{B})^{-1} \mathbf{w}.$$

Vantagem do modelo VEC: grande flexibilidade, todos os elementos de \mathbf{H}_t dependam de todos os produtos cruzados de $\text{vech}(\mathbf{u}_{t-1}\mathbf{u}'_{t-1})$ e de todos os elementos de \mathbf{H}_{t-1} .

Desvantagens (superam largamente as suas vantagens):

- O número de parâmetros a estimar é excessivamente alto. No GARCH(1,1) multivariado com m equações, o número de parâmetros a estimar é

m	$(m(m+1)/2)(1+(m(m+1)))$
2	21
3	78
4	210

- Por definição a matriz \mathbf{H}_t deve ser definida positiva, mas não é fácil garantir isso a partir das matrizes \mathbf{A} e \mathbf{B} . Se \mathbf{H}_t não é definida positiva, é possível, por exemplo, que $\rho_{ij,t} > 1$ ou $h_{t,ii} < 0$.

Exemplo Existem efeitos de rendimento e de volatilidade do PSI20 que possam ser antecipados através do Dow Jones (DJ)? Seja y_{1t} e y_{2t} o retorno diário associado, respectivamente, aos índices, PSI20 e DJ. As variáveis y_{1t} e y_{2t} foram previamente centradas (razão?). Depois de vários ensaios, definiu-se o seguinte modelo

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \mathbf{H}_t^{1/2} \boldsymbol{\varepsilon}_t$$

onde

$$\begin{aligned} \text{vech}(\mathbf{H}_t) &= \begin{pmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ w_{22} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & 0 & \alpha_{13} \\ 0 & 0 & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix} \begin{pmatrix} u_{1,t-1}^2 \\ u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}^2 \end{pmatrix} \\ &+ \begin{pmatrix} \beta_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix} + \begin{pmatrix} \gamma vol_{t-1} \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Usando os dados no período 31/12/92 a 15/03/99 (1496 observações) obteve-se,

$$\hat{y}_{1t} = \underset{(.028)}{.2343}y_{1t-1} + \underset{(.023)}{.1430}y_{2t-1}, \quad \hat{y}_{2t} = \underset{(.023)}{.0753}y_{2t-1}$$

$$\hat{h}_{11,t} = \underset{(.0466)}{.3132}\hat{u}_{1t-1}^2 + \underset{(.0151)}{.0466}\hat{u}_{2t-1}^2 + \underset{(.0459)}{.6053}\hat{h}_{1t-1} + \underset{(.0062)}{.0254}vol_{t-1}$$

$$\hat{h}_{22,t} = \underset{(5.2 \times 10^{-7})}{1.25 \times 10^{-6}} + \underset{(.0195)}{.0903}\hat{u}_{2t-1}^2 + \underset{(.0227)}{.897}\hat{h}_{2t-1}, \quad h_{12,t} = 0.$$

Escreva $\hat{\mathbf{A}}$ e $\hat{\mathbf{B}}$ e analise os resultados.

4.4 Modelo Diagonal VECH

Podem obter-se modelos VECH com menos parâmetros impondo que as matrizes \mathbf{A} e \mathbf{B} sejam diagonais. Por exemplo, no caso $m = 2$, vem

$$\begin{aligned} \text{vech}(\mathbf{H}_t) &= \begin{pmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} w_{11} \\ w_{12} \\ w_{22} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix} \begin{pmatrix} u_{1,t-1}^2 \\ u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}^2 \end{pmatrix} \\ &+ \begin{pmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix} \end{aligned}$$

(este princípio aplica-se naturalmente no caso de modelos multivariados GARCH(p,q)).

Com matrizes \mathbf{A} e \mathbf{B} diagonais pode optar por escrever o modelo *diagonal VECH* na forma equivalente

$$\mathbf{H}_t = \boldsymbol{\omega} + \mathbf{a} \circ \mathbf{u}_{t-1} \mathbf{u}'_{t-1} + \mathbf{b} \circ \mathbf{H}_{t-1}$$

onde $\boldsymbol{\omega}$, \mathbf{a} e \mathbf{b} são matrizes simétricas de tipo $m \times m$ e “ \circ ” é o produto de Hadamard . Por exemplo, no caso $m = 2$, o modelo anterior escreve-se

$$\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \circ \begin{pmatrix} u_{1,t-1}^2 & u_{1,t-1}u_{2,t-1} \\ u_{1,t-1}u_{2,t-1} & u_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix} \circ \begin{pmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{pmatrix}$$

onde $a_{11} = \alpha_{11}$, $a_{12} = \alpha_{22}$, $a_{22} = \alpha_{33}$, etc. Note-se, portanto, que

$$\begin{aligned} h_{11,t} &= \omega_{11} + a_{11}u_{1,t-1}^2 + b_{11}h_{11,t-1} \\ h_{12,t} &= \omega_{12} + a_{12}u_{1,t-1}u_{2,t-1} + b_{12}h_{12,t-1} \\ h_{22,t} &= \omega_{22} + a_{22}u_{2,t-1}^2 + b_{22}h_{22,t-1}. \end{aligned}$$

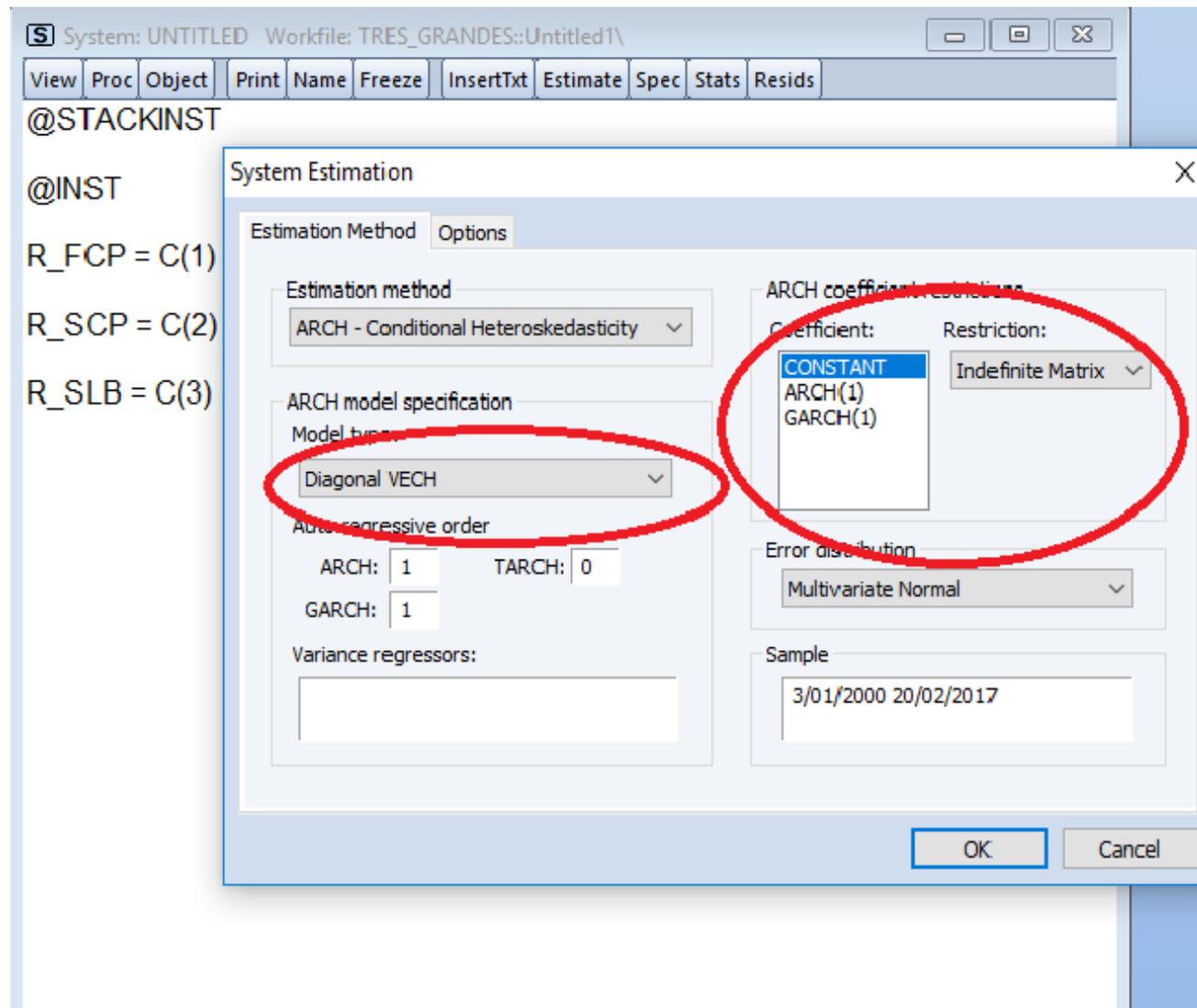
Vantagem do modelo:

- Menos parâmetros a estimar. Passam a ser $3m(m + 1) / 2$.

Desvantagens:

- Considere-se o caso $m = 2$ e GARCH(1,1). $h_{ii,t}$ só depende dos termos $u_{i,t-1}^2$ e $h_{ii,t-1}$, e $h_{12,t}$ só depende dos termos $u_{1,t-1}u_{2,t-1}$ e $h_{12,t-1}$. A especificação Diagonal VECH elimina a possibilidade de interação entre as diferentes variâncias e covariâncias condicionais.
- Por outro lado, a matriz \mathbf{H}_t , por construção, não resulta definida positiva.

No EVIEWS este modelo é estimado com a opção Diagonal VECH + Indefinite Matrix



Há várias formas de ultrapassar o problema de \mathbf{H}_t poder não ser definida positiva. Por exemplo:

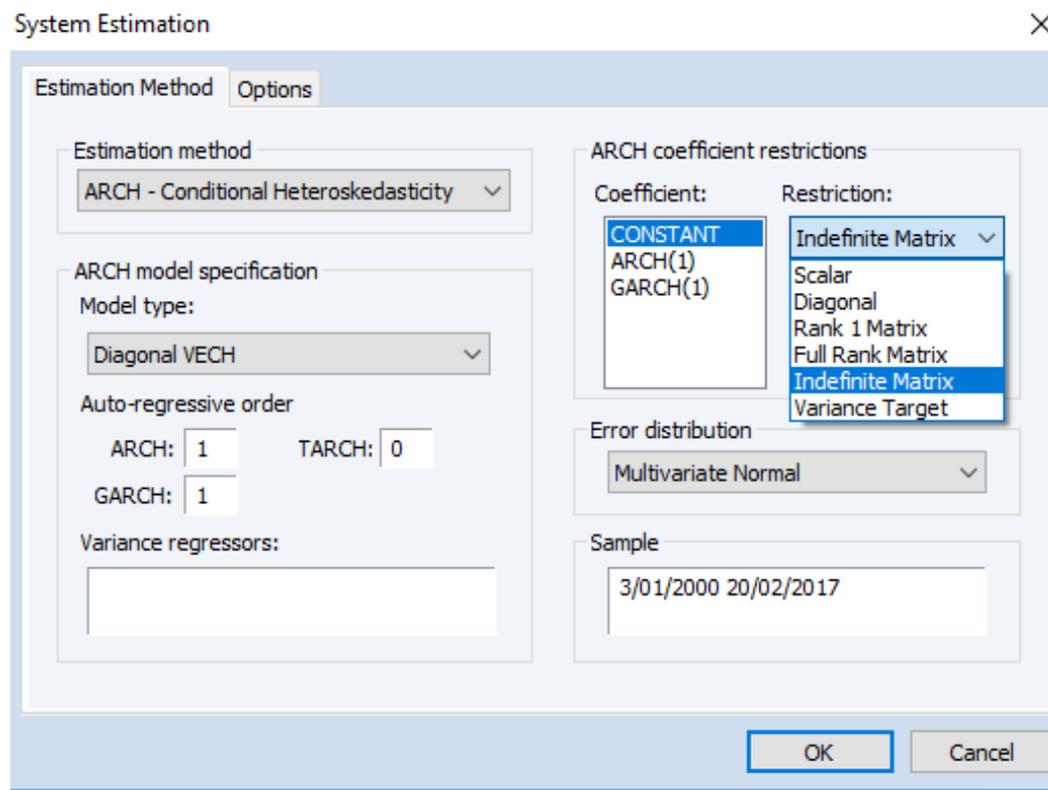
$$\mathbf{H}_t = \boldsymbol{\omega} + \mathbf{a} \circ \mathbf{u}_{t-1} \mathbf{u}'_{t-1} + \mathbf{b} \circ \mathbf{H}_{t-1}$$

mas agora impõe-se:

- $\boldsymbol{\omega} = \tilde{\boldsymbol{\omega}} \tilde{\boldsymbol{\omega}}'$
- $\mathbf{a} = \tilde{\mathbf{a}} \tilde{\mathbf{a}}'$
- $\mathbf{b} = \tilde{\mathbf{b}} \tilde{\mathbf{b}}'$

$\tilde{\boldsymbol{\omega}}$, $\tilde{\mathbf{a}}$ e $\tilde{\mathbf{b}}$ são matrizes quadradas de ordem m . As matrizes $\boldsymbol{\omega}$, \mathbf{a} e \mathbf{b} assim construídas implicam uma matriz \mathbf{H}_t definida positiva.

No EVIEWS as matrizes $\tilde{\boldsymbol{\omega}}$, $\tilde{\mathbf{a}}$ e $\tilde{\mathbf{b}}$ podem estar sujeitas a restrições com vista a diminuir ainda mais o número de parâmetros a estimar



O modelo mais restritivo:

$$\begin{pmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{pmatrix} = \begin{pmatrix} 1 - \lambda & 0 & 0 \\ 0 & 1 - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{pmatrix} \begin{pmatrix} u_{1,t-1}^2 \\ u_{1,t-1}u_{2,t-1} \\ u_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix}.$$

Passamos a ter apenas um parâmetro a estimar qualquer que seja o número de equações do modelo.

4.5 Modelo “Triangular”

4.5.1 Introdução e Formalização do Modelo

Seja $F_t^{y_i}$ o conjunto de informação de y_{it} . Em certas aplicações é admissível supor:

$$\begin{aligned} E\left(y_{1t} \mid F_{t-1}^{y_1} \cup F_{t-1}^{y_2}\right) &= E\left(y_{1t} \mid F_{t-1}^{y_1}\right) \\ \text{Var}\left(y_{1t} \mid F_{t-1}^{y_1} \cup F_{t-1}^{y_2}\right) &= \text{Var}\left(y_{1t} \mid F_{t-1}^{y_1}\right) \end{aligned} \quad (*)$$

A equação (*) implica que y_2 não causa à Granger y_1 .

Estas relações são possíveis, por exemplo, se y_{1t} é o retorno do NASDAQ e y_{2t} é o retorno do PSI20.

Para processos y_1 e y_2 com as características acima descritas, é possível definirem-se processos multivariados simplificados.

Considere-se o processo $\mathbf{y} = (y_1, y_2, y_3)$ e suponham-se as seguintes relações: $y_1 \rightsquigarrow y_2 \rightsquigarrow y_3$ onde “ $y_1 \rightsquigarrow y_2$ ” significa y_1 influencia y_2 dado $F_{t-1}^{y_2}$ e y_2 não influencia y_1 dado $F_{t-1}^{y_1}$.

Suponha-se $\mathbf{y} \sim \text{VAR}(1)$. Tem-se:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} \phi_{11} & 0 & 0 \\ \phi_{21} & \phi_{22} & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}.$$

A matriz dos coeficientes autoregressivos é triangular, porque na média condicional $y_{1,t}$ apenas depende de $y_{1,t-1}$, y_{2t} depende de $y_{1,t-1}$ e $y_{2,t-1}$ e y_{3t} depende de $y_{1,t-1}$, $y_{2,t-1}$ e $y_{3,t-1}$.

Estrutura de dependências do segundo momento condicional:

$$\begin{cases} u_{1t} = e_{1t} \\ u_{2t} = ae_{1t} + e_{2t} \\ u_{3t} = be_{1t} + ce_{2t} + e_{3t} \end{cases} \Leftrightarrow \underbrace{\begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}}_{\mathbf{u}_t} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}}_{\Psi} \underbrace{\begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix}}_{\mathbf{e}_t}$$

onde:

- (e_{1t}, e_{2t}, e_{3t}) são independentes entre si;
- $e_{it} | F_{t-1} \sim N(0, \sigma_{it}^2)$, $\sigma_{it}^2 = \omega_i + \alpha_i e_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2$.

u_{2t} depende de e_{2t} (efeitos idiossincrásicos) e ainda dos choques e_{1t} (*volatility spillover*).

u_{3t} depende de e_{3t} (efeitos idiossincrásicos) e ainda dos choques e_{1t} e e_{2t} .

A designação “modelo triangular” é agora óbvia:

$$\mathbf{y}_t = \mathbf{c} + \Phi \mathbf{y}_{t-1} + \Psi \mathbf{e}_t$$

sendo Φ e Ψ matrizes triangulares inferiores.

Dadas as hipótese sobre o vector \mathbf{e}_t , defina-se

$$\Sigma_t := \text{Var}(\mathbf{e}_t | F_{t-1}) = \begin{pmatrix} \sigma_{1,t}^2 & 0 & 0 \\ 0 & \sigma_{2,t}^2 & 0 \\ 0 & 0 & \sigma_{3,t}^2 \end{pmatrix}.$$

Logo,

$$\mathbf{H}_t = \text{Var}(\mathbf{u}_t | F_{t-1}) = \text{Var}(\Psi \mathbf{e}_t | F_{t-1}) = \Psi \text{Var}(\mathbf{e}_t | F_{t-1}) \Psi' = \Psi \Sigma_t \Psi'.$$

Pode-se provar que a condição de ESO do processo $\{\mathbf{u}_t\}$ é $\alpha_i + \beta_i < 1$, $i = 1, 2, 3$.

\mathbf{H}_t como função dos termos $\mathbf{u}_{t-1}\mathbf{u}'_{t-1}$ e \mathbf{H}_{t-1}

$$\mathbf{H}_t = \dots = \Psi \mathbf{W} \Psi' + \Psi \left(\mathbf{A} \circ \Psi^{-1} \mathbf{u}_{t-1} \mathbf{u}'_{t-1} (\Psi^{-1})' \right) \Psi' + \Psi \left(\mathbf{B} \circ \Psi^{-1} \mathbf{H}_{t-1} (\Psi^{-1})' \right) \Psi'.$$

Resulta, por exemplo, que

$$\begin{aligned} h_{22,t} &= a^2 \omega_1 + \omega_2 + a^2 (\alpha_1 + \alpha_2) u_{1,t-1}^2 \\ &\quad - 2a\alpha_2 u_{1,t-1} u_{2,t-1} + a^2 (\beta_1 + \beta_2) h_{11,t-1} - 2a\beta_2 h_{12,t-1} + \beta_2 h_{22,t-1} \end{aligned}$$

\mathbf{H}_t como função de $\sigma_{1,t}^2$, $\sigma_{2,t}^2$ e $\sigma_{3,t}^2$

$$\begin{aligned}
 \mathbf{H}_t &= \mathbf{\Psi} \mathbf{\Sigma}_t \mathbf{\Psi}' \\
 &= \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix} \begin{pmatrix} \sigma_{1,t}^2 & 0 & 0 \\ 0 & \sigma_{2,t}^2 & 0 \\ 0 & 0 & \sigma_{3,t}^2 \end{pmatrix} \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} \sigma_{1,t}^2 & a\sigma_{1,t}^2 & b\sigma_{1,t}^2 \\ a\sigma_{1,t}^2 & a^2\sigma_{1,t}^2 + \sigma_{2,t}^2 & ab\sigma_{1,t}^2 + c\sigma_{2,t}^2 \\ b\sigma_{1,t}^2 & ab\sigma_{1,t}^2 + c\sigma_{2,t}^2 & b^2\sigma_{1,t}^2 + c^2\sigma_{2,t}^2 + \sigma_{3,t}^2 \end{pmatrix}.
 \end{aligned}$$

Resulta:

$$\begin{aligned}
 h_{11,t} &= \text{Var}(y_{1t} | F_{t-1}) = \text{Var}(u_{1t} | F_{t-1}) = \sigma_{1,t}^2 \\
 h_{22,t} &= \text{Var}(y_{2t} | F_{t-1}) = \text{Var}(u_{2t} | F_{t-1}) = a^2\sigma_{1,t}^2 + \sigma_{2,t}^2 \\
 h_{33,t} &= \text{Var}(y_{3t} | F_{t-1}) = \text{Var}(u_{3t} | F_{t-1}) = b^2\sigma_{1,t}^2 + c^2\sigma_{2,t}^2 + \sigma_{3,t}^2 \\
 h_{12,t} &= \text{Cov}(y_{1t}, y_{2t} | F_{t-1}) = a\sigma_{1,t}^2 \\
 h_{13,t} &= \text{Cov}(y_{1t}, y_{3t} | F_{t-1}) = b\sigma_{1,t}^2 \\
 h_{23,t} &= \text{Cov}(y_{2t}, y_{3t} | F_{t-1}) = ab\sigma_{1,t}^2 + c\sigma_{2,t}^2
 \end{aligned}$$

Coefficientes de correlação condicionados:

$$\rho_{12,t} = \frac{a\sigma_{1,t}^2}{\sqrt{\sigma_{1,t}^2}\sqrt{a^2\sigma_{1,t}^2 + \sigma_{2,t}^2}} = \frac{a\sigma_{1,t}}{\sqrt{a^2\sigma_{1,t}^2 + \sigma_{2,t}^2}}$$

$$\rho_{13,t} = \frac{b\sigma_{1,t}^2}{\sqrt{\sigma_{1,t}^2}\sqrt{b^2\sigma_{1,t}^2 + c^2\sigma_{2,t}^2 + \sigma_{3,t}^2}} = \frac{b\sigma_{1,t}}{\sqrt{b^2\sigma_{1,t}^2 + c^2\sigma_{2,t}^2 + \sigma_{3,t}^2}}$$

$$\rho_{23,t} = \frac{ab\sigma_{1,t}^2 + c\sigma_{2,t}^2}{\sqrt{a^2\sigma_{1,t}^2 + \sigma_{2,t}^2}\sqrt{b^2\sigma_{1,t}^2 + c^2\sigma_{2,t}^2 + \sigma_{3,t}^2}}.$$

Os sinais dos coeficientes a , b e c são decisivos nos sinais dos coeficientes de correlação condicionados.

Rácios de Variância e Transmissão da Volatilidade ao longo do Tempo

Análise do mercado 2

Tendo em conta

$$\frac{a^2\sigma_{1,t}^2 + \sigma_{2,t}^2}{h_{22,t}} = 1$$

tem-se

$$RV_t^{1,2} = \frac{a^2\sigma_{1,t}^2}{h_{22,t}},$$
$$RV_t^{2,2} = \frac{\sigma_{2,t}^2}{h_{22,t}} = 1 - \frac{a^2\sigma_{1,t}^2}{h_{22,t}}$$

$RV_t^{1,2}$ a proporção da variância do mercado 2 que é causada pelo efeito de *volatility spillover* do mercado 1 (efeito do mercado 1 para 2).

Análise do mercado 3

Tendo em conta

$$\frac{b^2\sigma_{1,t}^2 + c^2\sigma_{2,t}^2 + \sigma_{3,t}^2}{h_{33,t}} = 1$$

tem-se

$$RV_t^{1,3} = \frac{b^2\sigma_{1t}^2}{h_{33,t}},$$
$$RV_t^{2,3} = \frac{c^2\sigma_{2t}^2}{h_{33,t}},$$
$$RV_t^{3,3} = 1 - \frac{b^2\sigma_{1t}^2}{h_{33,t}} - \frac{c^2\sigma_{2t}^2}{h_{33,t}}.$$

4.5.2 Estimação

Considere a representação $\mathbf{y}_t = \mathbf{c} + \Phi \mathbf{y}_{t-1} + \Psi \mathbf{e}_t$, isto é,

$$y_{1t} = c_1 + \phi_{11}y_{1,t-1} + e_{1t} \quad (1)$$

$$y_{2t} = c_2 + \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + ae_{1t} + e_{2t} \quad (2)$$

$$y_{3t} = c_3 + \phi_{31}y_{1,t-1} + \phi_{32}y_{2,t-1} + \phi_{33}y_{3,t-1} + be_{1t} + ce_{2t} + e_{3t} \quad (3)$$

onde $e_{it} | F_{t-1} \sim N(0, \sigma_{it}^2)$, $\sigma_{it}^2 = \omega_i + \alpha_i e_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2$.

Passos:

1. Estimar a eq. (1), pelo método da máxima verosimilhança, e obter os resíduos $\{\hat{e}_{1t}\}$.
2. Substituir, na eq. (2), e_{1t} por \hat{e}_{1t} e estimar o modelo. Obter os resíduos $\{\hat{e}_{2t}\}$.
3. Substituir, na eq. (3), e_{1t} por \hat{e}_{1t} e e_{2t} por \hat{e}_{2t} e estimar o modelo.

4.5.3 Exemplo

Pinto (2010) analisou a transmissão de volatilidade do mercado Norte-Americano (US) para o mercado Europeu (EU) e, em particular, as repercussões destes dois mercados no mercado Português (PT), através de um modelo triangular. Período: 4 de Janeiro de 1993 - 4 de Setembro de 2009.

r_{1t} - retorno do SP500, r_{2t} - retorno do DJ Euro 50, r_{3t} - retorno do PSI 20.

Passo 1

Dependent Variable: R1
 Method: ML - ARCH
 Sample (adjusted): 6/01/1993 4/09/2009
 Included observations: 4055 after adjustments

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.052202	0.013247	3.940658	0.0001
R1(-1)	-0.017161	0.017808	-0.963644	0.3352
Variance Equation				
C	0.007049	0.001077	6.546092	0.0000
RESID(-1)^2	0.066257	0.004820	13.74617	0.0000
GARCH(-1)	0.929667	0.005114	181.7812	0.0000

Passo 2

Dependent Variable: R2
 Method: ML - ARCH
 Sample (adjusted): 6/01/1993 4/09/2009
 Included observations: 4055 after adjustments

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.052704	0.013123	4.016235	0.0001
R1(-1)	0.449084	0.019369	23.18547	0.0000
R2(-1)	-0.179191	0.015456	-11.59382	0.0000
RES1	0.552914	0.015972	34.61746	0.0000

Variance Equation				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.009916	0.002944	3.367632	0.0008
RESID(-1)^2	0.068232	0.010502	6.497201	0.0000
GARCH(-1)	0.923938	0.010928	84.55149	0.0000

Passo 3

Dependent Variable: R3
 Method: ML - ARCH
 Sample (adjusted): 6/01/1993 4/09/2009
 Included observations: 4055 after adjustments

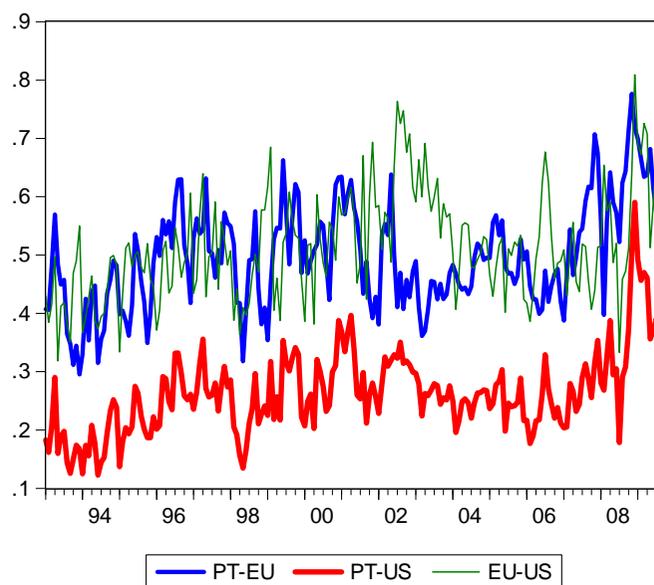
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.043610	0.010837	4.024173	0.0001
R1(-1)	0.222423	0.010966	20.28223	0.0000
R2(-1)	-0.093210	0.012230	-7.621611	0.0000
R3(-1)	0.143101	0.016376	8.738563	0.0000
RES1	0.233364	0.007759	30.07544	0.0000
RES2	0.406663	0.008429	48.24352	0.0000

Variance Equation				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.027709	0.002862	9.681126	0.0000
RESID(-1)^2	0.158430	0.009900	16.00381	0.0000
GARCH(-1)	0.815877	0.010650	76.60615	0.0000

Discutir:

Efeitos de *volatility spillover* do mercado US para o mercado EU?

Efeitos de *volatility spillover* dos mercado US e EU para o mercado PT?



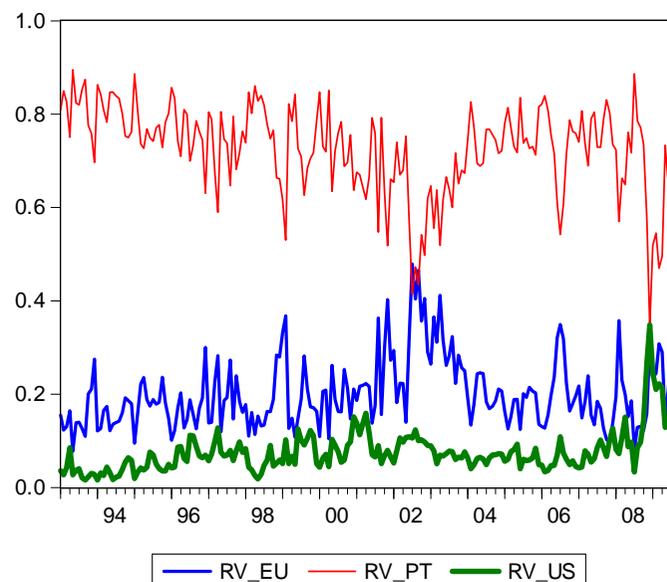
Coeficientes de Correlação Condicional

Transmissão da volatilidade dos mercados EU e US para o mercado PT através dos rácios de volatilidade. São considerados os seguintes rácios

$$RV_US_t = RV_t^{1,3} = \frac{\hat{b}^2 \hat{\sigma}_{1t}^2}{\hat{h}_{33,t}},$$

$$RV_EU_t = RV_t^{2,3} = \frac{\hat{c}^2 \hat{\sigma}_{2t}^2}{\hat{h}_{33,t}},$$

$$RV_PT_t = 1 - RV_US_t - RV_EU_t.$$



Rácios de Volatilidade

4.6 Testes de Diagnóstico

A hipótese de partida é $\mathbf{u}_t = \mathbf{H}_t^{1/2} \boldsymbol{\varepsilon}_t$ onde $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{I}_m)$. Se o modelo estiver correctamente especificado, $\{\boldsymbol{\varepsilon}_t\}$ deve ser uma sucessão de vectores i.i.d., com matriz de variâncias-covariâncias (contemporânea) dada por \mathbf{I}_m . Naturalmente que $\boldsymbol{\varepsilon}_t$ é desconhecido, mas pode ser estimado da seguinte forma

$$\hat{\boldsymbol{\varepsilon}}_t = \hat{\mathbf{H}}_t^{-1/2} \hat{\mathbf{u}}_t.$$

$\hat{\boldsymbol{\varepsilon}}_t$ é o vector dos resíduos standardizados (e $\hat{\mathbf{u}}_t$ é o vector dos resíduos). A matriz $\hat{\mathbf{H}}_t^{-1/2}$ pode obter-se a partir da decomposição Cholesky. Por exemplo, considere-se um sistema de duas equações ($m = 2$)

$$\mathbf{H}_t = \begin{bmatrix} \sigma_{1t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2t}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{1t}^2 & \rho_t \sigma_{1t} \sigma_{2t} \\ \rho_t \sigma_{1t} \sigma_{2t} & \sigma_{2t}^2 \end{bmatrix}.$$

A decomposição de Cholesky fornece

$$\hat{\mathbf{H}}_t^{1/2} = \begin{bmatrix} \sigma_{1t} & 0 \\ \rho_t \sigma_{2t} & \sigma_{2t} \sqrt{1 - \rho_t^2} \end{bmatrix}.$$

Assim,

$$\hat{\mathbf{H}}_t^{-1/2} = \begin{bmatrix} \frac{1}{\sigma_{1t}} & 0 \\ -\frac{\rho_t}{\sigma_{1t}\sqrt{1-\rho_t^2}} & \frac{1}{\sigma_{2t}\sqrt{1-\rho_t^2}} \end{bmatrix}.$$

e

$$\begin{bmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \end{bmatrix} = \begin{bmatrix} \frac{1}{\hat{\sigma}_{1t}} & 0 \\ -\frac{\hat{\rho}_t}{\hat{\sigma}_{1t}\sqrt{1-\hat{\rho}_t^2}} & \frac{1}{\hat{\sigma}_{2t}\sqrt{1-\hat{\rho}_t^2}} \end{bmatrix} \begin{bmatrix} \hat{u}_{1t} \\ \hat{u}_{2t} \end{bmatrix} = \begin{bmatrix} \frac{\hat{u}_{1t}}{\hat{\sigma}_{1t}} \\ \frac{\hat{u}_{2t}}{\hat{\sigma}_{2t}\sqrt{1-\hat{\rho}_t^2}} - \frac{\hat{u}_{1t}\hat{\rho}_t}{\hat{\sigma}_{1t}\sqrt{1-\hat{\rho}_t^2}} \end{bmatrix}.$$

Para avaliar se os efeitos de heterocedasticidades estão convenientemente modelados:

Primeiro passo: regressão de $\hat{\varepsilon}_{1t}^2$ sobre as seguintes variáveis (para além de um termo constante):

- resíduos quadráticos $\hat{\varepsilon}_{i,t-k}^2$, com $i = 1, \dots, m$ e $k = 1, \dots, L$ (L defasamentos) e
- termos cruzados $\hat{\varepsilon}_{i,t-k}\hat{\varepsilon}_{j,t-k}$, com $i, j = 1, \dots, m$ e $k = 1, \dots, L$.

Segundo passo: teste F de nulidade de todos os parâmetros com exceção do do termo independente. Se existir evidência estatística contra a hipótese nula, podemos suspeitar que a matriz \mathbf{H}_t não foi convenientemente modelada.

Nos passos seguintes repete-se o procedimento, tomando sucessivamente $\hat{\varepsilon}_{i,t}^2$ $i = 2, \dots, m$ como variável dependente na regressão auxiliar.

4.7 Aplicação Seleccção de Portfolios

4.7.1 Abordagem Marginal

Problema da determinação dos pesos óptimos de uma carteira constituída por m activos com risco. Considere-se um portfolio constituído pelas acções A1, A2 e A3. Vai ser tomada uma decisão de investimento no valor de 1 milhão de Euros. A questão é: quanto comprar de cada um desses activos? Suponha-se

$$\begin{bmatrix} r_{1t} \\ r_{2t} \\ r_{3t} \end{bmatrix} \sim N(\boldsymbol{\mu}, \mathbf{H})$$

onde

$$\boldsymbol{\mu} = \begin{pmatrix} 5.22 \\ 6.05 \\ 5.98 \end{pmatrix} \quad \mathbf{H} = \begin{pmatrix} 1.549 & 0.9534 & 0.465 \\ 0.9534 & 2.093 & 1.0416 \\ 0.465 & 1.0416 & 1.265 \end{pmatrix}.$$

As rendibilidades estão anualizadas e em percentagem.

Retorno do *portfolio*:

$$R_{pt} = \sum_{i=1}^m \omega_i R_{it} = \boldsymbol{\omega}' \mathbf{r}_t$$

Retorno esperado do *portfolio*

$$\mathbb{E}(R_{pt}) = \mathbb{E}(\boldsymbol{\omega}' \mathbf{r}_t) = \boldsymbol{\omega}' \boldsymbol{\mu}$$

Variância do *portfolio*

$$\begin{aligned} \text{Var}(R_{pt}) &= \text{Var}(\boldsymbol{\omega}' \mathbf{r}_t) \\ &= \boldsymbol{\omega}' \text{Var}(\mathbf{r}_t) \boldsymbol{\omega} = \boldsymbol{\omega}' \mathbf{H} \boldsymbol{\omega}. \end{aligned}$$

Por exemplo, suponha-se que o valor do *portfolio* de um investidor é de 1 milhão de Euros e está assim distribuída: 176000 Euros em A1, 104000 em A2 e 720000 em A3. Assim,

$$\omega = \begin{bmatrix} 0.176 \\ 0.104 \\ 0.720 \end{bmatrix}.$$

Logo

$$E(R_{pt}) = \omega' \mu = \begin{pmatrix} 0.176 & 0.104 & 0.72 \end{pmatrix} \begin{pmatrix} 5.22 \\ 6.05 \\ 5.98 \end{pmatrix} = 5.85$$

$$\begin{aligned} \text{Var}(R_{pt}) &= \omega' H \omega \\ &= \begin{pmatrix} 0.176 & 0.104 & 0.72 \end{pmatrix} \begin{pmatrix} 1.549 & 0.9534 & 0.465 \\ 0.9534 & 2.093 & 1.0416 \\ 0.465 & 1.0416 & 1.265 \end{pmatrix} \begin{pmatrix} 0.176 \\ 0.104 \\ 0.720 \end{pmatrix} \\ &= 1.03 \end{aligned}$$

Para outro valor ω resulta outro valor de $E(R_{pt})$ e outro de $\text{Var}(R_{pt})$.

Seleccção de Portfolios: problema da determinação dos pesos óptimos de uma carteira constituída por m activos com risco:

$$\begin{cases} \min_{\omega} \text{Var} (R_{p,n+1}) \\ \text{s.a } \mathbf{E} (R_{p,n+1}) = \mu_p \text{ e } \sum_{i=1}^m \omega_i = 1 \end{cases}$$

4.7.2 Abordagem Condicional

Considere-se vector do valor esperado condic. dos retornos:

$$\boldsymbol{\mu}_t = (\mathbf{E} (R_{1t} | F_{t-1}), \dots, \mathbf{E} (R_{mt} | F_{t-1}))';$$

Matriz das variâncias-covariâncias condicionais dos retornos:

$$\text{Var} (\mathbf{r}_t | F_{t-1}) = \mathbf{H}_t;$$

Vector dos pesos do *portfolio* no momento t :

$$\boldsymbol{\omega}_t = (\omega_{1t}, \dots, \omega_{mt})';$$

Retorno do portfólio:

$$R_{pt} = \sum_{i=1}^m \omega_{it} R_{it} = \boldsymbol{\omega}'_t \mathbf{r}_t;$$

Valor esperado condicional do portfólio:

$$\mathbf{E} \left(R_{pt} \mid F_{t-1} \right) = \mathbf{E} \left(\boldsymbol{\omega}'_t \mathbf{r}_t \mid F_{t-1} \right) = \boldsymbol{\omega}'_t \boldsymbol{\mu}_t;$$

Variância condicional do portfólio

$$\begin{aligned} \text{Var} \left(R_{pt} \mid F_{t-1} \right) &= \text{Var} \left(\boldsymbol{\omega}'_t \mathbf{r}_t \mid F_{t-1} \right) \\ &= \boldsymbol{\omega}'_t \text{Var} \left(\mathbf{r}_t \mid F_{t-1} \right) \boldsymbol{\omega}_t = \boldsymbol{\omega}'_t \mathbf{H}_t \boldsymbol{\omega}_t. \end{aligned}$$

O problema de otimização (horizonte de investimento $h = 1$):

$$\begin{cases} \min_{\omega_{n+1}} \text{Var} \left(R_{p,n+1} \mid F_n \right) \\ \text{s.a } \mathbf{E} \left(R_{p,n+1} \mid F_n \right) = \mu_p \text{ e } \sum_{i=1}^m \omega_{i,n+1} = 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} \min_{\omega_i} \omega'_{n+1} \mathbf{H}_{n+1} \omega_{n+1} \\ \text{s.a } \omega'_{n+1} \boldsymbol{\mu}_{n+1} = \mu_p \text{ e } \omega'_{n+1} \mathbf{1} = 1 \end{cases}$$

Problema Estatístico:

- Estimar $\boldsymbol{\mu}_{n+1}$ (modelo VAR, por exemplo)
- Estimar \mathbf{H}_{n+1} (modelo da família GARCH multivariado, por exemplo)

5 Modelos Não Lineares na Média

5.1 Introdução

Sucesso dos modelos ARMA Gaussianos:

- simplicidade: as equações às diferenças finitas são fáceis de tratar;
- o modelo ARMA Gaussiano é completamente caracterizado pela média, variância e pelas autocorrelações;
- são fáceis de aplicar e implementar (a maioria dos programas de estatísticas possui um módulo para tratar modelos ARMA);
- a teoria está completamente desenvolvida: todas as principais questões, relacionadas com a estimação, inferência e previsão estão resolvidas;
- apesar da simplicidade são relativamente flexíveis e úteis na análise previsão de um grande número de séries económicas.

Todavia:

- não são apropriados para dados que exibam súbitas alterações em períodos irregulares;
- não são apropriados para dados que exibam forte assimetria e achatamento;
- os coeficiente de autocorrelação e autocorrelação parcial de ordem k são apenas um dos aspectos da distribuição conjunta de (y_t, y_{t-k}) . Pode suceder ρ_k e ϕ_{kk} serem baixos, mas existir uma forte relação não linear entre y_t e y_{t-k} . Por exemplo $\rho_k(y^2)$ pode ser alto;
- não modelam dados que exibam “soluções periódicas estáveis”. Retomaremos este tema mais à frente.

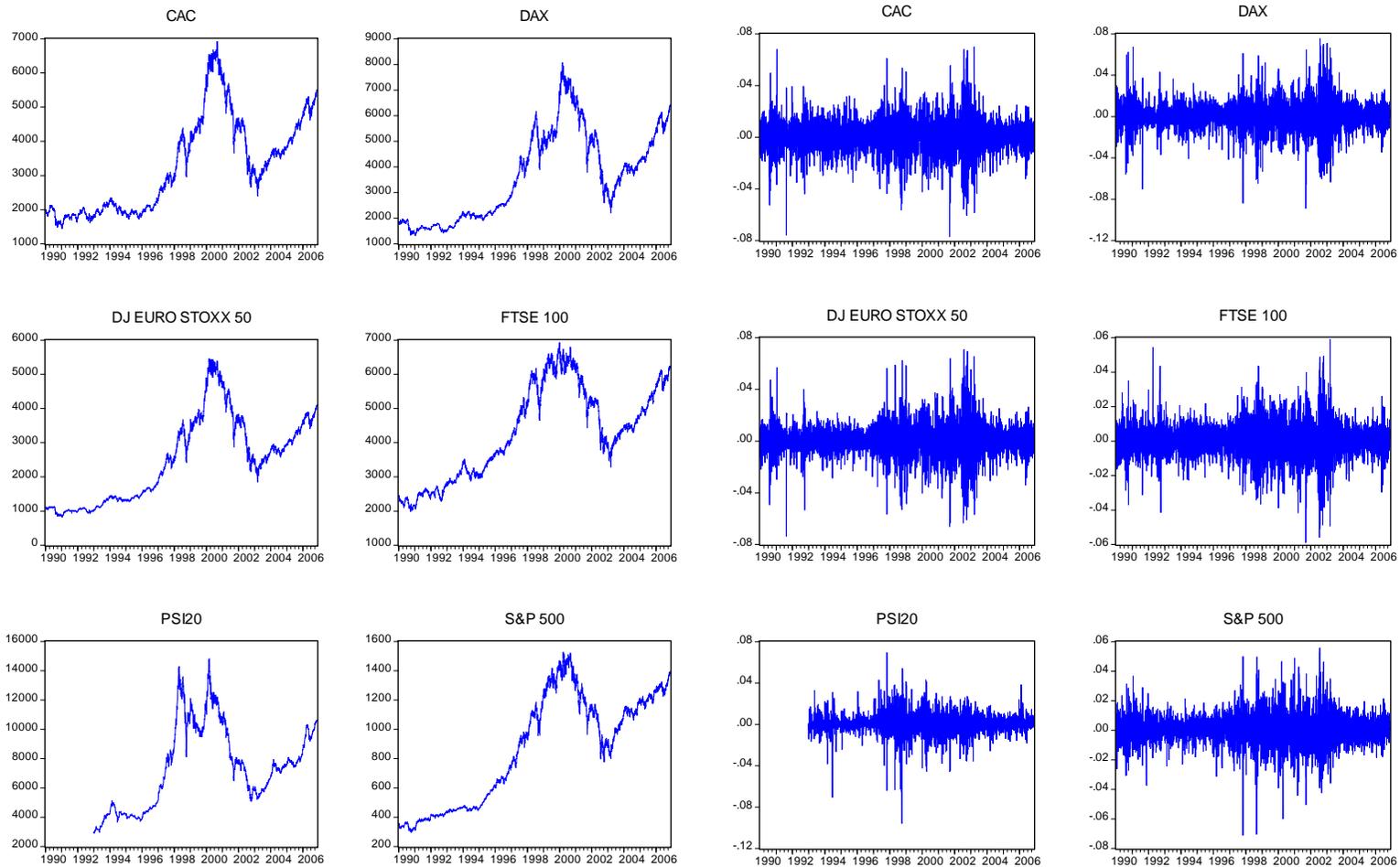
Discuta-se “não são apropriados para dados que exibam súbitas alterações em períodos irregulares”.

Modelo ARMA Gaussiano estacionário: possui momentos de todas as ordens, i.e., $\mathbf{E} [|y_t|^k] < \infty$ para qualquer $k \in \mathbb{N}$. Implicação? Se y possui o momento de ordem k , então (pela desigualdade de Markov) vem, para $c > 0$

$$P(|y| > c) \leq O(c^{-k}) \text{ quando } c \rightarrow \infty.$$

Consequências?

Exemplos



Índices Bolsistas

Retornos de Índices

Included observations: 5177

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.036	-0.036	6.6149	0.010
		2	-0.032	-0.034	11.993	0.002
		3	-0.043	-0.045	21.524	0.000
		4	-0.003	-0.007	21.564	0.000
		5	0.006	0.003	21.772	0.001
		6	0.001	-0.001	21.777	0.001
		7	0.004	0.004	21.878	0.003
		8	0.026	0.027	25.484	0.001
		9	-0.005	-0.002	25.601	0.002
		10	-0.014	-0.012	26.657	0.003
		11	-0.010	-0.009	27.168	0.004
		12	0.012	0.011	27.976	0.006
		13	0.016	0.015	29.370	0.006
		14	0.020	0.021	31.391	0.005
		15	-0.020	-0.016	33.404	0.004
		16	0.004	0.005	33.509	0.006
		17	0.003	0.005	33.569	0.010
		18	-0.017	-0.018	35.143	0.009
		19	-0.008	-0.009	35.506	0.012
		20	0.015	0.012	36.613	0.013

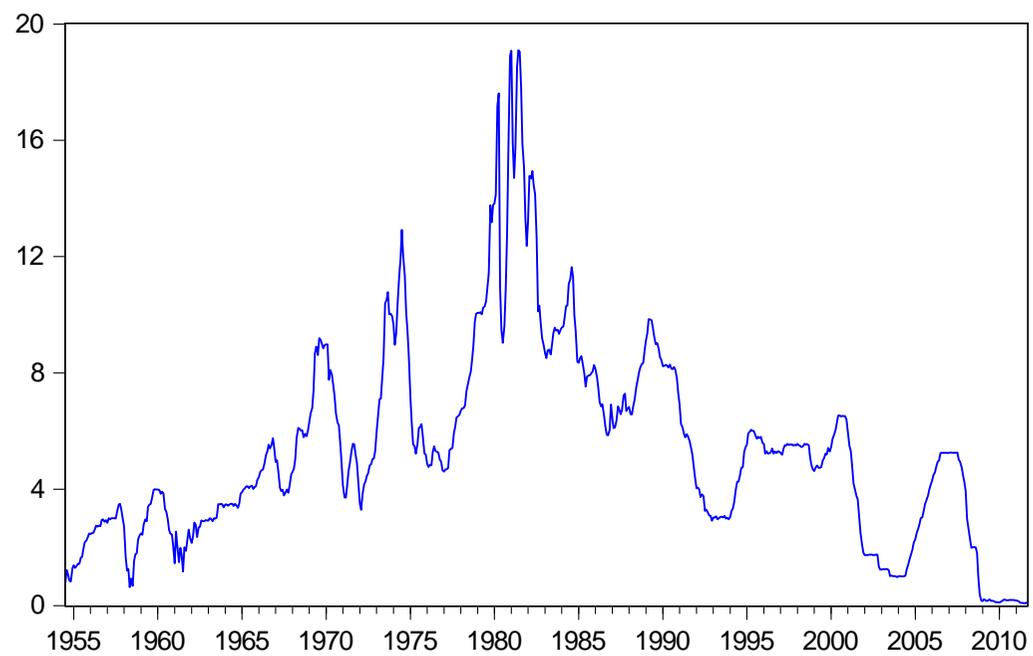
Funções de autocorrelação dos retornos diários (Microsoft 1986-2006)

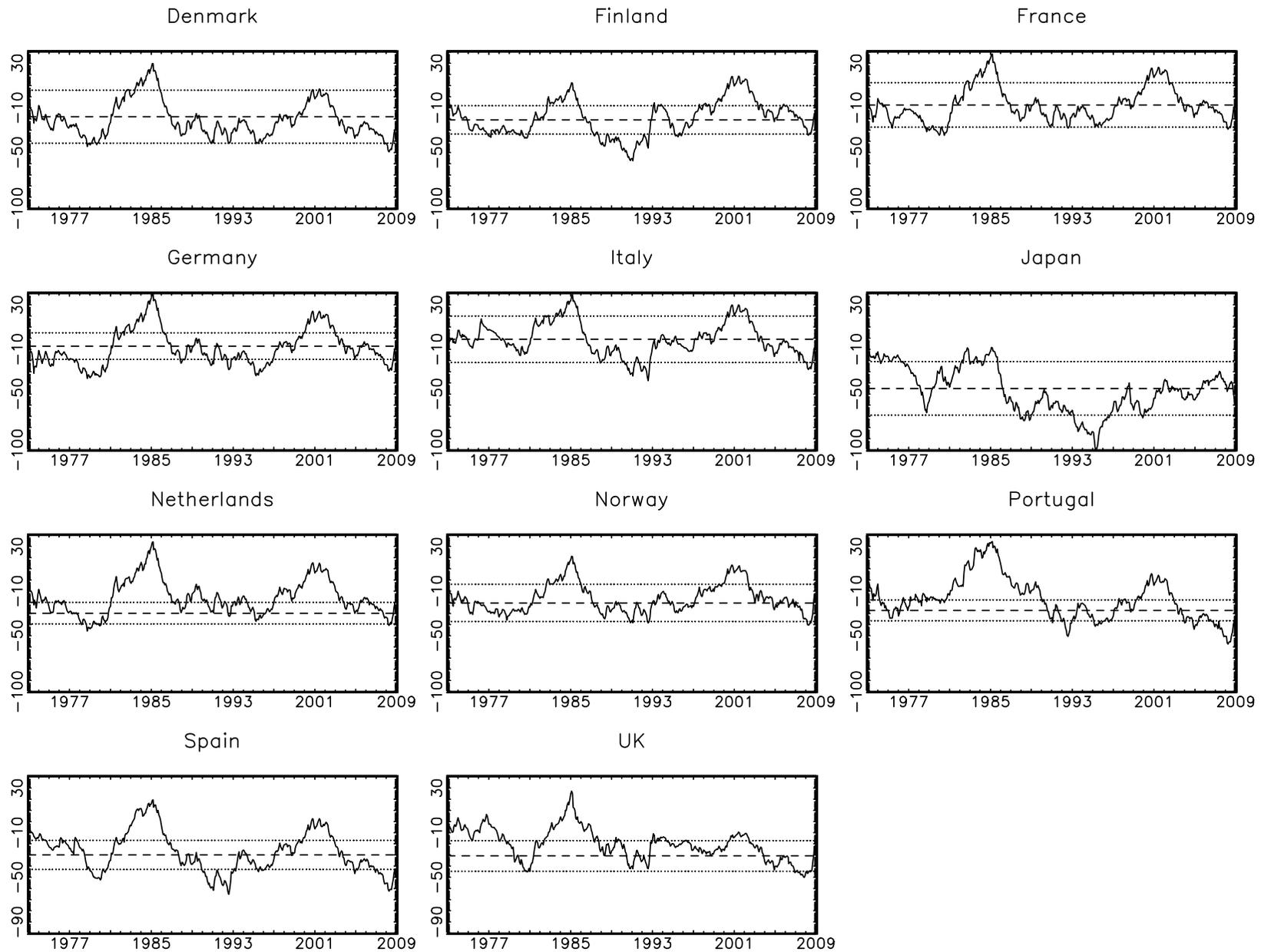
Included observations: 19405

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.205	0.205	819.60	0.000
		2	0.211	0.177	1687.0	0.000
		3	0.164	0.099	2207.3	0.000
		4	0.123	0.048	2500.6	0.000
		5	0.181	0.119	3138.2	0.000
		6	0.107	0.024	3361.2	0.000
		7	0.106	0.025	3578.1	0.000
		8	0.124	0.056	3876.5	0.000
		9	0.107	0.037	4098.3	0.000
		10	0.104	0.025	4306.9	0.000
		11	0.094	0.024	4479.6	0.000
		12	0.084	0.016	4616.4	0.000
		13	0.078	0.009	4734.4	0.000
		14	0.071	0.009	4832.9	0.000
		15	0.086	0.031	4977.0	0.000
		16	0.074	0.014	5084.2	0.000
		17	0.066	0.006	5168.3	0.000
		18	0.078	0.025	5287.8	0.000
		19	0.076	0.022	5398.6	0.000
		20	0.072	0.013	5500.6	0.000

Funções de autocorrelação dos quadrados dos retornos (Dow JOnes)

TX JURO (Federal funds rate)





Real Exchange Rates in the Period January 1973 - May 2009

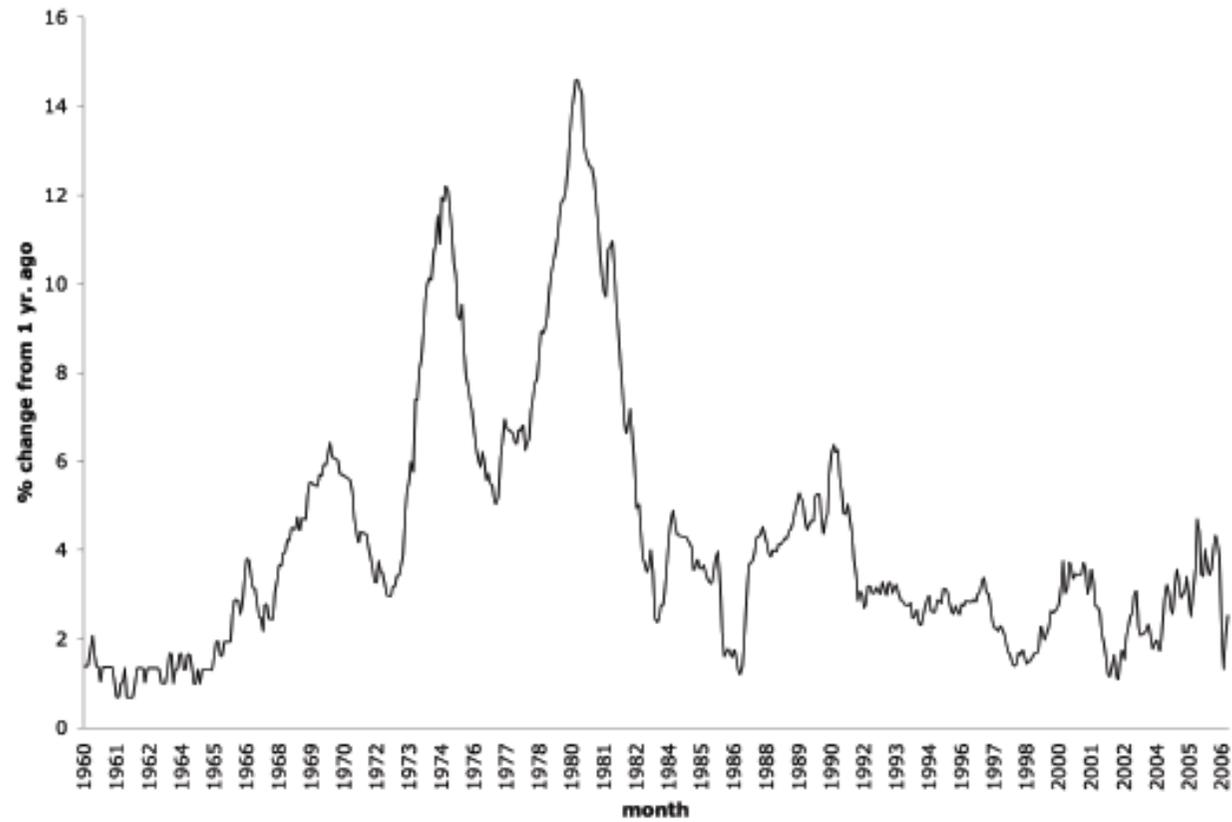


Figure 3
U.S. Inflation 1960-2006 (Source: *St. Louis Fed website*)

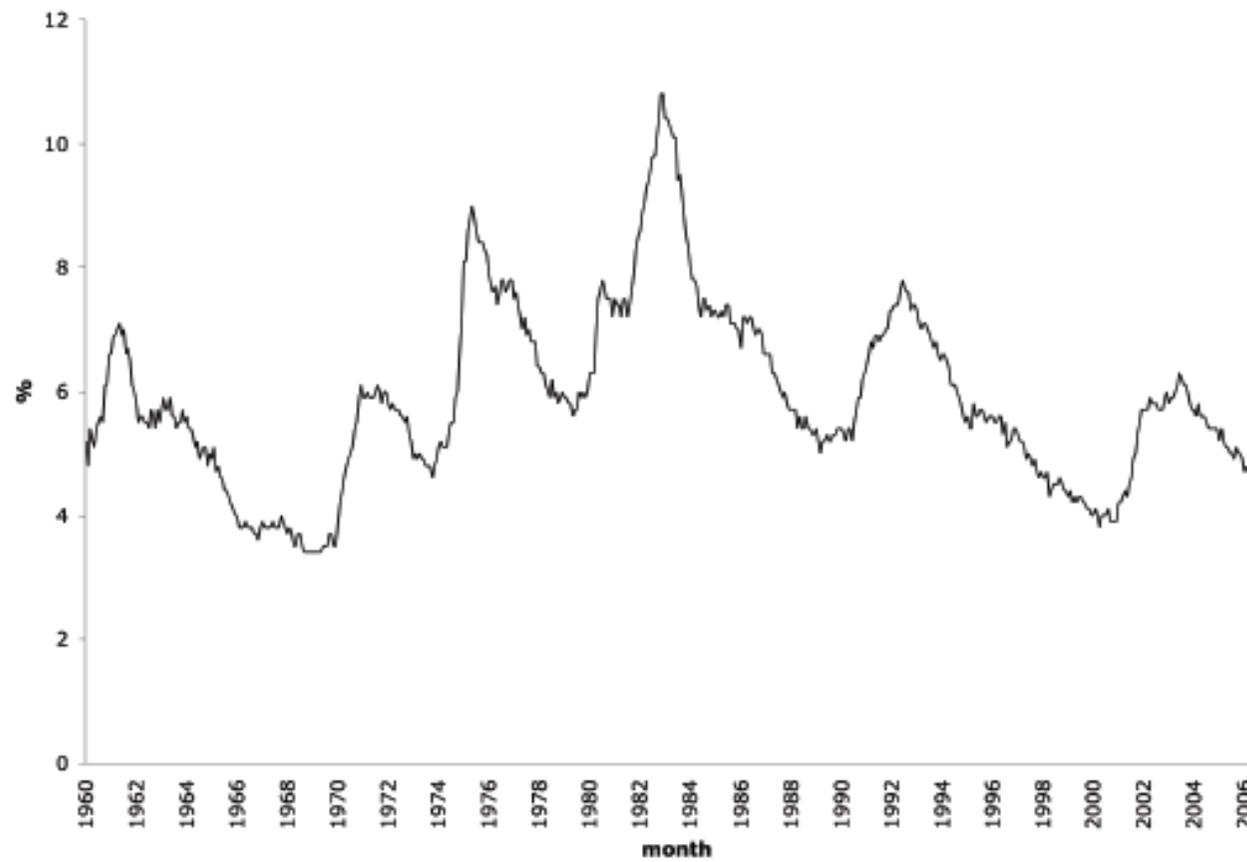


Figure 6

U.S. Civilian Unemployment Rate 1960-2006 (*Source: St. Louis Fed website*)

O que é um modelo de séries temporais não linear?

Considere-se o modelo

$$y_t = \mu_t + u_t, \quad u_t = \sigma_t \varepsilon_t$$

onde ε_t é um ruído branco,

$$\mu_t = g \left(y_{t-1}, y_{t-2}, \dots, y_{t-p}; u_{t-1}, u_{t-2}, \dots, u_{t-q} \right)$$

é a média condicional de y_t e

$$\sigma_t^2 = h \left(y_{t-1}, y_{t-2}, \dots, y_{t-p}; u_{t-1}, u_{t-2}, \dots, u_{t-q} \right) > 0$$

é a variância condicional de y_t .

O modelo é **não linear na média** se g é uma função não linear dos seus argumentos.

O modelo é **não linear na variância** se σ_t é não constante ao longo do tempo pois, neste caso, o processo $\{u_t\}$, definido por $u_t = \sigma_t \varepsilon_t$, é não linear, por ser um processo multiplicativo.

Veamos alguns exemplos.

- $y_t = \phi y_{t-1} + \log(1 + y_{t-1}^2) + u_t$

- $y_t = \xi u_{t-1} u_{t-2} + u_t$ (Bilinear)

- $y_t = \phi y_{t-1} + \sigma_t \varepsilon_t, \sigma_t^2 = \omega + \alpha_1 u_{t-1}^2$ (ARCH)

- $y_t = \begin{cases} \phi_{10} + \phi_{11} y_{t-1} + u_t & \text{se } y_{t-d} < \gamma \\ \phi_{20} + \phi_{21} y_{t-1} + u_t & \text{se } y_{t-d} > \gamma \end{cases}$ (TAR)

- $y_t = \mu + b_1 (y_{t-1} - \mu) + b_2 (y_{t-1} - \mu) F(y_{t-1}) + u_t, \quad F(x) = 1 - e^{-\theta^2(x-\mu)^2}$
(ESTAR: Exponential Smooth Autoregressive model)

- $y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + \sigma_1 \varepsilon_t & \text{se } S_t = 1 \\ c_2 + \phi_2 y_{t-1} + \sigma_2 \varepsilon_t & \text{se } S_t = 2 \end{cases}$ (Markov-Switching)

Uma classe importante de processos não lineares na média baseiam-se na ideia de *regime-switching*. Podem ser usados em duas circunstâncias gerais:

- existem alterações bruscas e inesperadas nas trajetórias dos processos (e.g., ataques especulativos, *crashes* bolsistas, anúncios públicos de medidas do governo, eventos políticos e, em geral, eventos extraordinários não antecipados).
- existem alterações da dinâmica do processo mas não ocorrerem alterações bruscas nas trajetórias.

Para este tipo de fenômenos, os modelos com alterações (estocásticas) de regime (ou *regime-switching*) podem ser, no essencial, de dois tipos:

- a mudança de regime é função de uma variável observável; são exemplos, modelos com variáveis impulso (*dummy*), os modelos limiares ou *threshold AR* (TAR), os modelos onde os coeficientes associados às componentes AR são funções não lineares dos valores passados y (STAR, *smoothed transition AR*), entre outros. Exemplo:

$$y_t = \begin{cases} \phi_{10} + \phi_{11}y_{t-1} + u_t & \text{se } y_{t-d} < \gamma \\ \phi_{20} + \phi_{21}y_{t-1} + u_t & \text{se } y_{t-d} > \gamma. \end{cases}$$

- a mudança de regime não é observada, incluindo-se, nesta classe, os modelos onde os regimes são independentes entre si (como, por exemplo, os modelos *simple switching* ou de Bernoulli) e os modelos onde existe dependência entre os regimes (como por exemplo, os modelos MS, *Markov-Switching*). Exemplo:

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + \sigma_1 \varepsilon_t & \text{se } S_t = 1 \\ c_2 + \phi_2 y_{t-1} + \sigma_2 \varepsilon_t & \text{se } S_t = 2 \end{cases}$$

5.2 Estabilidade em Equações às Diferenças Finitas Determinísticas

Considere-se a equação determinística

$$y_t = g(y_{t-1}).$$

Definição \bar{y} é um ponto fixo de g se $g(\bar{y}) = \bar{y}$.

No ponto fixo a solução y_t não varia (está em equilíbrio): $y_{t-1} = \bar{y} \Rightarrow \Delta y_t = 0$.

Exemplo Considere-se $y_0 = 2$ e $y_t = (1/2)y_{t-1} + 1$. Iterando ...

Exemplo Considere-se $y_t = 2y_{t-1}(1 - y_{t-1})$. Tem-se $g(x) = 2x(1 - x)$. Pontos fixos (de g)?

Obtenha-se o gráfico teia de aranha das equações **a)** $y_t = 0.5y_{t-1}$, com $y_0 = 4$; **b)** $y_t = y_{t-1}^2$ (o que sucede $y_0 \in (1, \infty)$? $y_0 \in (0, 1)$? $y_0 \in (-1, 0)$?).

Seja $y_t = g^t(y_0)$ a solução da EDF $y_t = g(y_{t-1})$, dada a condição inicial y_0

$$g^t(x) := \underbrace{g(\dots g(g(x)))}_{t \text{ vezes}}.$$

Definição O ponto fixo \bar{y} diz-se estável se para cada $\varepsilon > 0$ existe um $\delta = \delta(t_0, \varepsilon)$ tal que, para cada qualquer solução $y_t(y_0)$ a desigualdade $|y_0 - \bar{y}| \leq \delta$ implica $|y_t(y_0) - \bar{y}| < \varepsilon$ para todo o $t \geq t_0$. O ponto fixo \bar{y} diz-se assintoticamente estável se é estável e se existe um $\delta_0 > 0$ tal que a desigualdade $|y_0 - \bar{y}| \leq \delta_0$ implica $|y_t(y_0) - \bar{y}| \rightarrow 0$ quando $t \rightarrow \infty$. O ponto fixo \bar{y} diz-se instável se não é estável.

Grosso modo, um ponto fixo \bar{y} é estável se $y_t = g^t(y_0)$ permanecer "perto" de \bar{y} para todo o t sempre que y_0 se encontrar "perto" de \bar{y} . O ponto fixo \bar{y} é assintoticamente estável se for estável e toda a solução inicializada perto de \bar{y} converge para \bar{y} .

Exemplo Estude-se $y_t = c + \phi y_{t-1}$, com $c \neq 0$.

Definição p é um ponto periódico de período k se

$$g^k(p) = p \quad (*)$$

e k é o menor inteiro positivo tal que (*) se verifica.

Note-se que se p é um ponto periódico de período 2 então p é um ponto fixo de g^2 . O recíproco não é verdade. Por exemplo, um ponto fixo de g^2 pode ser também um ponto fixo de g e, neste caso, de acordo com a definição, este ponto tem período 1.

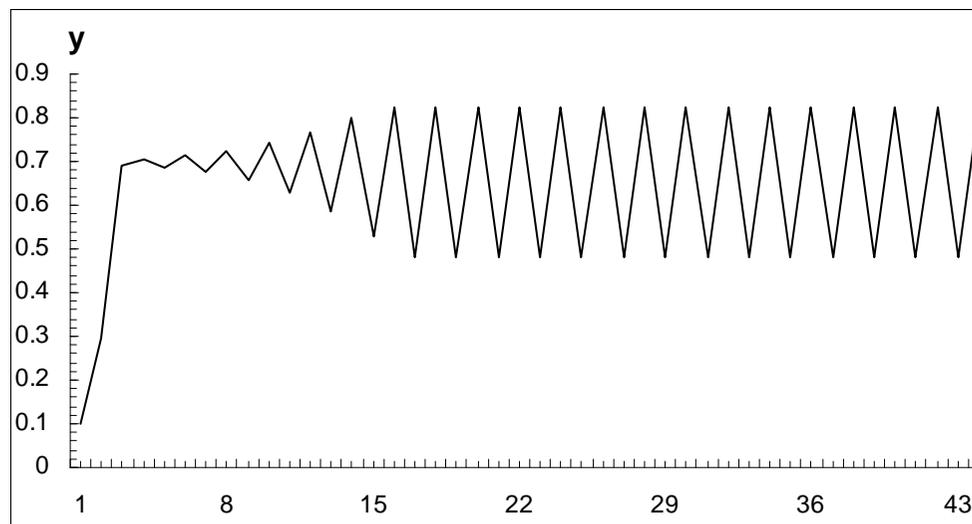
Considere-se uma órbita de valor inicial p , i.e., $\{p, g(p), g^2(p), \dots\}$. Se p é um ponto periódico de período 3, p deve repetir-se de três em três iterações. Por exemplo,

$$\{p, g(p), g^2(p), p, \dots\}.$$

Mas $g(p)$ e $g^2(p)$ também se repetem de três em três iterações,

$$\{\dots, p, g(p), g^2(p), p, g(p), g^2(p), p, \dots\}.$$

Exemplo Considere-se a equação $y_t = 3.3y_{t-1}(1 - y_{t-1})$. Investigue-se se existem pontos periódicos de período 2.



Trajectória de $y_t = 3.3y_{t-1}(1 - y_{t-1})$, $y_0 = 0.1$

t:	0	1	2	3	4	5	6	7	8	9	10	11	12	13
yt:	0.1000	0.2970	0.6890	0.7071	0.6835	0.7139	0.6740	0.7251	0.6577	0.7429	0.6303	0.7690	0.5863	0.8004
t:	14	15	16	17	18	19	20	21	22	23	24	25	26	27
yt:	0.5271	0.8226	0.4816	0.8239	0.4788	0.8235	0.4796	0.8236	0.4794	0.8236	0.4794	0.8236	0.4794	0.8236
t:	28	29	30	31	32	33	34	35	36	37	38	39	40	41
yt:	0.4794	0.8236	0.4794	0.8236	0.4794	0.8236	0.4794	0.8236	0.4794	0.8236	0.4794	0.8236	0.4794	0.8236

Sucessão $y_t = 3.3y_{t-1}(1 - y_{t-1})$, $y_0 = 0.1$, $t = 1, \dots, 41$

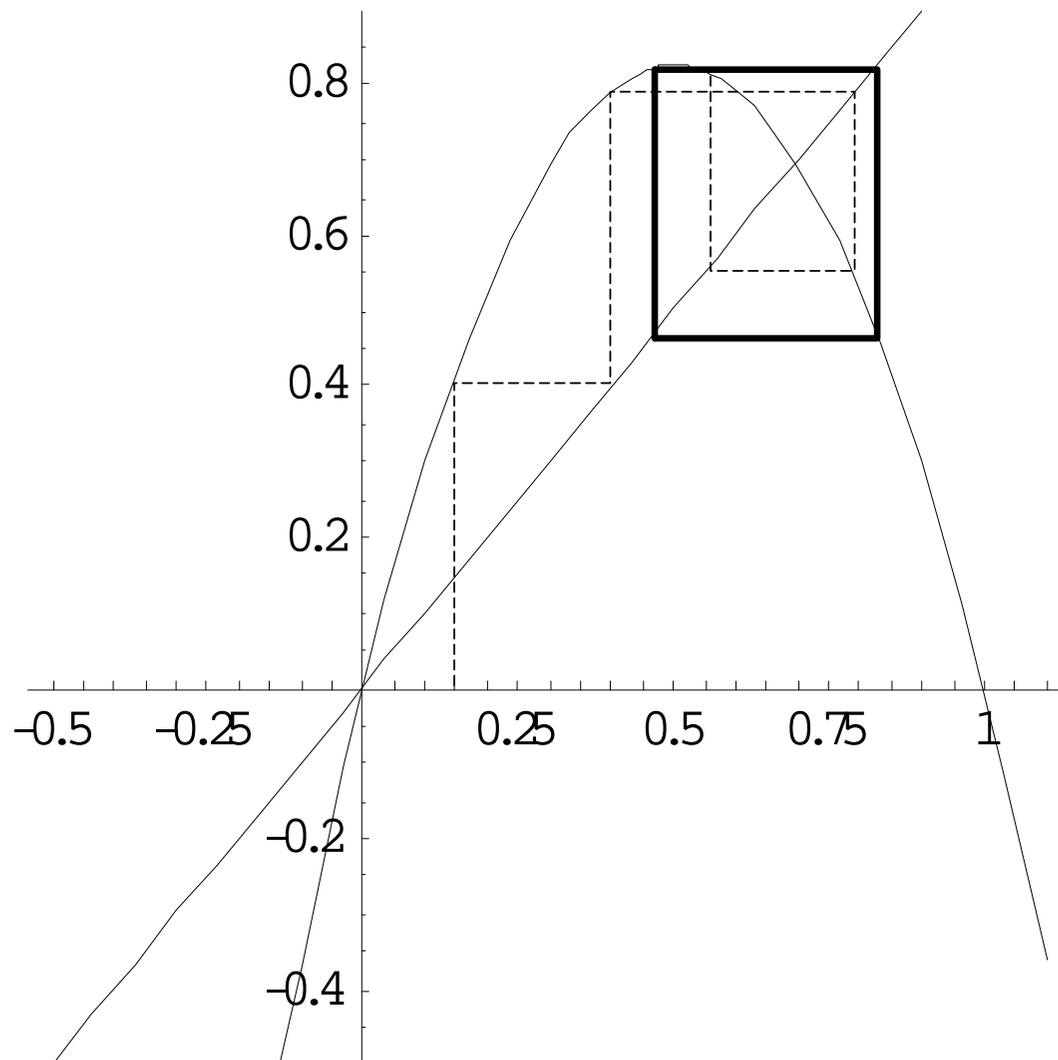


Gráfico Teia de Aranha da equação $y_t = 3.3y_{t-1}(1 - y_{t-1})$

Um modelo que pode gerar pontos periódicos é o modelo linear por troços (ou TAR). Assume-se um comportamento diferenciado do sistema dinâmico consoante o estado do sistema no momento $t - 1$, concretamente, consoante y_{t-1} se encontre abaixo ou acima de certo limiar γ . No caso mais simples o modelo corresponde a

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1}, & y_{t-1} < \gamma \\ c_2 + \phi_2 y_{t-1}, & y_{t-1} \geq \gamma. \end{cases}$$

5.3 Processos Não Lineares e Estacionaridade Estrita

5.3.1 Processos Ergódicos e Estritamente Estacionários

Definição y é um processo estacionário de segunda ordem (ESO) se

$$\begin{aligned} E(y_t) &= \mu, & (\text{não depende de } t) \\ \text{Var}(y_t) &= \sigma^2, & (\text{não depende de } t) \\ \text{Cov}(y_t, y_{t\pm h}) &= \gamma(h) & (\text{não depende de } t). \end{aligned}$$

Definição y é um processo estritamente estacionário (EE) se as distribuições conjuntas de

$$(y_1, y_2, \dots, y_s) \text{ e } (y_{k+1}, y_{k+2}, \dots, y_{k+s})$$

são iguais para qualquer $s \in \mathbb{N}$ e $k \in \mathbb{Z}$.

Definição y é um processo de Markov se

$$P(y_t < \xi | y_1, \dots, y_{t-1}) = P(y_t < \xi | y_{t-1}).$$

O processo AR(2), $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$, não é de Markov pois

$$P(y_t < \xi | y_1, \dots, y_{t-1}) = P(y_t < \xi | y_{t-2}, y_{t-1}) \neq P(y_t < \xi | y_{t-1}).$$

No entanto, é possível representá-lo na forma Markoviana [Quadro].

Também o processo não linear do tipo

$$y_t = g(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + u_t$$

onde $\{u_t\}$ é uma sequência de v.a. i.i.d. e independentes de y_{t-k} , $k \geq 1$, admite uma representação Markoviana.

Doravante trabalha-se com processos de Markov.

Definição A função de distribuição condicional a n passos de um processo de Markov é

$$P(y_{n+k} < y | y_k).$$

$f_n(y|x)$ é a densidade associada (que admitimos existir)

$$f_n(y|x) = \frac{\partial P(y_{n+k} < y | y_k)}{\partial y}.$$

Exemplo Considere-se

$$y_t = \phi y_{t-1} + u_t, \quad u_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2), \quad |\phi| < 1.$$

Calcule $f_1(y|x)$ e $f_n(y|x)$. Nota: Para processos não lineares, conhece-se geralmente $f_1(y|x)$ (uma vez especificada a distribuição dos erros), mas não $f_n(y|x)$, para $n > 1$.

Definição y é um processo homogêneo se

$$P(y_{t+k} < \xi | y_k = x) = P(y_{t+s} < \xi | y_s = x), \quad (t \in \mathbb{N})$$

para todo o k e $s \in \mathbb{Z}$.

Exemplo Retome-se o exemplo $y_t = \phi y_{t-1} + u_t$, $u_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $|\phi| < 1$. Discuta a homogeneidade, sabendo que

$$\begin{aligned} y_{t+k} | y_k &= x \sim N\left(\phi^t x, \frac{\sigma^2}{1 - \phi^2} + \frac{\sigma^2 \phi^{2t}}{\phi^2 - 1}\right) \\ y_{t+s} | y_s &= x \sim N\left(\phi^t x, \frac{\sigma^2}{1 - \phi^2} + \frac{\sigma^2 \phi^{2t}}{\phi^2 - 1}\right). \end{aligned}$$

Se y é um processo homogêneo, para avaliar probabilidades condicionadas acerca da posição do processo num dado momento, não interessa o momento em que essas probabilidades são calculadas; apenas interessa saber o valor de t .

Um processo em que pelo menos um dos momentos condicionais depende explicitamente de t (por exemplo, $E[y_t | y_{t-1}] = y_{t-1}/t$) poderá ser um processo homogêneo?

Comente: "A homogeneidade é uma condição necessária, mas não suficiente para garantir estacionaridade (estrita ou de segunda ordem)" no contexto do

Exemplo Seja $y_t = y_{t-1} + u_t$, onde $u_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$. Tem-se,

$$y_{t+k} = y_k + \sum_{i=k+1}^{t+k} u_i \Big| y_k = x \sim$$

$$y_{t+s} = y_s + \sum_{i=s+1}^{t+s} u_i \Big| y_s = x \sim$$

A definição de ergodicidade (à “Harris”) que se apresenta a seguir envolve um conceito de proximidade entre duas funções de distribuição, H e G , baseada na norma de *variação total* $\|\cdot\|$ que, caso existam as funções densidades associadas a H e G , é dada por

$$\|H - G\| = \int |h(x) - g(x)| dx.$$

Considere-se o processo estocástico y com fdp condicional a n passos dada por $f_n(y|x)$.

Definição Se existir uma função de densidade f e uma constante $\rho \in (0, 1]$ tal que

$$\rho^{-n} \int |f_n(y|x) - f(y)| dy \rightarrow 0$$

para todo o x , então y é ergódico se $\rho = 1$ e geometricamente ergódico se $\rho < 1$. f é a função de densidade estacionária.

Proposição Suponha-se que y é ergódico. Então existe uma distribuição estacionária F tal que o processo y , inicializado com a distribuição F , é estritamente estacionário.

Para processos nas condições da proposição anterior, a lei forte dos grandes números verifica-se sempre:

Proposição Suponha-se $E[|h(y)|] < \infty$. Nas condições anteriores verifica-se

$$\frac{1}{n} \sum_{t=1}^n h(y_t) \xrightarrow{p} E[h(y)]$$

(também convergência quase certa ou com probabilidade 1) qualquer que seja o valor inicial do processo y .

Em geral, é difícil verificar directamente $\rho^{-n} \int |f_n(y|x) - f(y)| dy \rightarrow 0$, a não ser para casos relativamente simples, como o do exemplo que se apresenta a seguir.

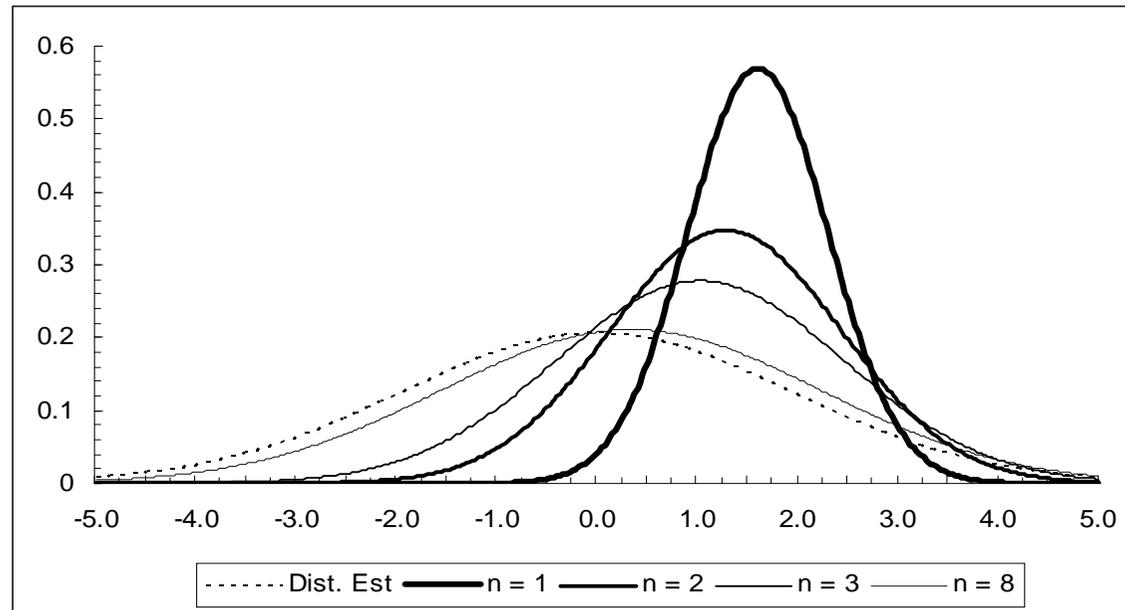
Exemplo Retome-se o exemplo $y_t = \phi y_{t-1} + u_t$, $u_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $|\phi| < 1$. Concluimos que

$$y_n | F_0 \sim N \left(\phi^n y_0, \frac{\sigma^2}{1 - \phi^2} + \frac{\sigma^2 \phi^{2n}}{\phi^2 - 1} \right)$$

$$f_n(y|x) =$$

Calcule, $\lim_{n \rightarrow \infty} E[y_n | x]$, $\lim_{n \rightarrow \infty} \text{Var}[y_n | x]$ e $\lim_{n \rightarrow \infty} f_n(y|x)$.

Exemplo Para exemplificar, considere-se $x = 2$, $\sigma^2 = 0.7$ e $\phi = 0.8$.



$\{f_n(y|x=2), n=1, 2, 3, 8\}$ e $f(y)$

No exemplo anterior conseguimos discutir

$$\rho^{-n} \int |f_n(y|x) - f(y)| dy \rightarrow 0$$

porque y é linear. Para modelos não linear é geralmente impossível obter $f_n(y|x)$ e $f(y)$. Felizmente, é possível, em certas circunstâncias, provar-se $\rho^{-n} \int |f_n(y|x) - f(y)| dy \rightarrow 0$ de forma indirecta, através de resultados auxiliares. É isso que veremos a seguir.

5.3.2 Modelos Não Lineares do tipo $y_t = g(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + u_t$

Proposição Considere-se o processo $\{y\}$ definido por

$$y_t = g(y_{t-1}, \dots, y_{t-p}) + u_t$$

onde $g : \mathbb{R}^p \rightarrow \mathbb{R}$, $\{u_t\}$ é uma sequência de v.a. i.i.d. e independentes de y_{t-k} , $k \geq 1$, com função de densidade positiva e $\mathbf{E}[u_t] = 0$. O processo $\{y\}$ é geometricamente ergódico se qualquer uma das seguintes condições se verificar:

(a) g é uma função limitada em conjuntos limitados,

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{|g(\mathbf{x}) - (b_1 x_1 + \dots + b_p x_p)|}{\|\mathbf{x}\|} = 0$$

e as raízes do polinómio $1 - b_1 z - \dots - b_p z^p$ estão fora do círculo unitário.

(b) Existe uma constante $\lambda \in (0, 1)$ e c tal que

$$|g(\mathbf{x})| \leq \lambda \max\{|x_1|, \dots, |x_p|\} + c.$$

(c) Existe uma constante $\rho \in (0, 1)$, c e $a_i \geq 0$ e $a_1 + \dots + a_p = 1$ tal que

$$|g(\mathbf{x})| \leq \rho (a_1 |x_1| + \dots + a_p |x_p|) + c.$$

Exemplo Considere-se o processo AR(2),

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

onde u é um ruído branco. [Quadro]

Exemplo Considere-se o processo

$$y_t = \phi \frac{y_{t-1}^2}{1 + y_{t-1}^2} + u_t$$

onde u é um ruído branco. [Quadro]

Exemplo Considere-se o processo

$$y_t = \begin{cases} -0.5y_{t-1} + \varepsilon_t & \text{se } y_{t-1} < 0 \\ 0.5y_{t-1} + \varepsilon_t & \text{se } y_{t-1} \geq 0. \end{cases}$$

onde ε é um ruído branco. [Quadro]

5.3.3 Modelos Não Lineares do tipo $\mathbf{y}_t = \mathbf{A}_t \mathbf{y}_{t-1} + \mathbf{B}_t$

Definição Uma equação vectorial autoregressiva de dimensão d com coeficientes aleatórios i.i.d. não negativos é uma equação da forma

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{y}_{t-1} + \mathbf{B}_t$$

onde $\{(\mathbf{A}_t, \mathbf{B}_t), t \in \mathbb{Z}\}$ é uma sucessão de matrizes i.i.d..

Inúmeros processos podem ser representados na forma $\mathbf{y}_t = \mathbf{A}_t \mathbf{y}_{t-1} + \mathbf{B}_t$.

Proposição No caso escalar, $d = 1$,

$$y_t = A_t y_{t-1} + B_t,$$

assuma-se $-\infty \leq \mathbf{E} [\log |A_1|] < 0$ e $\mathbf{E} [\log^+ |B_1|] < \infty$. Então y converge (com probabilidade um) e a sua solução é estritamente estacionária.

Exemplo Discuta-se a estacionaridade do processo

$$y_t = \phi y_{t-1} + u_t, \quad \phi > 0$$

onde u é ruído branco Gaussiano. [Quadro]

Exemplo Discuta-se a estacionaridade do processo

$$y_t = e^{\mu + e_t} y_{t-1} + u_t$$

onde μ é uma constante e e e u são ruídos brancos com variância igual a 1, Gaussianos e independentes. [Quadro]

Exemplo Considere-se o processo y_t com heterocedasticidade condicionada de ordem (1, 1), (GARCH(1,1))

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \omega, \alpha_1, \beta_1 > 0. \end{aligned}$$

onde ε é um ruído branco estritamente estacionário e independente de u_{t-k} , $k \in \mathbb{N}$. [Quadro]

5.4 Modelo Limiar Autoregressivo (Threshold AR) TAR

5.4.1 Introdução

Apresentam-se modelos onde a mudança de regime depende de uma variável observável. Provavelmente o modelo mais conhecido desta família é o modelo *Threshold AR* ou TAR. Exemplo:

$$y_t = \begin{cases} \phi_{10} + \phi_{11}y_{t-1} + u_t & q_{t-d} < \gamma \\ \phi_{20} + \phi_{21}y_{t-1} + u_t & q_{t-d} > \gamma \end{cases}$$

γ : valor do limiar (*threshold values*), q_{t-d} : variável limiar e $d > 0$ é um inteiro.

Se $q_{t-d} = y_{t-d}$ o modelo designa-se por *Self-Exciting TAR* ou SETAR (é o próprio processo y que determina a mudança de regime).

A ideia: se certo fenómeno y é não linear na média, então é incorrecto assumir que y se comporta linearmente em todo o seu domínio. A solução que está implícita no modelo TAR consiste em assumir linearizações diferenciadas, consoante o valor de y . Em lugar de se ter

uma aproximação linear global, têm-se várias linearizações em sub-intervalos do espaço de estados.

Outra interpretação do TAR:

$$\begin{aligned}
 y_t &= \left(\phi_{10} I_{\{q_{t-d} \leq \gamma\}} + \phi_{11} I_{\{q_{t-d} \leq \gamma\}} y_{t-1} \right) \\
 &\quad + \left(\phi_{20} I_{\{q_{t-d} > \gamma\}} + \phi_{21} I_{\{q_{t-d} > \gamma\}} y_{t-1} \right) + u_t. \\
 &= \underbrace{\left(\phi_{10} I_{\{q_{t-d} \leq \gamma\}} + \phi_{20} I_{\{q_{t-d} > \gamma\}} \right)}_{\phi_0(q_{t-d})} + \underbrace{\left(\phi_{11} I_{\{q_{t-d} \leq \gamma\}} + \phi_{21} I_{\{q_{t-d} > \gamma\}} \right)}_{\phi_1(q_{t-d})} y_{t-1} + u_t \\
 &= \phi_0(q_{t-d}) + \phi_1(q_{t-d}) y_{t-1} + u_t.
 \end{aligned}$$

TAR: processo AR de coeficientes aleatórios com dependência em q_{t-d} .

5.4.2 Soluções Periódicas

Considere-se o processo linear, $y_t = c + \phi y_{t-1} + u_t$, $|\phi| < 1$. Ponto fixo único da equação determinística: $\bar{y} = c / (1 - \phi)$

Processos não lineares podem admitir *esqueletos* com vários pontos fixos e com periodicidade superior a um. Esta possibilidade permite modelar dados que exibam “soluções periódicas estáveis”.

Para concretizar considere-se o exemplo

$$y_t = \begin{cases} 1 + 0.5y_{t-1} + u_t & y_{t-1} < 0 \\ -1 - 0.5y_{t-1} + u_t & y_{t-1} \geq 0. \end{cases}$$

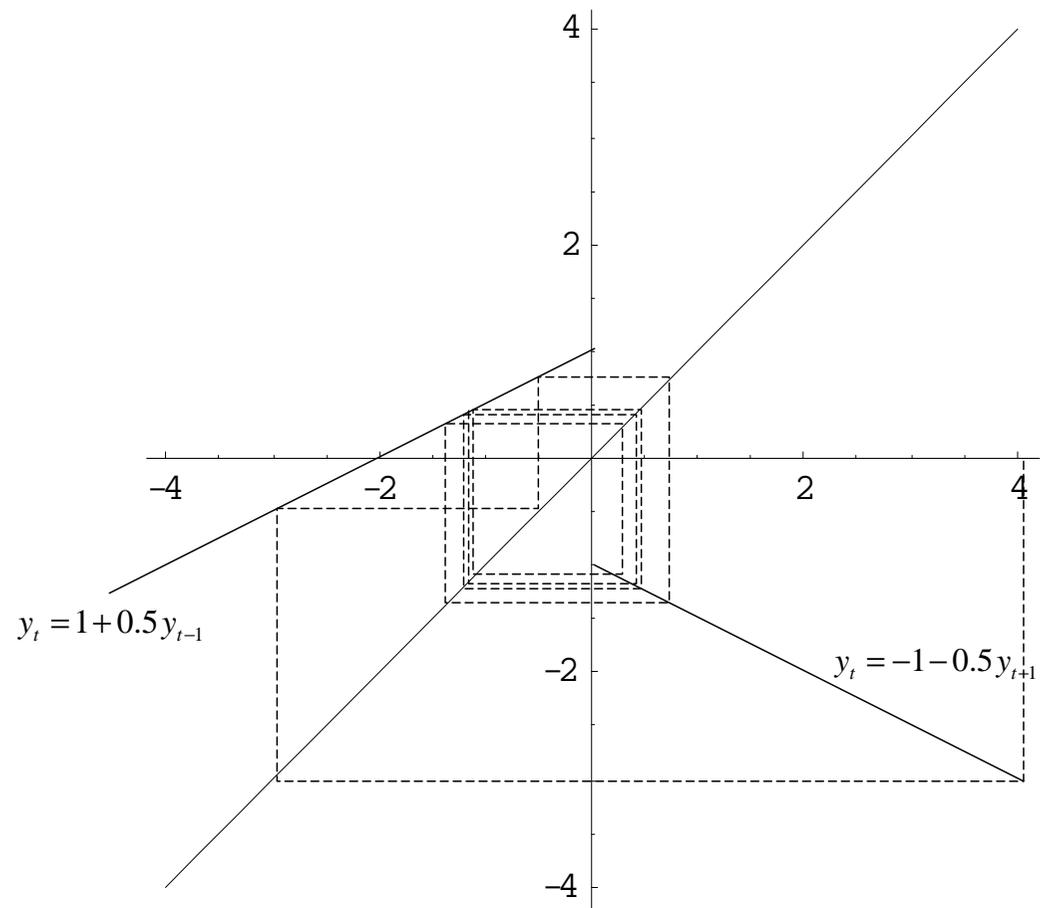
O *esqueleto* da equação é

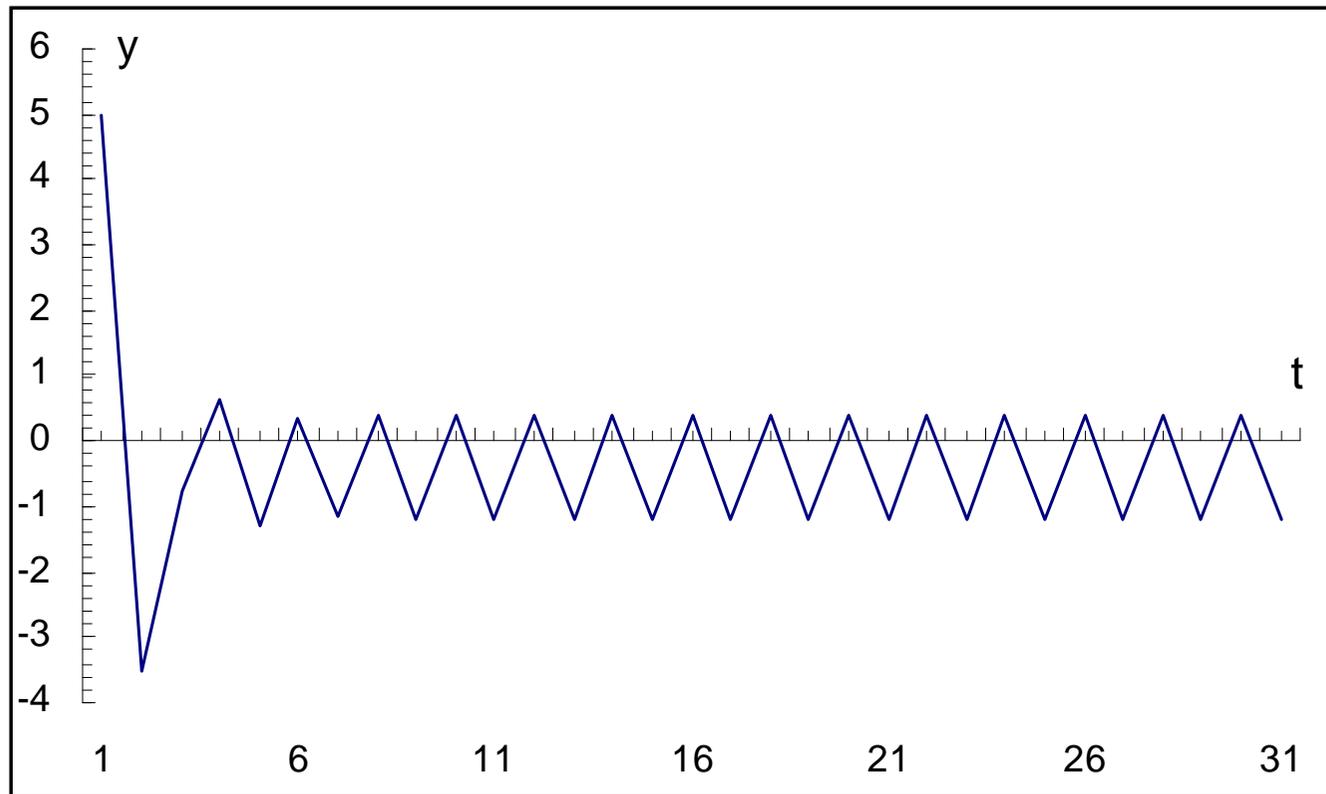
$$\tilde{y}_t = \begin{cases} 1 + 0.5\tilde{y}_{t-1}, & y_{t-1} < 0 \\ -1 - 0.5\tilde{y}_{t-1}, & y_{t-1} \geq 0 \end{cases}$$

e a função $g(x)$ representa-se na forma

$$g(x) = \begin{cases} 1 + 0.5x & x < 0 \\ -1 - 0.5x & x \geq 0. \end{cases}$$

Pode-se provar que 0.4 e -1.2 são os pontos periódicos de período 2. Estes pontos são assintoticamente estáveis. Qualquer que seja o valor de inicialização, a solução \tilde{y}_t é atraída para a órbita $\{0.4, -1.2\}$.

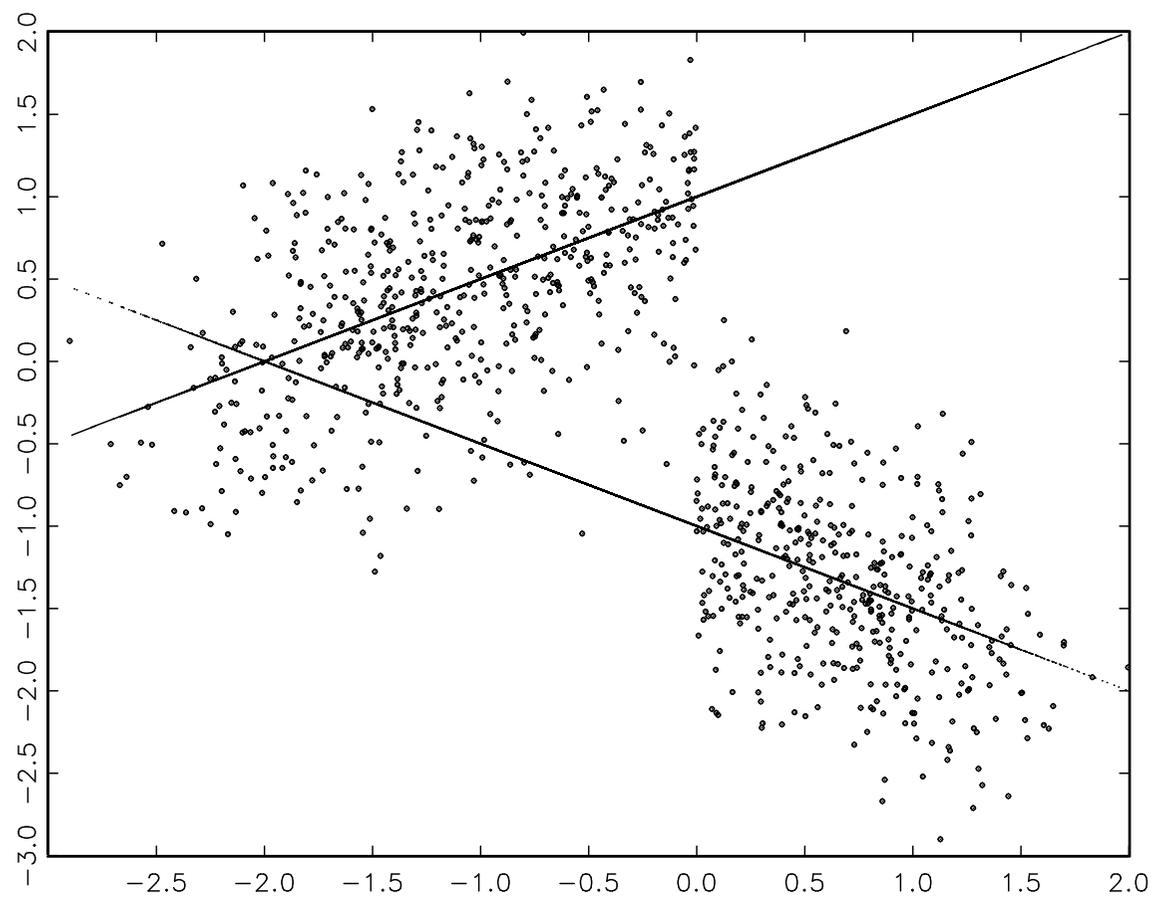




O comportamento periódico subjacente ao *esqueleto* está parcialmente presente no modelo estocástico. Na figura seguinte apresenta-se o gráfico *scatter* - gráfico dos pares (y_{t-1}, y_t) , após se ter gerado uma trajetória do processo

$$y_t = \begin{cases} 1 + 0.5y_{t-1} + u_t & y_{t-1} < 0 \\ -1 - 0.5y_{t-1} + u_t & y_{t-1} \geq 0. \end{cases}$$

com $u_t \sim N(0, 1)$. Traçam-se também as rectas $1 + .5x$ e $-1 - 0.5x$.



Como o processo é estritamente estacionário, que tipo de distribuição estacionária f define o processo? f é desconhecida, mas pode estimar-se de várias formas.

- simula-se uma trajectória suficientemente longa;
- estima-se a função de densidade marginal $f(x)$ através do estimador

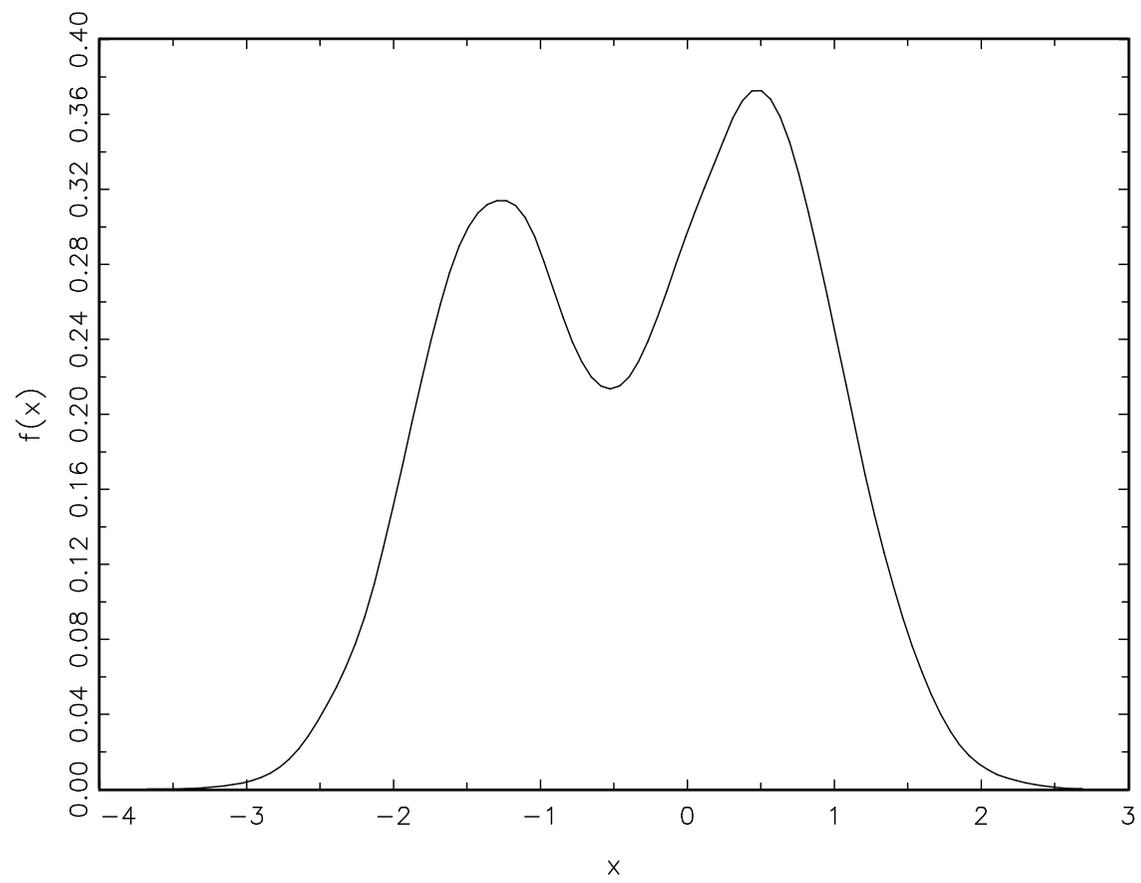
$$\hat{f}(x) = \frac{1}{n_S h} \sum_{i=1}^{n_S} K\left(\frac{x - y_i}{h}\right)$$

onde n_S é o número de observações simuladas usadas na estimação e K é, por exemplo, o Kernel (ou núcleo) Gaussiano

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

Sob certas condições sabe-se que $\hat{f}(x) \xrightarrow{p} f(x)$ onde f não é mais do que a densidade estacionária.

Naturalmente é necessário considerar vários valores de x para se ter uma ideia do comportamento de f no espaço de estados da variável. Como n_S está sob nosso controle, a função f pode ser estimada de forma arbitrariamente precisa.



5.4.3 Estacionaridade

A estacionaridade estrita pode ser analisada no contexto das proposições apresentadas anteriormente. Por exemplo, no modelo mais geral

$$y_t = \begin{cases} c_1 + \phi_{11}y_{t-1} + \dots + \phi_{1p}y_{t-p} + u_t & y_{t-d} < \gamma_1 \\ \dots & \dots \\ c_k + \phi_{k1}y_{t-1} + \dots + \phi_{kp}y_{t-p} + u_t & y_{t-d} > \gamma_k \end{cases}$$

se a condição $\max_{1 \leq i \leq k} \sum_{j=1}^p |\phi_{ji}| < 1$ se verificar, então o processo é estritamente estacionário. Esta condição é apenas suficiente, mas não necessária. Como tal, pode ser demasiado exigente.

Para certos modelos particulares são conhecidas as condições N&S:

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + u_t & y_{t-d} < \gamma_1 \\ c_2 + \phi_2 y_{t-1} + u_t & \gamma_1 \leq y_{t-d} \leq \gamma_2 \\ c_3 + \phi_3 y_{t-1} + u_t & y_{t-d} > \gamma_2 \end{cases}$$

Proposição O processo $\{y_t\}$ é estritamente estacionário sse qualquer uma das seguintes condições se verificar:

- (i) $\phi_1 < 1$, $\phi_3 < 1$, $\phi_1 \phi_3 < 1$; (ii) $\phi_1 = 1$, $\phi_3 < 1$, $c_1 > 0$;
 (iii) $\phi_1 < 1$, $\phi_3 = 1$, $c_3 < 0$; (iv) $\phi_1 = 1$, $\phi_3 = 1$, $c_3 < 0 < c_1$;
 (v) $\phi_1 \phi_3 = 1$, $\phi_1 < 1$, $c_1 + \phi_3 c_1 > 0$;

Há uma dado curioso nestas condições. Qual é?

5.4.4 Exemplo (Bounded Random Walk)

Um caso muito interessante é o seguinte:

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + u_t & y_{t-1} < \gamma_1 \\ c_2 + y_{t-1} + u_t & \gamma_1 \leq y_{t-1} \leq \gamma_2 \\ c_3 + \phi_3 y_{t-1} + u_t & y_{t-1} > \gamma_2 \end{cases}$$

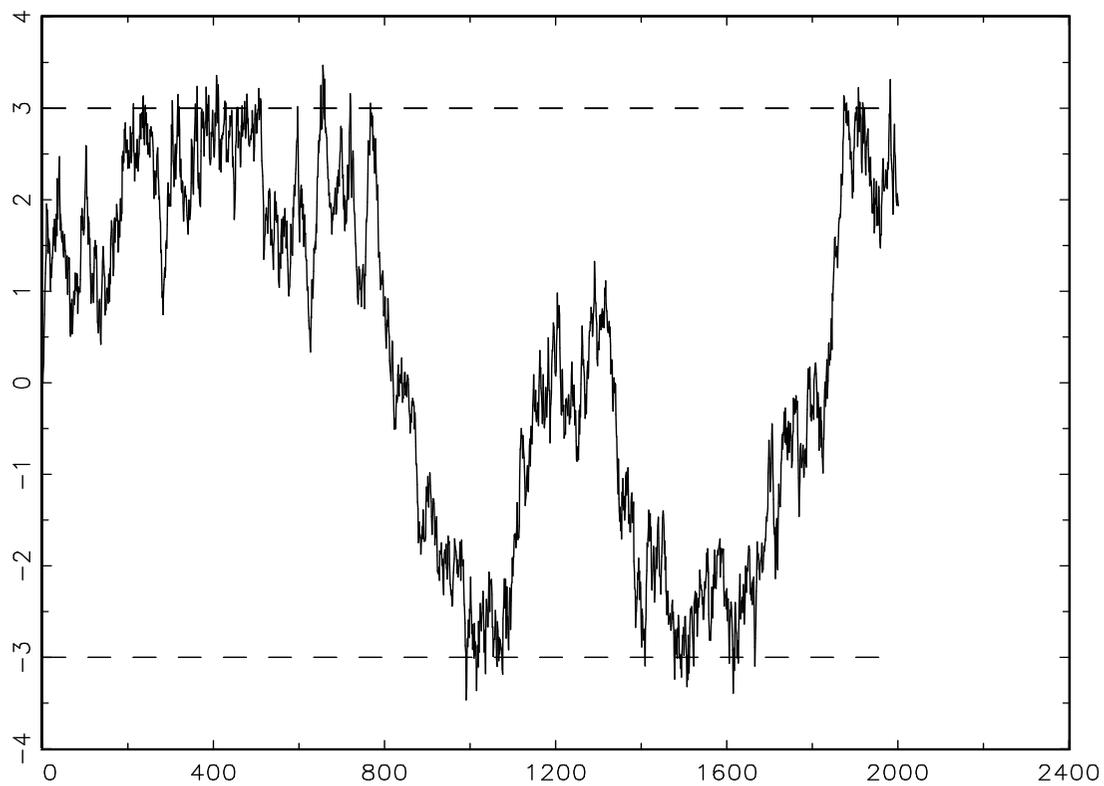
Assuma-se $|\phi_1| < 1$ e $|\phi_3| < 1$. O processo no regime 2 é integrado de ordem 1, no entanto y é estritamente estacionário.

Trajectória simulada a partir do processo

$$y_t = \begin{cases} 0.9y_{t-1} + u_t & y_{t-1} < -3 \\ y_{t-1} + u_t & -3 \leq y_{t-1} \leq 3 \\ 0.9y_{t-1} + u_t & y_{t-1} > 3 \end{cases}$$

onde $\{u_t\}$ é uma sequência de v.a. i.i.d. com distribuição $N(0, 0.2^2)$.

Não é fácil distinguir um passeio aleatório de um processo estacionário. Discuta a potência do teste Dickey-Fuller ...



Trajectória Simulada

5.4.5 Estimação

Considere-se, sem perda de generalidade, o modelo com dois regimes:

$$y_t = \begin{cases} \phi_{10} + \phi_{11}y_{t-1} + \dots + \phi_{1p}y_{t-p} + u_t & q_{t-d} \leq \gamma \\ \phi_{20} + \phi_{21}y_{t-1} + \dots + \phi_{2p}y_{t-p} + u_t & q_{t-d} > \gamma \end{cases}$$

onde se assume que $\{u_t\}$ é um ruído branco com $\text{Var}(u_t) = \sigma^2$.

Parâmetros desconhecidos:

$$d, \gamma, \sigma^2, \phi = \begin{pmatrix} \phi'_1 & \phi'_2 \end{pmatrix}'$$

onde $\phi_i = \begin{pmatrix} \phi_{i0} & \phi_{i1} & \dots & \phi_{ip} \end{pmatrix}'$.

Escreva-se o modelo na forma $y_t = \mathbf{x}_t(\gamma)' \phi + u_t$.

Se $\{u_t\}$ é um ruído branco Gaussiano então

$$y_t | F_{t-1} \sim N \left(\mathbf{x}_t(\gamma)' \phi, \sigma^2 \right)$$

MM \Leftrightarrow método dos mínimos quadrados condicionados.

Estimação de ϕ

$$\hat{\phi}(\gamma) = (\mathbf{X}(\gamma)' \mathbf{X}(\gamma))^{-1} \mathbf{X}(\gamma) \mathbf{y}$$

onde $\mathbf{X}(\gamma)$ é a matriz das variáveis “explicativas” endógenas desfasadas (dado γ), tendo por linha genérica o vector-linha $\mathbf{x}_t(\gamma)$ e \mathbf{y} é o vector das observações $\{y_t\}$.

Claro que o estimador $\hat{\phi}(\gamma)$ depende de γ , que é desconhecido.

Exemplo Considere

$$y_t = \begin{cases} \phi_{10} + \phi_{11}y_{t-1} + u_t & \text{se } y_{t-1} \leq 0.1 \\ \phi_{20} + \phi_{21}y_{t-1} + u_t & \text{se } y_{t-1} > 0.1 \end{cases}$$

Note-se que $d = 1$, $\gamma = 0.1$. Suponha que se observa

$$y' = (.2 \quad -.2 \quad -.1 \quad .5 \quad .3 \quad .05)$$

(na prática não se têm tão poucas observações). Como se perde a primeira observação, devido à estrutura autoregressiva do processo, o vector y a ser usado efectivamente na estimação é

$$y' = (y_2 \quad y_3 \quad y_4 \quad y_5 \quad y_6) = (-.2 \quad -.1 \quad .5 \quad .3 \quad .05).$$

A matriz \mathbf{X} é

$$\mathbf{X} = \begin{bmatrix} I_{\{y_1 \leq .1\}} & y_1 I_{\{y_1 \leq .1\}} & I_{\{y_1 > .1\}} & y_1 I_{\{y_1 > .1\}} \\ I_{\{y_2 \leq .1\}} & y_2 I_{\{y_2 \leq .1\}} & I_{\{y_2 > .1\}} & y_2 I_{\{y_2 > .1\}} \\ I_{\{y_3 \leq .1\}} & y_3 I_{\{y_3 \leq .1\}} & I_{\{y_3 > .1\}} & y_3 I_{\{y_3 > .1\}} \\ I_{\{y_4 \leq .1\}} & y_4 I_{\{y_4 \leq .1\}} & I_{\{y_4 > .1\}} & y_4 I_{\{y_4 > .1\}} \\ I_{\{y_5 \leq .1\}} & y_5 I_{\{y_5 \leq .1\}} & I_{\{y_5 > .1\}} & y_5 I_{\{y_5 > .1\}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & .2 \\ 1 & -.2 & 0 & 0 \\ 1 & -.1 & 0 & 0 \\ 0 & 0 & 1 & .5 \\ 0 & 0 & 1 & .3 \end{bmatrix}.$$

Estimação de γ

Seja

$$\hat{\sigma}^2(\gamma) = \frac{\sum_{t=1}^n \left(y_t - \mathbf{x}_t(\gamma)' \hat{\phi}(\gamma) \right)^2}{n}$$

a soma dos quadrados dos resíduos. Supondo d conhecido, toma-se para estimativa de γ o valor que minimiza $\hat{\sigma}^2(\gamma)$, i.e.,

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} \hat{\sigma}^2(\gamma).$$

Sejam $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ são as estatística de ordem tais que

$$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}.$$

Como $\hat{\sigma}^2(\gamma)$ não se altera quando γ varia entre duas estatísticas de ordem, o problema de otimização é equivalente a

$$\hat{\gamma} = \arg \min_{\gamma \in \tilde{\Gamma}} \hat{\sigma}^2(\gamma), \quad \tilde{\Gamma} = \{y_{(1)}, \dots, y_{(n)}\}.$$

Estimação de d

Como $\hat{\sigma}^2$ depende de d , escolhe-se d num certo conjunto $D = \{1, 2, \dots, d^*\}$ de forma a minimizar $\hat{\sigma}^2$. Logo,

$$(\hat{\gamma}, \hat{d}) = \arg \min_{\gamma \in \tilde{\Gamma}, d \in D} \hat{\sigma}^2(\gamma, d)$$

onde, agora,

$$\hat{\sigma}^2(\gamma, d) = n^{-1} \sum_{t=1}^n \left(y_t - \mathbf{x}_t(\gamma)' \hat{\phi}(\gamma, d) \right)^2$$

e

$$\hat{\phi}(\gamma, d) = \left(\mathbf{X}(\gamma, d)' \mathbf{X}(\gamma, d) \right)^{-1} \mathbf{X}(\gamma, d) \mathbf{y}.$$

Em suma, os passos na estimação dos parâmetros são o seguintes:

1. para cada $\gamma \in \hat{\Gamma}$ e para cada $d \in D$, construir $\mathbf{X}(\gamma, d)$ e estimar $\hat{\phi}(\gamma, d)$ e $\hat{\sigma}^2(\gamma, d)$;
2. tomar para estimativa de γ e de d o par (γ, d) que minimiza $\hat{\sigma}^2(\gamma, d)$; seja esse par dado por $(\hat{\gamma}, \hat{d})$.
3. tomar para estimativa de ϕ , o valor $\hat{\phi}(\hat{\gamma}, \hat{d})$.

Inferência

Pode mostrar-se que o estimador para d é super-consistente. Podemos, para efeitos de inferência estatística sobre os demais parâmetros, assumir que d é conhecido.

Inferência sobre ϕ

Sob certas condições (incluindo estacionaridade estrita) tem-se

$$\sqrt{n} \left(\hat{\phi}(\hat{\gamma}) - \phi_0 \right) \xrightarrow{d} N \left(\mathbf{0}, I(\phi_0)^{-1} \right),$$

Em grandes amostras, a inferência estatística sobre os parâmetros ϕ_{ij} conduz-se da forma habitual.

Inferência sobre a existência de Threshold

Teste $H_0 : \phi_1 = \phi_2$. Sob esta hipótese, o TAR reduz-se ao AR.

Não é possível ensaiar H_0 , no quadro do teste de Wald: pois sob a hipótese nula o parâmetro γ não identificável.

Hansen (2000) sugere o seguinte procedimento:

1. obter $F_n = n \left(\tilde{\sigma}_n^2 - \hat{\sigma}_n^2 \right) / \hat{\sigma}_n^2$ onde $\tilde{\sigma}_n^2$ é a variância dos erros de regressão do modelo AR (sob H_0).
2. simular u_t^* , $t = 1, \dots, n$ com distribuição i.i.d. $N(0, 1)$;
3. $y_t^* = u_t^*$;
4. fazer a regressão de y_t^* sobre $x_t' = \left(1 \quad y_{t-1} \quad \dots \quad y_{t-p} \right)$ e obter $\tilde{\sigma}_n^{*2}$

5. obter $\hat{\gamma} = \arg \min_{\gamma \in \tilde{\Gamma}} \hat{\sigma}_n^{*2}(\gamma)$ onde $\hat{\sigma}_n^{*2}(\gamma)$ resulta da regressão de y_t^* sobre $\mathbf{x}_t(\gamma)$;
6. obter $F_n^* = n (\tilde{\sigma}_n^{*2} - \hat{\sigma}_n^{*2}) / \hat{\sigma}_n^{*2}$
7. repetir os passos 2-6 B vezes
8. valor-p = percentagem de vezes (em B) em que $F_n^* \geq F_n$.

A fundamentação do teste baseia-se na prova segundo a qual F_n^* possui a mesma distribuição assintótica de F_n , sob H_0 . F_n^* é construída sob a hipótese de ausência de *threshold*. Assim, existe indício de violação da hipótese nula quando $\tilde{\sigma}_n^2$ é significativamente superior a $\hat{\sigma}_n^2$, de tal forma que F_n tende a ser geralmente alto quando comparado com F_n^* . Nestas circunstâncias raramente ocorre $F_n^* \geq F_n$ (nas B simulações) e o valor-p vem baixo.

5.4.6 Exemplo

Taxa desemprego EUA Jan/1948 - Jul/2011

$$\Delta y_t = \phi_{10} + \phi_{11} \Delta y_{t-1} + u_t$$

```
-----
RESULTS AR(P)
-----
missing data found - using listwise deletion
Valid cases:           761      Dependent variable:           Y
Missing cases:         1       Deletion method:           Listwise
Total SS:              35.213   Degrees of freedom:         759
R-squared:             0.015    Rbar-squared:              0.014
Residual SS:          34.683   Std error of est:          0.214
F(1,759):             11.603   Probability of F:          0.001

Variable      Estimate      Standard      Prob      Standardized      Cor with
              Estimate      Error        >|t|      Estimate           Dep Var
-----
CONSTANT     0.006031     0.007754     0.777836   0.437              ---
X1           0.122458     0.035951     3.406287   0.001              0.122706
```

dmax=5; p=1; alfa=.1;

$$\Delta y_t = \begin{cases} \phi_{10} + \phi_{11}\Delta y_{t-1} + u_t, & \Delta y_{t-d} \leq \gamma \\ \phi_{20} + \phi_{21}\Delta y_{t-1} + u_t, & \Delta y_{t-d} > \gamma \end{cases}$$

RESULTS TAR

d optimal = 2.0000000
g optimal = 0.0000000
min sqr/n = 0.041530392 ==> Std error of est = (sqr/n)^.5 = 0.20379007
Fobs = 74.611716

Valid cases:	760	Dependent variable:	Y
Missing cases:	0	Deletion method:	None
Total SS:	35.176	Degrees of freedom:	756
R-squared:	0.103	Rbar-squared:	0.099
Residual SS:	31.563	Std error of est:	0.204
F(4,756):	21.633	Probability of F:	0.000

Variable	Estimate	Standard Error	t-value	Prob > t	Standardized Estimate	Cor with Dep Var
X1	-0.029502	0.009359	-3.152167	0.002	-0.108702	-0.104061
X2	-0.122878	0.049211	-2.496957	0.013	-0.086107	-0.080248
X3	0.054338	0.012297	4.418757	0.000	0.153778	0.186662
X4	0.318124	0.048741	6.526791	0.000	0.227141	0.249403

Teste $H_0 : \phi_1 = \phi_2$. p-value = 0

5.5 Modelo Markov-Switching

5.5.1 Introdução

Considere-se o modelo

$$y_t = c_1 + \delta_1 d_t + \phi y_{t-1} + u_t$$

onde d_t é uma variável *dummy*

$$d_t = \begin{cases} 1 & \text{se evento A ocorre} \\ 0 & \text{se o evento A não ocorre.} \end{cases}$$

Sabemos exactamente quando o evento A ocorre. Suponhamos agora que A não é observável. Como modelar, estimar e prever y ? São estas as questões que abordaremos ao longo deste ponto.

$N \rightarrow n^o$ regimes. Por exemplo, $N = 2$

$$y_t = \begin{cases} c_1 + \phi y_{t-1} + u_t & \text{se } y_t \text{ está no regime 1} \\ c_1 + \delta_1 + \phi y_{t-1} + u_t & \text{se } y_t \text{ está no regime 2} \end{cases}$$

Questão: quando é que y está no regime 1 ou 2?

Seja S_t o regime em que o processo y se encontra no momento t . No modelo anterior, que valores pode S assumir?

$\{S_t\}$ é claramente um processo estocástico com espaço de estados (finito) $E = \{1, 2, \dots, N\}$.

Modelos para S_t ? Admita-se $E = \{1, 2\}$.

- $\{S_t\}$ é uma sequência de v.a. i.i.d. com distribuição de Bernoulli de parâmetro p (convençionalmente chamar-se “sucesso”, por exemplo, a $S_t = 1$);
- $\{S_t\}$ é uma sequência de v.a. dependentes

Exemplo?

5.5.2 Cadeias de Markov em tempo discreto com espaço de estados discretos finito

Modelo para $\{S_t\}$ num cenário de dependência temporal: cadeia de Markov homogénea em tempo discreto com espaço de estados discretos finito.

Processo de Markov

Uma cadeia de Markov é um processo de Markov no seguinte sentido:

$$P(S_t = j | S_{t-1} = i, \dots, S_0 = i_0) = P(S_t = j | S_{t-1} = i)$$

isto é, S_t é independente de $S_{t-2}, S_{t-3}, \dots, S_0$ dado S_{t-1} .

Matriz de Probabilidades de Transição

Matriz de probabilidades de transição a um passo no caso de dois regimes:

$$\mathbf{P} = \begin{bmatrix} P(S_t = 1 | S_{t-1} = 1) & P(S_t = 2 | S_{t-1} = 1) \\ P(S_t = 1 | S_{t-1} = 2) & P(S_t = 2 | S_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}.$$

No caso geral tem-se

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{bmatrix}.$$

A matriz \mathbf{P} respeita as condições:

- $0 \leq p_{ij} \leq 1$, $(i, j = 1, \dots, N)$;
- $\sum_{j=1}^N p_{ij} = 1$, $(i = 1, \dots, N)$ (as linhas soma 1).

Exemplo Convencione-se: $S_t = 1 \rightarrow$ “economia em recessão no momento t ”; $S_t = 2 \rightarrow$ “economia em expansão no momento t ”. Suponha-se:

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}.$$

Discutir ...

Podemos estar interessados em calcular uma probabilidade de transição a dois passos, por exemplo,

$$P(S_t = 1 | S_{t-2} = 2) = \dots = \text{elemento } (2, 1) \text{ da matriz } P^2 = P \times P.$$

Calcule-se a matriz P^2 . Tem-se:

$$\begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.55 & 0.45 \\ 0.3 & 0.7 \end{bmatrix}.$$

Assim, $P(S_t = 1 | S_{t-2} = 2) =$

O resultado geral é dado pelas designadas equações de Chapman-Kolmogorov:

$$P(S_{m+n} = j | S_0 = i) = \sum_{k=1}^{\#E} P(S_m = k | S_0 = i) P(S_{m+n} = j | S_m = k).$$

Deduz-se da expressão anterior que

$$P(S_{n+h} = j | S_n = i) = \text{elemento } (i, j) \text{ da matriz } P^h.$$

P: Matriz de probabilidades homogénea. Homogénea porquê?

Distribuição Estacionária

No exemplo anterior, é interessante saber qual é a probabilidade da economia se encontrar em expansão (ou recessão) no longo prazo, dado que no momento presente a economia se encontra em expansão. Tendo em conta a observação anterior, esta probabilidade corresponde ao elemento (2, 2) da matriz P^h quando $h \rightarrow \infty$.

Observe-se

$$\begin{aligned} \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}^{10} &= \begin{bmatrix} 0.40059 & 0.59941 \\ 0.39961 & 0.60039 \end{bmatrix} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}^{20} &= \begin{bmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{bmatrix} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}^{40} &= \begin{bmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{bmatrix}. \end{aligned}$$

É intuitivo admitir-se que

$$\begin{aligned} \lim_{n \rightarrow \infty} P(S_{t+n} = 2 | S_t = 2) &= \\ \lim_{n \rightarrow \infty} P(S_{t+n} = 2 | S_t = 1) &= \end{aligned}$$

Como obter $\lim_{h \rightarrow \infty} \mathbf{P}^h$ sem termos de calcular \mathbf{P}^2 , \mathbf{P}^3 , etc.?

Sob certas condições pode-se provar a proposição seguinte.

Proposição Se S é recorrente positiva aperiódica com espaço de estados finito $\{1, 2, \dots, N\}$ então, o vector (linha) das probabilidades estacionárias

$$\pi = \left[\pi_1 \quad \pi_2 \quad \cdots \quad \pi_N \right]$$

onde

$$\begin{aligned} \pi_i &= \lim_{n \rightarrow \infty} P(S_{t+n} = i | S_t = j) \text{ (para qualquer } j) \\ &= P(S_t = i) \geq 0 \end{aligned}$$

satisfaz as equações

$$\pi = \pi \mathbf{P}, \quad \sum_{i=1}^N \pi_i = 1.$$

$\{\pi_i; i = 1, 2, \dots, N\}$ é a distribuição estacionária da cadeia S .

Exemplo Considere-se

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}.$$

Considerando $\pi = \pi\mathbf{P}$, obtém-se, depois de algumas contas,

$$\begin{cases} \pi_1 = 0.4 \\ \pi_2 = 0.6. \end{cases}$$

De uma forma geral as probabilidades estacionárias associadas a

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix} \quad 0 < p_{ii} < 1.$$

são dadas por

$$\pi_1 = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \quad \pi_2 = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}.$$

5.5.3 Modelos Markov-Switching

Um exemplo com dois regimes $E = \{1, 2\}$:

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + \sigma_1 \varepsilon_t & \text{se } S_t = 1 \\ c_2 + \phi_2 y_{t-1} + \sigma_2 \varepsilon_t & \text{se } S_t = 2 \end{cases}$$

onde S_t é uma cadeia de Markov homogénea (escondida) com

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}.$$

Uma representação equivalente é

$$y_t = c_1 + (c_2 - c_1) I_{\{S_t=2\}} + \left(\phi_1 + (\phi_2 - \phi_1) I_{\{S_t=2\}} \right) y_{t-1} \\ + \left(\sigma_1 + (\sigma_2 - \sigma_1) I_{\{S_t=2\}} \right) \varepsilon_t.$$

Outra representação sugestiva:

$$y_t = c(S_t) + \phi(S_t) y_{t-1} + \sigma(S_t) \varepsilon_t$$

onde

$$c(S_t) = \begin{cases} c_1 & \text{se } S_t = 1 \\ c_2 & \text{se } S_t = 2 \end{cases}, \quad \phi(S_t) = \begin{cases} \phi_1 & \text{se } S_t = 1 \\ \phi_2 & \text{se } S_t = 2 \end{cases}, \quad \sigma(S_t) = \begin{cases} \sigma_1 & \text{se } S_t = 1 \\ \sigma_2 & \text{se } S_t = 2. \end{cases}$$

5.5.4 Função densidade de probabilidade de y

Vamos obter a fdp $f(y_t | F_{t-1})$ que é essencial no âmbito da estimação, inferência e previsão.

Retome-se o modelo

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + \sigma_1 \varepsilon_t & \text{se } S_t = 1 \\ c_2 + \phi_2 y_{t-2} + \sigma_2 \varepsilon_t & \text{se } S_t = 2. \end{cases}$$

Se $\varepsilon_t \sim N(0, 1)$ então:

$$f(y_t | F_{t-1}, S_t = 1) = f_{1t} = \frac{1}{\sqrt{\sigma_1^2 2\pi}} e^{-\frac{1}{2\sigma_1^2} (y_t - c_1 - \phi_1 y_{t-1})^2}$$

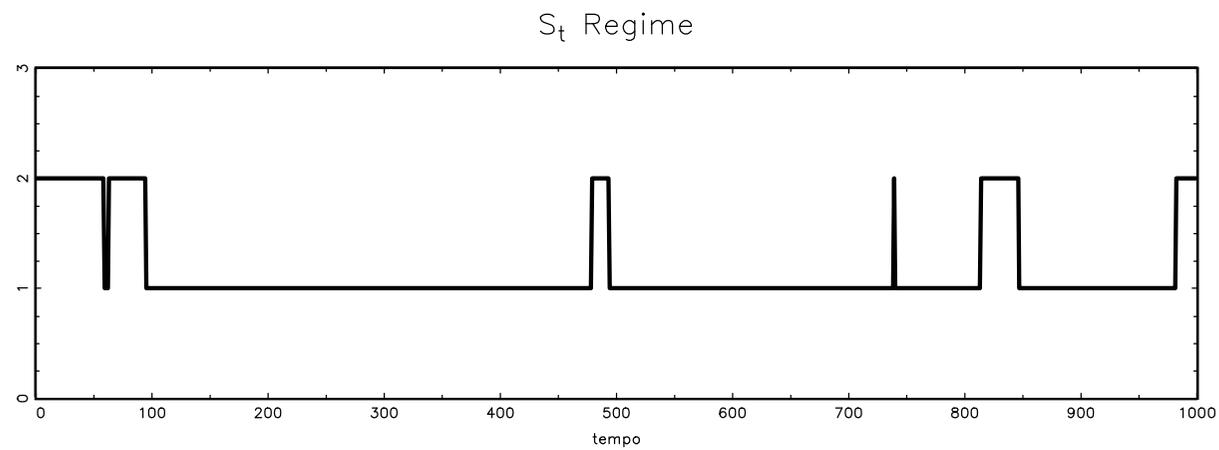
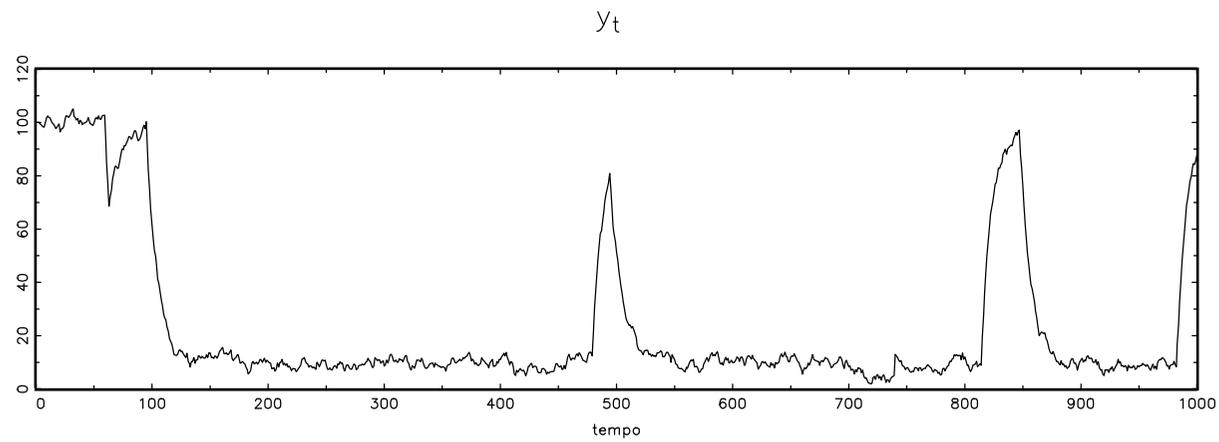
$$f(y_t | F_{t-1}, S_t = 2) = f_{2t} = \frac{1}{\sqrt{\sigma_2^2 2\pi}} e^{-\frac{1}{2\sigma_2^2} (y_t - c_2 - \phi_2 y_{t-1})^2}.$$

Como obter $f(y_t | F_{t-1})$?

$$f(y_t | F_{t-1}) = f_{1t} P(S_t = 1 | F_{t-1}) + f_{2t} P(S_t = 2 | F_{t-1})$$

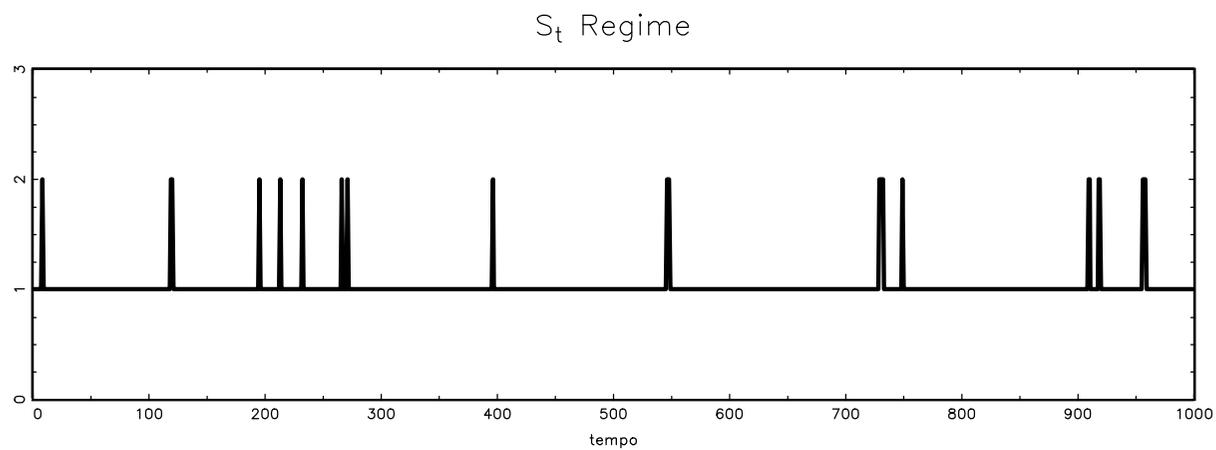
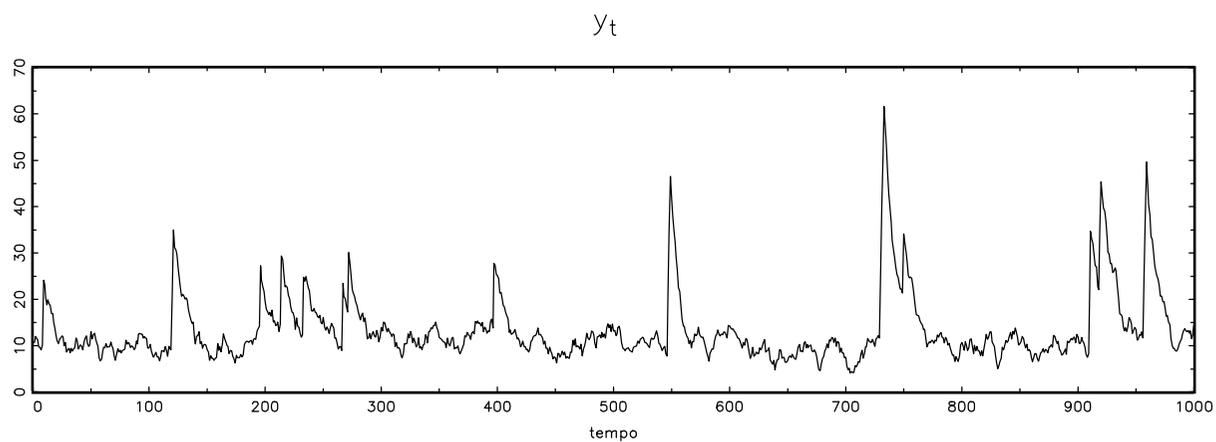
Modelo 1

```
/* S=1 */ c1=1;phi1=.9;sigma1=1;p11=0.99;  
/* S=2 */ c2=10;phi2=0.9; sigma2=1;p22=0.99;
```



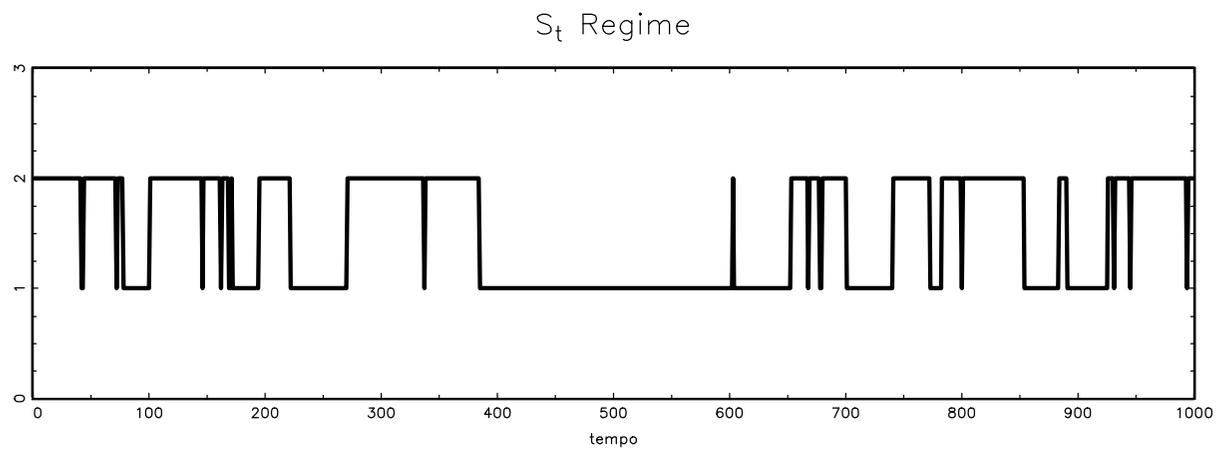
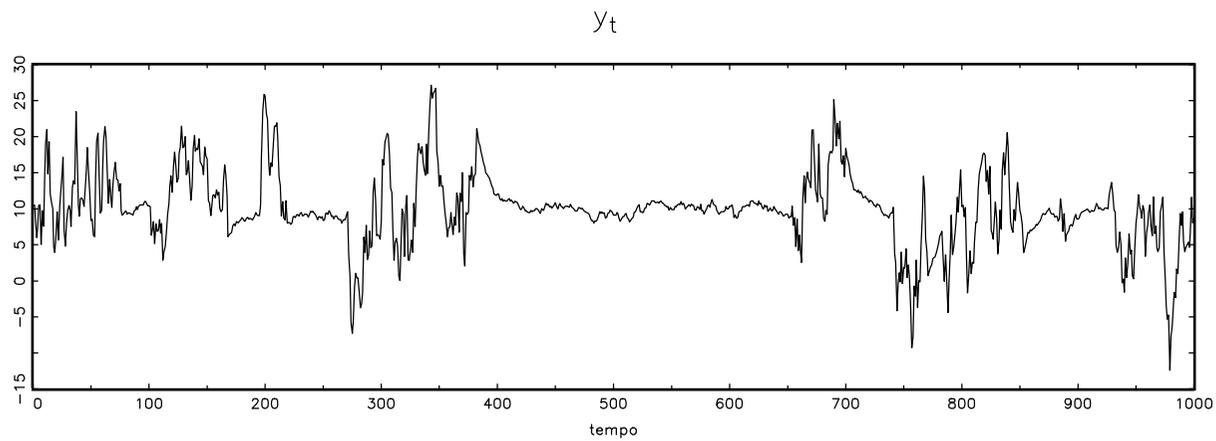
Modelo 2

```
/* S=1 */ c1=1;phi1=.9;sigma1=1;p11=0.99;  
/* S=2 */ c2=15;phi2=0.9; sigma2=1;p22=0.5;
```



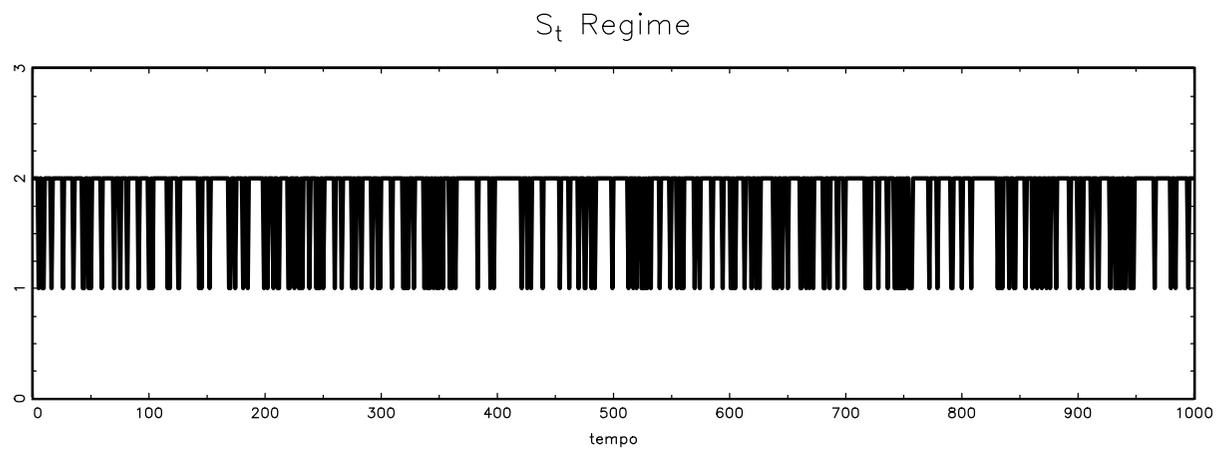
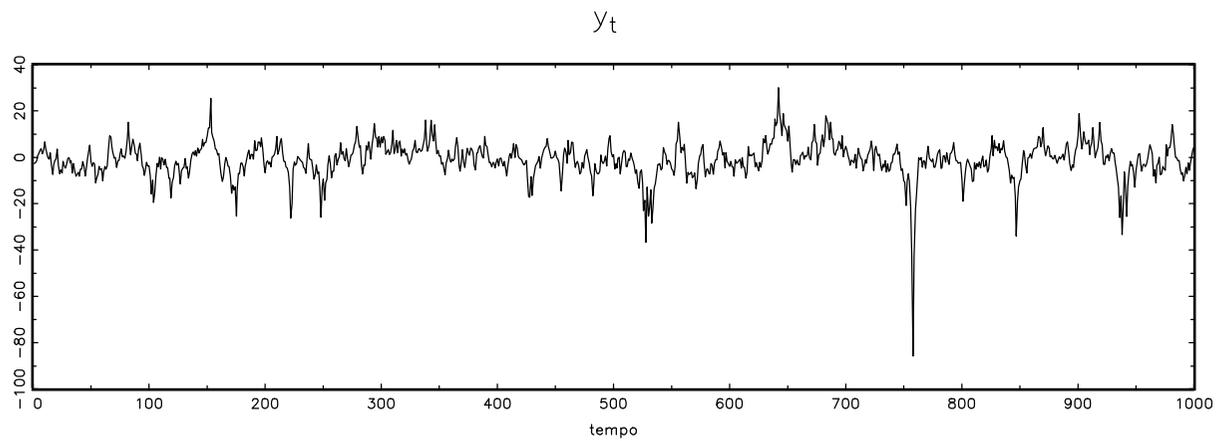
Modelo 3

```
/* S=1 */ c1=1;phi1=0.9;sigma1=.1;p11=0.99;  
/* S=2 */ c2=1;phi2=0.9; sigma2=10;p22=0.99;
```



Modelo 4

```
/* S=1 */ c1=0;phi1=2;sigma1=.1;p11=0.1;  
/* S=2 */ c2=0;phi2=0.6; sigma2=10;p22=0.8;
```



5.5.5 Probabilidades Associadas aos Regimes

É necessário caracterizar $P(S_t = i | F_{t-1})$, $i = 1, 2, \dots, N$

Regimes Independentes

O caso de regimes independentes é um caso particular da cadeia de Markov:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{11} & p_{12} \end{bmatrix} \quad (\text{linhas iguais})$$

$$P(S_t = i | F_{t-1}) = P(S_t = i) = p_{ii}.$$

Nota: No EVIEWS a opção é switching type:simple, Probability regressors: c

Regimes Seguem uma Cadeia de Markov (apenas o caso $N = 2$)

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

Pode-se provar:

$$P(S_t = 1 | F_{t-1}) = p_{11}P(S_{t-1} = 1 | F_{t-1}) + p_{21}P(S_{t-1} = 2 | F_{t-1})$$

A expressão anterior depende de $P(S_{t-1} = i | F_{t-1})$. Ver livro.

Nota: No EVIEWS a opção é switching type: Markov, Probability regressors: c

Regimes Seguem uma Cadeia de Markov Não homogénea (apenas o caso $N = 2$)

Para simplificar suponha-se: $P(S_t = j | S_{t-1} = i, F_{t-1}) = P(S_t = j | S_{t-1} = i, y_{t-1})$

$$P(S_t = 1 | S_{t-1} = 1, F_{t-1}) = \frac{e^{\alpha_0 + \alpha_1 y_{t-1}}}{1 + e^{\alpha_0 + \alpha_1 y_{t-1}}}$$

$$P(S_t = 2 | S_{t-1} = 2, F_{t-1}) = \frac{e^{\beta_0 + \beta_1 y_{t-1}}}{1 + e^{\beta_0 + \beta_1 y_{t-1}}}$$

$$P_t = \begin{bmatrix} p_{11t} & p_{12t} \\ p_{21t} & p_{22t} \end{bmatrix} = \begin{bmatrix} \frac{e^{\alpha_0 + \alpha_1 y_{t-1}}}{1 + e^{\alpha_0 + \alpha_1 y_{t-1}}} & 1 - \frac{e^{\alpha_0 + \alpha_1 y_{t-1}}}{1 + e^{\alpha_0 + \alpha_1 y_{t-1}}} \\ 1 - \frac{e^{\beta_0 + \beta_1 y_{t-1}}}{1 + e^{\beta_0 + \beta_1 y_{t-1}}} & \frac{e^{\beta_0 + \beta_1 y_{t-1}}}{1 + e^{\beta_0 + \beta_1 y_{t-1}}} \end{bmatrix}$$

Resulta

$$P(S_t = 1 | F_{t-1}) = p_{11t}P(S_{t-1} = 1 | F_{t-1}) + p_{21t}P(S_{t-1} = 2 | F_{t-1})$$

Nota: No EVIEWS a opção é switching type: Markov, Probability regressors: c y(-1)

5.5.6 Informação Adicional sobre as Probabilidades Associadas aos Regimes

Informação disponível no EVIEWS:

- One-step ahead predictions

$$P(S_t = i | F_{t-1})$$

- Filtered

$$P(S_t = i | F_t)$$

- Smoothed

$$P(S_t = i | F_T)$$

5.5.7 Estacionaridade

Proposição Considere-se $y_t = c(S_t) + \phi_1(S_t)y_{t-1} + u_t$ com dois regimes $N = 2$. Sejam π_1 e π_2 as probabilidades estacionárias da cadeia de Markov S . Se $\{u_t\}$ é um processo EE com segundo momento finito e $|\phi_1|^{\pi_1} |\phi_2|^{\pi_2} < 1$ então y é EE.

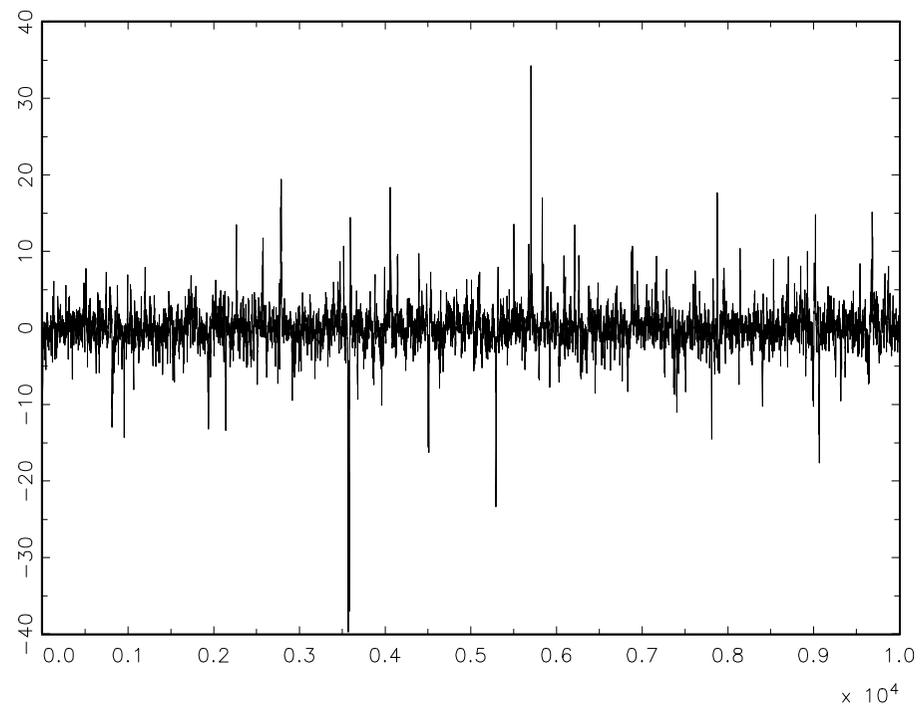
Exemplo Discutir a estacionaridade do seguinte processo:

$$y_t = \begin{cases} 2y_{t-1} + u_t & \text{se } S_t = 1 \\ 0.6y_{t-1} + u_t & \text{se } S_t = 2 \end{cases}$$

onde S_t é uma cadeia de Markov com

$$\mathbf{P} = \begin{pmatrix} 0.1 & 0.9 \\ 0.2 & 0.8 \end{pmatrix}$$

Uma trajectória simulado do processo com $u_t \sim N(0, 1)$:



5.5.8 Estimação e Inferência

Considere-se

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + \sigma_1 \varepsilon_t & \text{se } S_t = 1 \\ c_2 + \phi_2 y_{t-2} + \sigma_2 \varepsilon_t & \text{se } S_t = 2 \end{cases}$$

Como habitualmente, o estimador de máxima verosimilhança é dado por

$$\hat{\theta}_n = \arg \max_{\theta} \sum_{t=1}^n l_t(\theta),$$

onde

$$l_t(\theta) = \log f(y_t | F_{t-1}; \theta) = \log(f_{1t} p_{1t} + f_{2t}(1 - p_{1t})).$$

Sob certas condições, pode-se mostrar:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

Os ensaios individuais: procedimento habitual. O ensaio 1 regime versus 2 regimes envolve um problema ...

5.5.9 Previsão

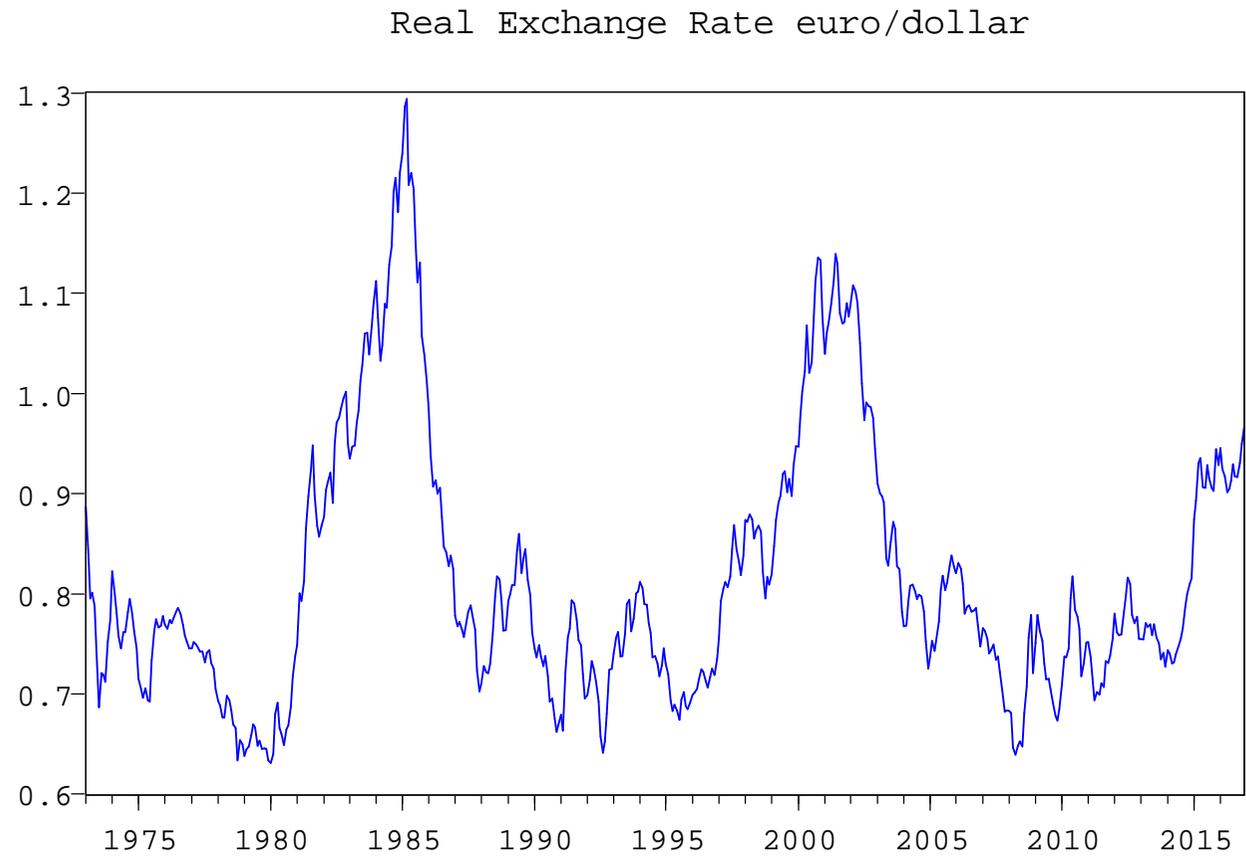
Seja

$$y_t = \begin{cases} c_1 + \phi_1 y_{t-1} + \sigma_1 \varepsilon_t & \text{se } S_t = 1 \\ c_2 + \phi_2 y_{t-1} + \sigma_2 \varepsilon_t & \text{se } S_t = 2. \end{cases}$$

Tem-se

$$\begin{aligned} \mathbf{E}(y_{n+1} | F_n) &= \mathbf{E}(y_{n+1} | F_n, S_{n+1} = 1) P(S_{n+1} = 1 | F_n) \\ &\quad + \mathbf{E}(y_{n+1} | F_n, S_{n+1} = 2) P(S_{n+1} = 2 | F_n) \\ &= (c_1 + \phi_1 y_n) P(S_{n+1} = 1 | F_n) + (c_2 + \phi_2 y_n) P(S_{n+1} = 2 | F_n). \end{aligned}$$

5.5.10 Aplicação



Null Hypothesis: REER has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic - based on SIC, maxlag=18)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.637838	0.0860
Test critical values:		
1% level	-3.441757	
5% level	-2.866464	
10% level	-2.569452	

*MacKinnon (1996) one-sided p-values.

Dependent Variable: REER
 Method: Least Squares
 Sample (adjusted): 1970M02 2016M12
 Included observations: 563 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011079	0.005486	2.019390	0.0439
REER(-1)	0.986416	0.006556	150.4678	0.0000
R-squared	0.975821	Mean dependent var		0.825912
Adjusted R-squared	0.975777	S.D. dependent var		0.133869
S.E. of regression	0.020835	Akaike info criterion		-4.900838
Sum squared resid	0.243524	Schwarz criterion		-4.885444
Log likelihood	1381.586	Hannan-Quinn criter.		-4.894829
F-statistic	22640.55	Durbin-Watson stat		1.442351
Prob(F-statistic)	0.000000			

Dependent Variable: REER
 Method: Markov Switching Regression (BFGS / Marquardt steps)
 Sample (adjusted): 1970M02 2016M12
 Included observations: 563 after adjustments
 Number of states: 2
 Initial probabilities obtained from ergodic solution
 Standard errors & covariance computed using observed Hessian
 Random search: 25 starting values with 10 iterations using 1 standard
 deviation (rng=kn, seed=545557300)
 Convergence achieved after 18 iterations

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Regime 1				
C	0.019506	0.008148	2.393927	0.0167
REER(-1)	0.978568	0.009354	104.6099	0.0000
LOG(SIGMA)	-3.730352	0.039512	-94.40980	0.0000
Regime 2				
C	0.002693	0.005800	0.464316	0.6424
REER(-1)	0.992331	0.007230	137.2530	0.0000
LOG(SIGMA)	-4.562396	0.068751	-66.36140	0.0000
Transition Matrix Parameters				
P11-C	3.898927	0.540226	7.217212	0.0000
P21-C	-3.178058	0.498707	-6.372594	0.0000
Mean dependent var	0.825912	S.D. dependent var	0.133869	
S.E. of regression	0.020819	Sum squared resid	0.241414	
Durbin-Watson stat	1.452359	Log likelihood	1421.740	
Akaike info criterion	-5.022167	Schwarz criterion	-4.960593	
Hannan-Quinn criter.	-4.998130			

Equation: MS01

Transition summary: Constant Markov transition
probabilities and expected durations

Sample (adjusted): 1970M02 2016M12

Included observations: 563 after adjustments

Constant transition probabilities:

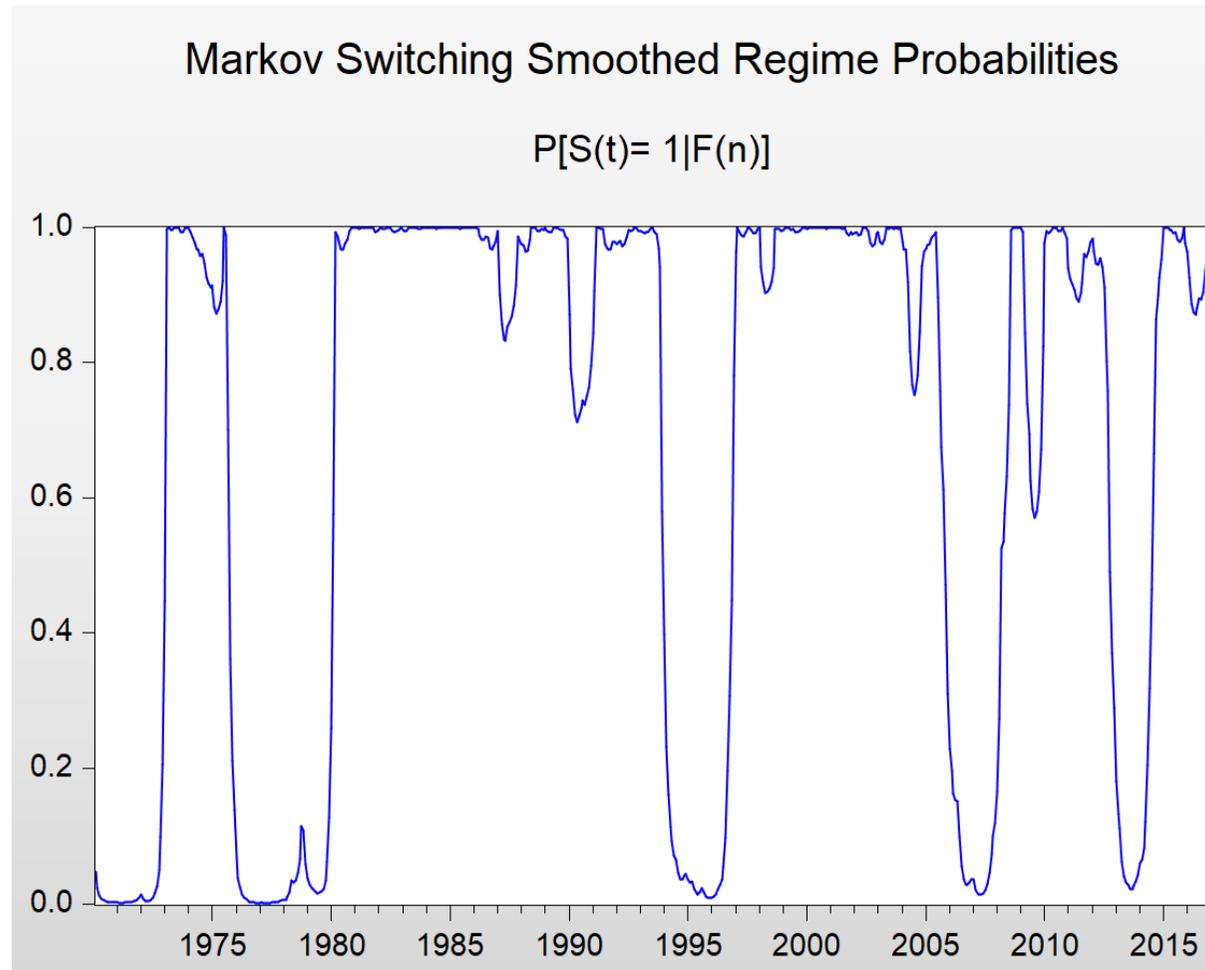
$P(i, k) = P(s(t) = k \mid s(t-1) = i)$

(row = i / column = j)

	1	2
1	0.980139	0.019861
2	0.040000	0.960000

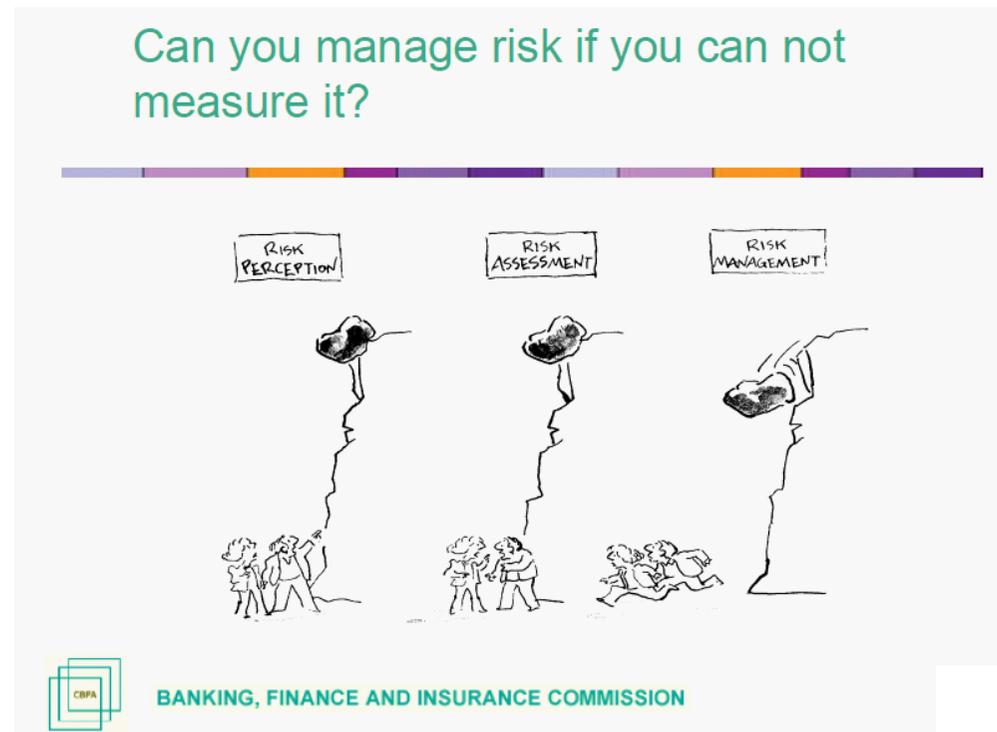
Constant expected durations:

	1	2
	50.34949	25.00010



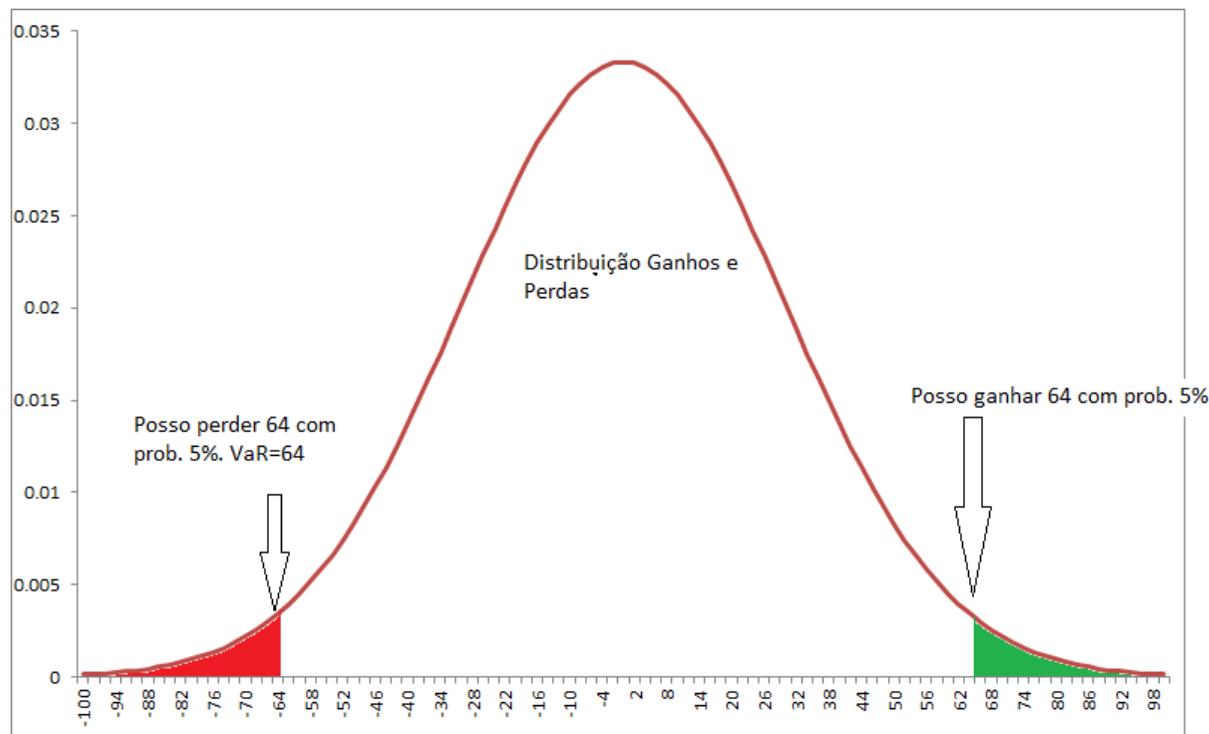
6 Risco de Mercado e o Valor em Risco

6.1 Introdução



Quanto é que podemos perder com probabilidade $\alpha > 0$ num certo período de tempo? Por exemplo, imagine-se que se investe 1000 unidade monetárias. Ao fim de uma semana quanto é que podemos perder com probabilidade α , e.g. $\alpha = 0.01$? Suponhamos que a resposta é 64. Então o Valor em Risco (1 semana, $\alpha = 0.01$) é precisamente 64.

VaR é a perda que pode ocorrer num lapso de tempo determinado, com certa probabilidade α , supondo que o *portfolio* não é gerido durante o período de análise.



Em termos probabilísticos, o VaR é o quantil de ordem α (em módulo) da distribuição teórica de ganhos e perdas (ou variação do capital)

$$V_{n+h} - V_n = \Delta V_{n+1} = \frac{\Delta V_{n+h}}{V_n} V_n = R_{n+h}(h) V_n$$

Onde:

- V_n é o valor do investimento no momento presente
- V_{n+h} é o valor do investimento no futuro (h períodos à frente)
- $R_{n+h}(h) = \frac{\Delta V_{n+h}}{V_n}$ retorno do investimento.

Exemplo Considere-se a compra de 200,000 acções de um título no período n , ao preço de $P_n = 5$ Euros. O investimento é de $V_n = 5 \times 200,000 = 1,000,000$ Euros. Suponha-se que no momento $n + h$ o preço do título passa para $P_{n+h} = 5.5$ Euros. A variação do capital é de $\Delta V_{n+h} = 1,100,000 - 1,000,000 = 100,000$. O retorno (em tempo discreto) é $R_{n+h}(h) = (P_{n+h} - P_n) / P_n = 0.1$. É importante notar que o retorno pode também ser obtido através da expressão $R_{n+h}(h) = \Delta V_{n+h} / V_n$.

Os valores que V vier a assumir no período $n + 1, n + 2, \dots, n + h$ dizem respeito ao período de investimento e, portanto, V_{n+h} , é desconhecido para $h \geq 1$.

O VaR a $100\alpha\%$ baseado na distribuição marginal de ganhos e perdas é o valor VaR tal que

$$P(\Delta V_{n+h} < -VaR) = \alpha.$$

O VaR a $100\alpha\%$ baseado na distribuição condicional é o valor VaR tal que

$$P(\Delta V_{n+h} < -VaR | \mathcal{F}_n) = \alpha.$$

O VaR deve ser baseado na distribuição de ganhos e perdas marginal ou condicional?

Em princípio deve ser baseado na distribuição condicional. Por que não usar \mathcal{F}_n se esta melhora a previsão?

Exemplo Determine o VaR a 5% a dois dias sabendo que $\Delta V_{n+2} | \mathcal{F}_n \sim N(10, 225)$.

Resposta: 14.67

Exemplo Determine o VaR a 5% a dois dias sabendo que $R_{n+2} | \mathcal{F}_n \sim N(0.01, 0.000225)$ e $V_n = 1000$. Resposta: 14.67

Na prática estas distribuições são desconhecidas.

6.2 Abordagem Não Paramétrica

Método 1 (directamente a partir da série ΔV_{n+1})

Tendo em conta

$$P(\Delta V_{n+1} < -VaR) = \alpha$$

tem-se

$$VaR_{n,n+1,\alpha} = -q_{\alpha}^{\Delta V}$$

$$\widehat{VaR}_{n,n+1,\alpha} = -\tilde{q}_{\alpha}^{\Delta V}$$

$\tilde{q}_{\alpha}^{\Delta V}$ é o quantil (estimado) de ordem α da distribuição de ΔV_{n+1} .

Exemplo Ver ficheiro VaR.XLS.

Método 2 (a partir dos retornos)

Por definição tem-se

$$P\left(\Delta V_{n+1} < -VaR_{n,n+1,\alpha}\right) = \alpha$$

ou, como $\Delta V_{n+1} = \frac{\Delta V_{n+1}}{V_n} V_n = R_{n+1} V_n$,

$$P\left(R_{n+1} V_n < -VaR_{n,n+1,\alpha}\right) = \alpha,$$

$$P\left(R_{n+1} < -\frac{VaR_{n,n+1,\alpha}}{V_n}\right) = \alpha,$$

$$P\left(R_{n+1} < q_{\alpha}^R\right) = \alpha.$$

Resulta

$$VaR_{n,n+1,\alpha} = -q_{\alpha}^R V_n$$

$$\widehat{VaR}_{n,n+1,\alpha} = -\tilde{q}_{\alpha}^R V_n$$

onde q_{α}^R é o quantil de ordem α da distribuição de R_{n+1} .

Exercício Considere a série que se encontra no ficheiro `petroleo.wk1` (Crude Oil WTI Cushing U\$/BBL). Calcule o VaR a 5% a um dia para um investimento de 1000 u.m. Use o método não paramétrico.

No caso $h > 1$ ver o manual.

Se a análise incide sobre um *portfolio* linear, na construção da série histórica dos retornos

$$R_{p,t} = \omega_1 R_{1,t} + \omega_2 R_{2,t} + \dots + \omega_n R_{n,t},$$

onde ω_i são os pesos do capital investido no activo i ($\sum_{i=1}^n \omega_i = 1$), os pesos devem permanecer fixos durante todo o período histórico.

Abordagem não paramétrica - desvantagens:

- A distribuição relevante para obter o VaR é a distribuição condicional de R_{n+h} e não a distribuição marginal.
- Quando α é muito baixo por exemplo $\alpha = 0.01$ ou inferior o estimador \tilde{q}_α^R é muito impreciso. Justificação teórica: sob certas condições o quantil empírico \tilde{q}_α (isto, é a estatística de ordem $[n\alpha]$) tem distribuição assintótica dada por

$$\sqrt{n} (\tilde{q}_\alpha - q_\alpha) \xrightarrow{d} N \left(0, \frac{\alpha(1-\alpha)}{(f(q_\alpha))^2} \right).$$

A variância assintótica de \tilde{q}_α “explode” quando α é muito alto (perto de 1) ou muito baixo (perto de zero).

- Até onde coligir os dados? Considerar todo o passado disponível? Ou só o passado recente? A metodologia exposta atribui o mesmo peso a todas as observações.

Nota Adicional: Estimação do VaR a partir de uma distribuição com caudas Pareto

Quadro

6.3 Abordagem Paramétrica

Considere-se novamente a expressão para o caso $h = 1$:

$$\Delta V_{n+1} = \frac{\Delta V_{n+1}}{V_n} V_n = R_{n+1} V_n \simeq r_{n+1} V_n.$$

Esta expressão mostra que a distribuição condicional de $r_{n+1} V_n$ apenas depende da distribuição de r_{n+1} (no momento n , V_n é conhecido). Podemos assim concentrar-nos apenas sobre a distribuição de r_{n+1} .

Assuma-se

$$r_{n+1} | \mathcal{F}_n \sim N(\mu_{n+1}, \sigma_{n+1}^2)$$

Tem-se

$$P(\Delta V_{n+1} < -VaR_{n,n+1,\alpha} | \mathcal{F}_n) = \alpha$$

$$P\left(r_{n+1} < -\frac{VaR_{n,n+1,\alpha}}{V_n} \middle| \mathcal{F}_n\right) = \alpha.$$

Estandarizando r_{n+1}

$$Z_{n+1} = \frac{r_{n+1} - \mu_{n+1}}{\sigma_{n+1}}$$

vem

$$P\left(Z_{n+1} < \frac{-\frac{VaR_{n,n+1,\alpha}}{V_n} - \mu_{n+1}}{\sigma_{n+1}} \middle| \mathcal{F}_n\right) = \alpha$$

$$\frac{-\frac{VaR}{V_n} - \mu_{n+1}}{\sigma_{n+1}} = q_\alpha^Z \Rightarrow \boxed{VaR_{n,n+1,\alpha} = -(\mu_{n+1} + q_\alpha^Z \sigma_{n+1}) V_n}$$

(q_α^Z é o quantil de ordem α da distribuição da variável Z_{n+1}).

No caso $h > 1$ tem-se

$$\begin{aligned}
 \Delta V_{n+h} &= \frac{\Delta V_{n+h}}{V_n} V_n \\
 &= R_{n+h}(h) V_n \\
 &\approx r_{n+h}(h) V_n. \\
 &= (r_{n+1} + r_{n+2} + \dots + r_{n+h}) V_n, \quad r_{n+i} = \log(P_{n+i}/P_{n+i-1})
 \end{aligned}$$

Pode-se mostrar

$$\text{VaR}_{n,n+h,\alpha} = - \left(\mathbf{E}(r_{n+h}(h) | \mathcal{F}_n) + q_\alpha^Z \sqrt{\text{Var}(r_{n+h}(h) | \mathcal{F}_n)} \right) V_n$$

onde q_α^Z é o quantil de ordem α da distribuição de

$$Z = \frac{r_{n+h}(h) - \mathbf{E}(r_{n+h}(h) | \mathcal{F}_n)}{\sqrt{\text{Var}(r_{n+h}(h) | \mathcal{F}_n)}}.$$

Exemplo Determine o VaR a 5% a dois dias sabendo que $R_{n+2} | \mathcal{F}_n \sim N(0.01, 0.000225)$ e $V_n = 1000$. Resposta: 14.67

Na prática é necessário definir um modelo econométrico que nos permita obter uma estimativa de $\mathbf{E}(r_{n+h}(h) | \mathcal{F}_n)$ e de $\text{Var}(r_{n+h}(h) | \mathcal{F}_n)$.

6.3.1 Modelo Gaussiano Simples

O caso mais simples consiste em assumir

$$r_t = c + u_t,$$

onde $\{u_t\}$ é um ruído branco Gaussiano, $u_t \sim N(0, \sigma^2)$. Assim, $r_{n+1} \sim N(c, \sigma^2)$.

No caso $h = 1$, tem-se, pela fórmula anterior

$$VaR_{n,n+1,\alpha} = - \left(c + q_\alpha^Z \sigma \right) V_n$$

onde q_α^Z é o quantil de ordem α da distribuição $N(0, 1)$. No excel: $q_\alpha^Z = \text{NORM.S.INV}(\alpha)$

α	0.001	0.01	0.05
q_α^Z	-3.09	-2.33	-1.64

Estimação:

$$\widehat{VaR}_{n,n+1,\alpha} = - \left(\hat{c} + q_\alpha^Z \hat{\sigma}_r \right) V_n.$$

Exercício Considere a série que se encontra no ficheiro petroleo.wk1 (Crude Oil WTI Cushing U\$/BBL). Calcule o VaR a 5% a um dia para um investimento de 1000 u.m. Use o método paramétrico baseado no modelo Gaussiano simples.

Caso $h > 1$. É necessário deduzir-se a distribuição condicional de $r_{n+h}(h)$. Como, por hipótese, $r_{n+h}(h)$ não depende \mathcal{F}_n , a distribuição condicional coincide com a distribuição marginal. Pelas propriedades habituais da distribuição normal vem

$$r_{n+h}(h) = r_{n+1} + r_{n+2} + \dots + r_{n+h} \sim N(ch, \sigma^2 h)$$

e, portanto, pela aplicação da fórmula vem

$$\begin{aligned} VaR_{n,n+h,\alpha} &= -\left(ch + \sqrt{h}\sigma q_\alpha^Z\right) V_n, \\ \widehat{VaR}_{n,n+h,\alpha} &= -\left(\hat{c}h + \sqrt{h}\hat{\sigma}q_\alpha^Z\right) V_n. \end{aligned}$$

Exercício Considere a série que se encontra no ficheiro petroleo.wk1 (Crude Oil WTI Cushing U\$/BBL). Calcule o VaR a 5% a cinco dias para um investimento de 1000 u.m. Use o método paramétrico baseado no modelo Gaussiano simples.

6.3.2 Modelo RiskMetrics

As hipóteses de normalidade e variância condicional constante, assumidas no modelo anterior, são, como se sabe, bastante limitativas. Modelo RiskMetrics desenvolvido pela J.P. Morgan para o cálculo do VaR:

$$r_t = c + \sigma_t \varepsilon_t, \quad \sigma_t^2 = (1 - \lambda) u_{t-1}^2 + \lambda \sigma_{t-1}^2$$

O modelo assenta na hipótese IGARCH(1,1) com termo constante nulo, $\omega = 0$, e média condicional $\mu_t = c$.

No caso $h \geq 1$ temos de aplicar a fórmula:

$$VaR_{n,n+h,\alpha} = - \left(\mathbf{E} (r_{n+h} (h) | \mathcal{F}_n) + q_{\alpha}^Z \sqrt{\mathbf{Var} (r_{n+h} (h) | \mathcal{F}_n)} \right) V_n.$$

Temos de calcular $\mathbf{E} (r_{n+h} (h) | \mathcal{F}_n)$ e $\mathbf{Var} (r_{n+h} (h) | \mathcal{F}_n)$.

Caso $h \geq 1$,

$$\begin{aligned}
 \mathbf{E} (r_{n+h} (h) | \mathcal{F}_n) &= ch \\
 \mathbf{Var} (r_{n+h} (h) | \mathcal{F}_n) &= \mathbf{Var} (r_{n+1} + \dots + r_{n+h} | \mathcal{F}_n) \\
 &= \mathbf{Var} (r_{n+1} | \mathcal{F}_n) + \dots + \mathbf{Var} (r_{n+h} | \mathcal{F}_n) \\
 &= \mathbf{E} (\sigma_{n+1}^2 | \mathcal{F}_n) + \dots + \mathbf{E} (\sigma_{n+h}^2 | \mathcal{F}_n) \\
 &= h\sigma_{n+1}^2 \text{ (ou } h\sigma_{n+1,n}^2 \text{ - ver previsão IGARCH)}
 \end{aligned}$$

Assim, pela fórmula $VaR_{n,n+h,\alpha} = - \left(\mathbf{E} (r_{n+h} (h) | \mathcal{F}_n) + q_\alpha^Z \sqrt{\mathbf{Var} (r_{n+h} (h) | \mathcal{F}_n)} \right) V_n$ vem

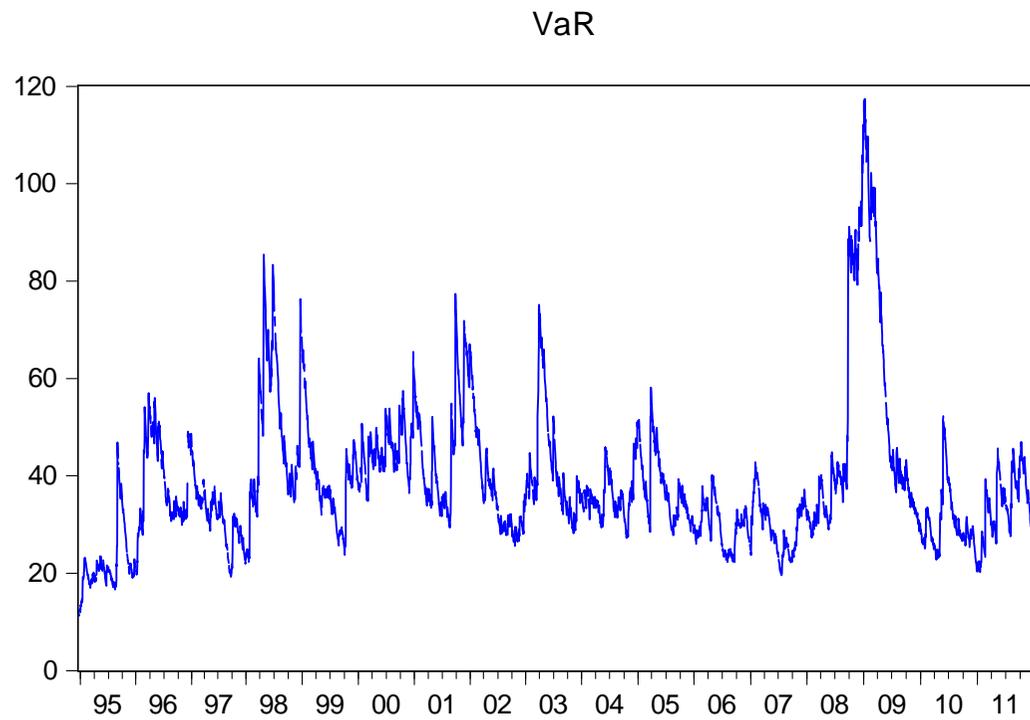
$$VaR_{n,n+h,\alpha} = - \left(ch + q_\alpha^Z \sqrt{h} \sigma_{n+1} \right) V_n.$$

Estimativa:

$$\widehat{VaR}_{n,n+h,\alpha} = - \left(\hat{c}h + q_\alpha^Z \sqrt{h} \hat{\sigma}_{n+1} \right) V_n.$$

Exercício Considere a série que se encontra no ficheiro petroleo.wk1 (Crude Oil WTI Cushing U\$/BBL). Calcule o $\widehat{VaR}_{t,t+1,0.05}$ para $t = 2, \dots, n$ supondo que em cada momento possui 1000 u.m. de petróleo ($h = 1$, $\alpha = 0.05$).

Na figura seguinte apresenta-se o $\widehat{VaR}_{t,t+1,0.05}$, $t = 1, \dots, n$



6.3.3 Modelo GARCH (média condicional constante)

Assuma-se agora

$$\begin{aligned} r_t &= c + u_t \\ u_t &= \sigma_t \varepsilon_t, \quad u_t \sim \text{GARCH}. \end{aligned}$$

Tal como no caso anterior, temos de obter a distribuição de $r_{n+h}(h) | \mathcal{F}_n$. Tem-se

$$\mathbb{E}(r_{n+h}(h) | \mathcal{F}_n) = ch$$

Para obter $\text{Var}(r_{n+h}(h) | \mathcal{F}_n)$ considere-se

$$\begin{aligned} & \text{Var}(r_{n+h}(h) | \mathcal{F}_n) \\ &= \text{Var}(r_{n+1} | \mathcal{F}_n) + \dots + \text{Var}(r_{n+h} | \mathcal{F}_n) \\ &= \sum_{k=1}^h \sigma_{n+k,n}^2 \end{aligned}$$

Tem-se

$$VaR_{n,n+h,\alpha} = - \left(ch + q_{\alpha}^Z \sqrt{\sum_{k=1}^h \sigma_{n+k,n}^2} \right) V_n$$

Exemplo Considere-se o modelo

$$r_t = c + \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Tendo em conta que

$$\sigma_{n+k,n}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{k-1} (\alpha_1 u_n^2 + \beta_1 \sigma_n^2),$$

resulta

$$\begin{aligned}
 & \text{Var} (r_{n+h} (h) | \mathcal{F}_n) \\
 = & \sum_{k=1}^h \sigma_{n+k,n}^2 \\
 = & \sum_{k=1}^h \left(\frac{\omega}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{k-1} (\alpha_1 u_n^2 + \beta_1 \sigma_n^2) \right) \\
 = & \frac{1}{1 - \alpha_1 - \beta_1} \left(h\omega - (\alpha_1 u_n^2 + \beta_1 \sigma_n^2) ((\alpha_1 + \beta_1)^h - 1) \right).
 \end{aligned}$$

Vem

$$\begin{aligned}
 & VaR_{n,n+h,\alpha} \\
 = & - \left(ch + q_\alpha^Z \sqrt{\frac{1}{1 - \alpha_1 - \beta_1} \left(h\omega - (\alpha_1 u_n^2 + \beta_1 \sigma_n^2) ((\alpha_1 + \beta_1)^h - 1) \right)} \right) V_n.
 \end{aligned}$$

Resultados da estimação GARCH dos retornos do Dow Jones (28317 observações diárias).

Dependent Variable: retornos do Dow Jones
Method: ML - ARCH
Included observations: 28317 after adjusting endpoints

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000416	4.54E-05	9.155216	0.0000
Variance Equation				
C	1.17E-06	5.17E-08	22.67097	0.0000
ARCH(1)	0.085080	0.001263	67.38911	0.0000
GARCH(1)	0.905903	0.001542	587.6203	0.0000
R-squared	-0.000450	Mean dependent var		0.000188
Adjusted R-squared	-0.000556	S.D. dependent var		0.010753
S.E. of regression	0.010756	Akaike info criterion		-6.640694
Sum squared resid	3.275823	Schwarz criterion		-6.639529
Log likelihood	94026.27	Durbin-Watson stat		1.921149

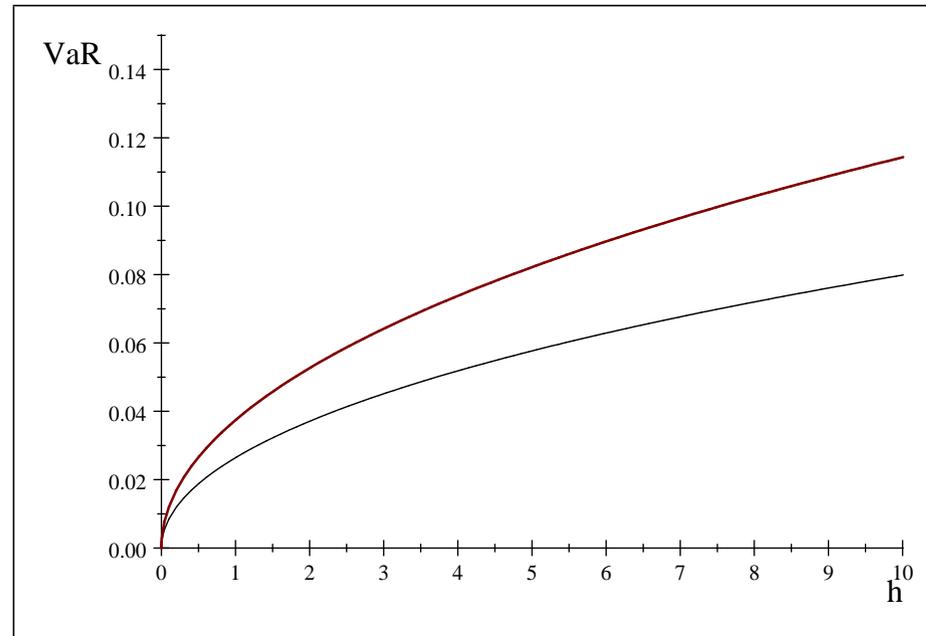
Sabe-se que $r_n = -0.0101$, $\hat{\sigma}_n^2 = 0.00014$.

$VaR_{n,n+h,\alpha}$ estimado, supondo $V_n = 1$, corresponde à expressão:

$$-\left(\hat{c}h + q_\alpha^Z \sqrt{\frac{1}{1 - \hat{\alpha}_1 - \hat{\beta}_1} \left(h\hat{\omega} - (\hat{\alpha}_1 \hat{u}_n^2 + \hat{\beta}_1 \sigma_n^2) \left((\hat{\alpha}_1 + \hat{\beta}_1)^h - 1 \right) \right)}\right)$$

onde $\hat{u}_n = r_n - \hat{c} = -0.0101 - 0.000416 = -0.01051$.

Na figura seguinte representa-se $\widehat{VaR}_{n,n+h,\alpha}$ para $\alpha = 0.01$ e para $\alpha = 0.05$ em função de h .



6.4 Generalização: Portfolio com m Activos

Obtenha-se agora o VaR supondo que o *portfolio* é constituído por m acções. Para simplificar assumamos que

$$\mathbf{r}_t = (R_{1t}, \dots, R_{mt})' \Big| \mathcal{F}_{t-1} \sim \mathbf{N}(\boldsymbol{\mu}_t, \mathbf{H}_t)$$

onde $\boldsymbol{\mu}_t := \mathbf{E}(\mathbf{r}_t | \mathcal{F}_{t-1})$ e $\text{Var}(\mathbf{r}_t | \mathcal{F}_{t-1}) := \mathbf{H}_t$. No momento n o *portfolio* vale V_n , por exemplo, uma alocação de 1 milhão de Euros no título 1 e 2 milhões no título 2, traduz-se por um investimento no valor de $V_n = 3$ milhões, sendo $\omega_1 = 1/3$ desse valor investido no título 1 e $\omega_2 = 2/3$ investido no título 2. A obtenção do VaR é similar ao do caso de um *portfolio* com apenas 1 activo:

$$\begin{aligned} P\left(\Delta V_{n+1} < -VaR_{n,n+1,\alpha} \Big| \mathcal{F}_n\right) &= \alpha \\ P\left(\frac{\Delta V_{n+1}}{V_n} V_n < -VaR_{n,n+1,\alpha} \Big| \mathcal{F}_n\right) &= \alpha \\ P\left(R_{p,n+1} < -\frac{VaR_{n,n+1,\alpha}}{V_n} \Big| \mathcal{F}_n\right) &= \alpha. \end{aligned}$$

Estandarizando $R_{p,n+1}$, isto é, considerando

$$Z_{n+1} = \frac{R_{p,n+1} - \mathbf{E} \left(R_{p,n+1} \mid \mathcal{F}_n \right)}{\sqrt{\mathbf{Var} \left(R_{p,n+1} \mid \mathcal{F}_n \right)}}$$

vem

$$P \left(Z_{n+1} < \frac{-\frac{VaR_{n,n+1,\alpha}}{V_n} - \mathbf{E} \left(R_{p,n+1} \mid \mathcal{F}_n \right)}{\sqrt{\mathbf{Var} \left(R_{p,n+1} \mid \mathcal{F}_n \right)}} \mid \mathcal{F}_n \right) = \alpha$$

$$\frac{-\frac{VaR_{n,n+1,\alpha}}{V_n} - \mathbf{E} \left(R_{p,n+1} \mid \mathcal{F}_n \right)}{\sqrt{\mathbf{Var} \left(R_{p,n+1} \mid \mathcal{F}_n \right)}} = q_\alpha^Z$$

$$\Rightarrow VaR_{n,n+1,\alpha} = - \left(\mathbf{E} \left(R_{p,n+1} \mid \mathcal{F}_n \right) + q_\alpha^Z \sqrt{\mathbf{Var} \left(R_{p,n+1} \mid \mathcal{F}_n \right)} \right) V_n$$

Tendo em conta que $R_{p,n+1} = \sum_{i=1}^m \omega_i R_{i,n+1} = \boldsymbol{\omega}' \mathbf{r}_{n+1}$ tem-se

$$R_{p,n+1} = \mathbf{E} \left(R_{p,n+1} \middle| \mathcal{F}_n \right) = \boldsymbol{\omega}' \boldsymbol{\mu}_{n+1},$$

e

$$\text{Var} \left(R_{p,n+1} \middle| \mathcal{F}_n \right) = \boldsymbol{\omega}' \mathbf{H}_{n+1} \boldsymbol{\omega}.$$

Finalmente

$$\boxed{VaR_{n,n+1,\alpha} = - \left(\boldsymbol{\omega}' \boldsymbol{\mu}_{n+1} + q_{\alpha}^Z \sqrt{\boldsymbol{\omega}' \mathbf{H}_{n+1} \boldsymbol{\omega}} \right) V_n}$$

onde q_{α}^Z é o quantil de ordem α da distribuição de $Z_{n+1} \middle| \mathcal{F}_n$.

Exemplo Considere-se um *portfolio*, no momento n , constituído por de 1 milhão de Euros no título 1 e 2 milhões no título 2. Admita-se a seguinte distribuição

$$\begin{pmatrix} R_{1,n+1} \\ R_{2,n+1} \end{pmatrix} \Big| \mathcal{F}_n \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.01 & 0.002 \\ 0.002 & 0.005 \end{pmatrix} \right).$$

Tem-se para $\alpha = 0.05$

$$VaR_{n,n+1,\alpha} = 1.645 \sqrt{\begin{pmatrix} 1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 0.01 & 0.002 \\ 0.002 & 0.005 \end{pmatrix} \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}} \times 3 = 0.32 \text{ milhões.}$$

O valor em risco com uma probabilidade de 0.05 para um horizonte temporal de $h = 1$ período é de cerca de 0.32 milhões de euros.

Se considerássemos os dois activos separadamente teríamos:

$$VaR \text{ do título 1} = 1.645 \sqrt{0.01} \times 1 = 0.164$$

$$VaR \text{ do título 2} = 1.645 \sqrt{0.005} \times 2 = 0.232.$$

Observa-se que a soma dos VaR individuais, $0.164 + 0.232 = 0.396$, é maior do que o VaR do *portfolio*.

6.5 Avaliação do VaR (Backtesting)

Neste ponto avalia-se a qualidade da estimativa proposta para o VaR. Esta avaliação é importante por várias razões. As empresas (sobretudo bancos) que usam o VaR são pressionadas interna e externamente (por directores, auditores, reguladores, investidores) para produzirem VaR precisos. Um VaR preciso é fundamental na gestão e controle do risco e na alocação de capital. Por essa razão é essencial que empresa teste regularmente as suas medidas de risco, na linha das recomendações adoptadas em acordo internacionais. Por outro lado, embora a definição de VaR seja muito precisa e objectiva, existem diferentes métodos de estimação do VaR, que produzem diferentes estimativas (algumas bastante díspares); por isso, é importante identificar a melhor abordagem para o problema concreto em análise.

6.5.1 Cobertura Marginal e Condicional

Como $\widehat{VaR}_{t-1,t,\alpha}$ é construído no período $t-1$ não se sabe, antecipadamente, se no período t se tem ou não $\Delta V_t < -\widehat{VaR}_{t-1,t,\alpha}$. Intuitivamente, é natural esperar que a desigualdade

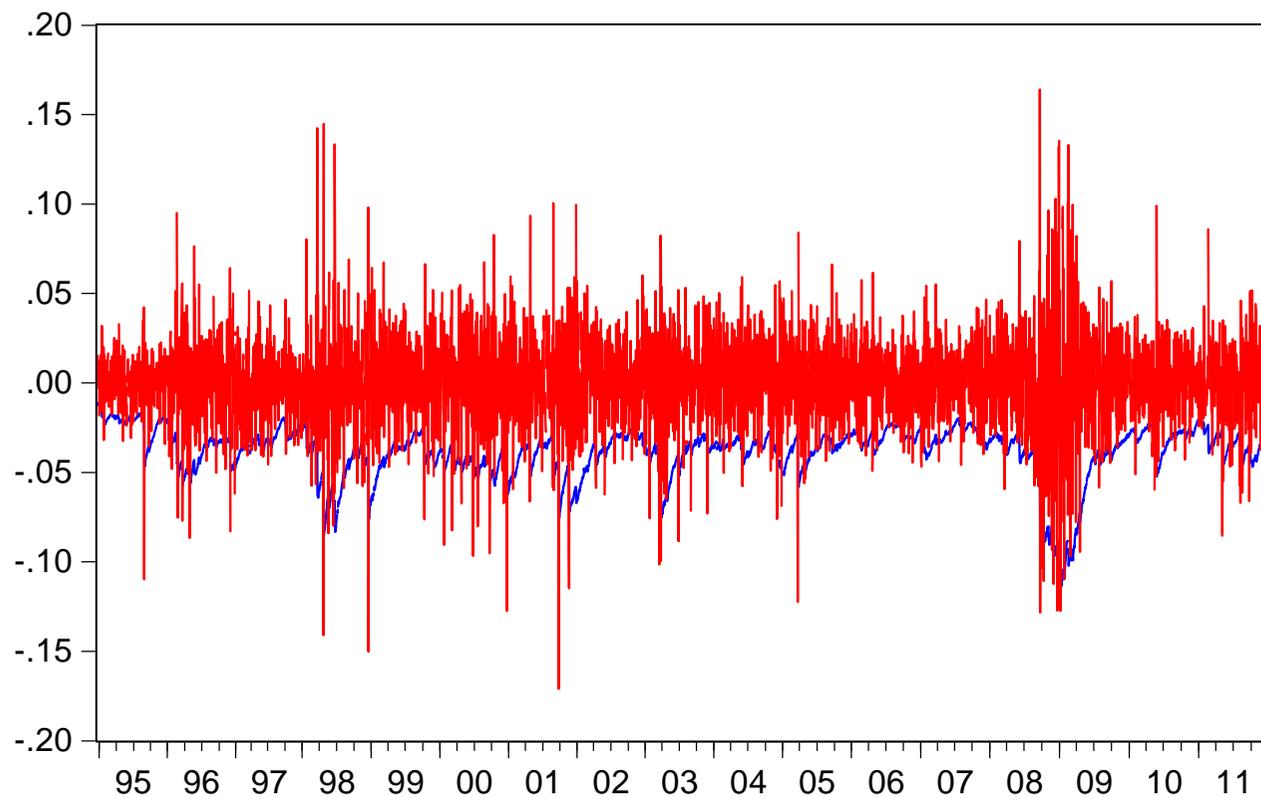
$$\Delta V_t < -\widehat{VaR}_{t-1,t,\alpha} \Leftrightarrow r_t < -\frac{\widehat{VaR}_{t-1,t,\alpha}}{V_{t-1}}$$

ocorra $\alpha 100\%$ das vezes. Por outras palavras, seja

$$I_t = \begin{cases} 1 & \text{se } \Delta V_t < -\widehat{VaR}_{t-1,t,\alpha} \text{ ou } r_t < -\frac{\widehat{VaR}_{t-1,t,\alpha}}{V_{t-1}} \\ 0 & \text{no caso contrário.} \end{cases}$$

É natural esperar que a % de uns da sequência I_t seja de $\alpha 100\%$.

Na figura seguinte representa-se r_t e $-VaR_{t-1,t,\alpha}/V_{t-1}$ (dados ficheiro:petroleo; metodologia do cálculo do VaR: RiskMetrics)



A previsão do VaR produz uma cobertura marginal correcta se

$$P(I_t = 1) = \alpha \Leftrightarrow \mathbf{E}(I_t) = \alpha.$$

A previsão do VaR produz uma cobertura condicional correcta se

$$P(I_t = 1 | \mathcal{F}_{t-1}) = \alpha \Leftrightarrow \mathbf{E}(I_t | \mathcal{F}_{t-1}) = \alpha.$$

Nota: $\mathbf{E}(I_t | \mathcal{F}_{t-1}) = \alpha \Rightarrow \mathbf{Cov}(I_t, I_{t-k}) = 0$.

Por que razão a cobertura condicional é importante?

6.5.2 Teste Estatístico

Suponha-se que I_t segue uma cadeia de Markov com a seguinte matriz de probabilidades de transição

$$\mathbf{P} = \begin{pmatrix} 1 - p_{01} & p_{01} \\ 1 - p_{11} & p_{11} \end{pmatrix}$$

onde $p_{ij} = P(I_t = j | I_{t-1} = i)$.

Se $p_{01} = p_{11} \Rightarrow \{I_t\}$ é uma sucessão de v.a. i.i.d. (com distribuição de Bernoulli). Logo a hipótese a testar é

$H_0: p_{01} = p_{11} = \alpha$ [A previsão do VaR produz uma cobertura condicional e marginal correcta]. Sob H_0 pode-se provar:

$$RV = -2 \log \frac{(1 - \alpha)^{n_{00} + n_{10}} \alpha^{n_{01} + n_{11}}}{(1 - \hat{p}_{01})^{n_{00}} \hat{p}_{01}^{n_{01}} (1 - \hat{p}_{11})^{n_{10}} \hat{p}_{11}^{n_{11}}} \xrightarrow{d} \chi^2_{(2)}.$$

n_{ij} é o número de vezes em que I passou de i para j

Exemplo Estimou-se o $VaR_{t-1,t,\alpha}$, com $\alpha = 0.05$ (dados ficheiro:petroleo; metodologia do cálculo do VaR: RiskMetrics). Teste

$$H_0 : p_{01} = p_{11} = 0.05$$

Sabe-se:

Freq. Absolutas I_t			
	$I_t = 0$	$I_t = 1$	Total
$I_{t-1} = 0$	3772	218	3990
$I_{t-1} = 1$	218	16	234
Total	3990	234	4224

\hat{P}		
	$I_t = 0$	$I_t = 1$
$I_{t-1} = 0$	$\frac{3772}{3990} = 0.945$	$\frac{218}{3990} = 0.055$
$I_{t-1} = 1$	$\frac{218}{234} = 0.932$	$\frac{16}{234} = 0.068$

$$\begin{aligned}
 RV_{obs} &= -2 \log \frac{(1 - \alpha)^{n_{00} + n_{10}} \alpha^{n_{01} + n_{11}}}{(1 - \hat{p}_{01})^{n_{00}} \hat{p}_{01}^{n_{01}} (1 - \hat{p}_{11})^{n_{10}} \hat{p}_{11}^{n_{11}}} \\
 &= -2 \log \frac{(1 - 0.05)^{3990} 0.05^{234}}{(1 - 0.055)^{3772} (0.055)^{218} (1 - 0.068)^{218} (0.068)^{16}} \\
 &= 3.244
 \end{aligned}$$

$$pvalue = P(RV > 3.244 | H_0) = 0.197$$

Os testes baseados na cadeia de Markov baseiam-se em dependências de primeira ordem. Ora, pode suceder que I_t dado I_{t-1} dependa ainda de I_{t-2} . Neste caso, os testes baseados no modelo da cadeia de Markov, podem perder bastante potência.

A hipótese $E(I_t | \mathcal{F}_{t-1}) = \alpha$ sugere que I_t dado \mathcal{F}_{t-1} apenas depende de uma constante, α . Assim, se I_t depende de alguma variável \mathcal{F}_{t-1} mensurável, a hipótese $E(I_t | \mathcal{F}_{t-1}) = \alpha$ deve ser rejeitada. Procedimento alternativo aos testes anteriores:

$$I_t = \alpha_0 + \sum_{i=1}^k \alpha_i I_{t-i} + \mathbf{x}'_{t-1} \boldsymbol{\beta} + u_t$$

onde \mathbf{x}'_{t-1} é um vector $1 \times m$ de variáveis \mathcal{F}_{t-1} mensuráveis e $\boldsymbol{\beta}$ é um vector de parâmetros $m \times 1$. A hipótese de correcta cobertura marginal e condicional envolve o ensaio $H_0 : \alpha_1 = 0, \dots, \alpha_k = 0, \boldsymbol{\beta} = \mathbf{0}, \alpha_0 = \alpha$. Naturalmente a estatística F pode ser usada.