21.4. PROCESSOS ESTOCÁSTICOS INTEGRADOS

O passeio aleatório é apenas um caso particular de uma classe de processos estocásticos conhecidos como **processos integrados**. Já vimos que o passeio aleatório é um processo não estacionário, mas sua 1ª. diferença é estacionária. Dizemos que ele é um processo integrado de ordem 1, denotado I(1).

Em geral, uma série temporal é integrada de ordem d, e escrevemos Y_t ~ I(d) se a série é não estacionária mas pode ser transformada numa série estacionária após d diferenças. Uma série estacionária é dita "integrada de ordem zero", e escrevemos Y_t ~ I(0).

Grande parte das séries temporais em Econometria é I(1), ou seja, pode ser feita estacionária após a aplicação de uma diferença.

Propriedades das séries integradas

Sejam X_t, Y_t e Z_t séries temporais.

1) Se X_t é I(0) (é estacionária) e Y_t é I(1), então $Z_t = X_t + Y_t$ é I(1).
2) Se X_t é I(d) então $Z_t = aX_t + b$ é I(d) também, onde a e b são constantes. Em particular, se X_t é estacionária, uma combinação linear dela também é estacionária.
3) Se X_t é $I(d_1)$ e Y_t é $I(d_2)$, então $Z_t = aX_t + bY_t$ é $I(d_2)$, onde $d_1 < d_2$. Ou seja, a combinação linear de duas séries integradas de ordens diferentes é integrada com a maior ordem dentre as duas.
4) Se X_t é I(d) e Y_t é I(d), ou seja, as duas séries são integradas com a mesma ordem, então $Z_t = aX_t + bY_t$ é $I(d^*)$, onde $d^* \leq d$. Este resultado será particularmente útil quando discutirmos a idéia de cointegração, onde uma combinação linear de duas séries I(1) poderá ser uma série estacionária (isto é, I(0)).

21.5. REGRESSÃO ESPÚRIA

Uma questão importante em Economia é determinar o comportamento de longo prazo de variáveis econômicas. Especificamente, a questão de interesse é: o efeito de um choque sobre a variável de interesse é temporário ou permanente? No primeiro caso, a série reverte para sua média de longo prazo após o choque e no segundo, o efeito do choque é permanente, indicando que a variável se comporta como um passeio aleatório (*random walk*), e a média e/ou variância da série apresentam uma dependência no tempo.

A questão é importante não apenas do ponto de vista econômico mas também tem impactos sobre as previsões, pois no caso de séries estacionárias uma previsão de longo prazo da série deverá convergir para a média incondicional da mesma. No primeiro caso, a tendência da série pode ser removida diretamente, enquanto no segundo é necessário diferenciar a série para torná-la estacionária.

Além disso, se Y_t e X_t são duas séries não estacionárias do tipo passeio aleatório, a regressão de uma variável na outra leva a resultados inconsistentes, a chamada "**regressão espúria"** (Granger e Newbold (1974)), em que os testes de significância convencionais apontam a existência de relações entre as variáveis que, de fato, inexistem.

Freqüentemente regressões espúrias são caracterizadas por um alto R^2 e um baixo valor da estatística de Durbin-Watson. Para evitar o problema, a solução recomendada era trabalhar com as séries diferenciadas (ao invés das séries em nível). No entanto, esta solução traz outros problemas, pois elimina a tendência das séries e esconde as relações de longo prazo entre as variáveis envolvidas. Isso traduzia uma controvérsia entre econometristas e analistas de séries temporais. Aqueles, em geral, argumentavam que as diferenças sucessivas aplicadas às séries de interesse poderiam levar a problemas de má-especificação dos modelos, e dificultar a interpretação de possíveis relações de equilíbrio de longo prazo entre as variáveis. Já os seguidores de Box e Jenkins eram favoráveis à diferenciação das variáveis e à aplicação de modelos de regressão às séries diferenciadas.

Regressão espúria é um fenômeno que acontece na prática quanto tentamos fazer a regressão de uma série não estacionária em outra série, também não estacionária. A regressão espúria é uma regressão "sem sentido", em que observamos uma significância estatística que, na verdade, não existe. Além disso, este fenômeno ocorre até mesmo em grandes amostras, e assim aumentar o tamanho das séries não resolve o problema.

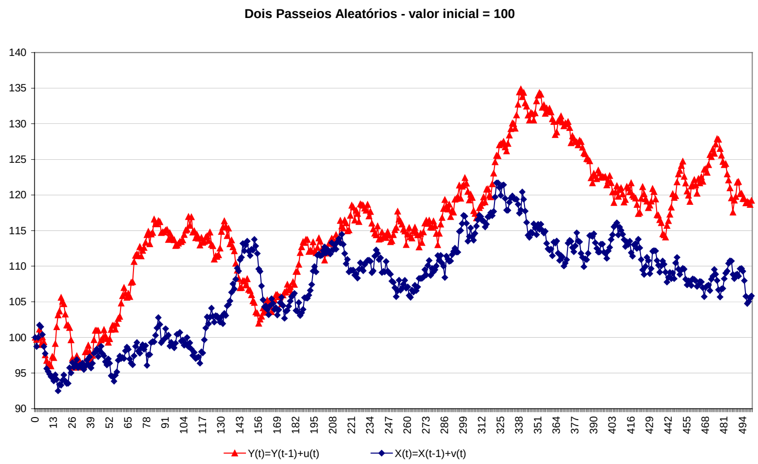
Exemplo – simulação – criando uma regressão espúria

Neste exemplo criamos duas séries de erros iid N(0,1) com 500 valores cada. Estas séries serão usadas para gerar dois passeios aleatórios, ambos com valor inicial 100, de acordo com as equações:

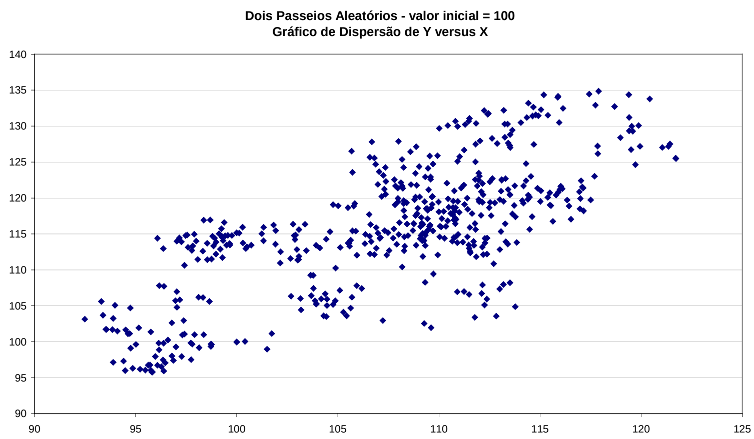
$$Y_t = Y_{t-1} + u_t \qquad (21.7.1)$$

$$X_t = X_{t-1} + v_t \qquad (21.7.2)$$

O gráfico das duas séries ao longo do tempo é mostrado a seguir.



O gráfico de dispersão das duas séries mostra que, se rodarmos uma regressão de Y contra X, é bem possível encontrarmos valores significantes para o coeficiente angular, indicando uma relação entre as variáveis que, de fato, não existe.



A regressão de Y em X foi ajustada, levando aos seguintes resultados:

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21699,290	1	21699,290	590,630	,000(a)
	Residual	18332,876	499	36,739		
	Total	40032,166	500			

Termo	Coeficiente	Erro Padrão	Estat. t	Significância
CONST	10,618	4,328	2,454	0,986
X	0,977	0,040	24,317	1,000

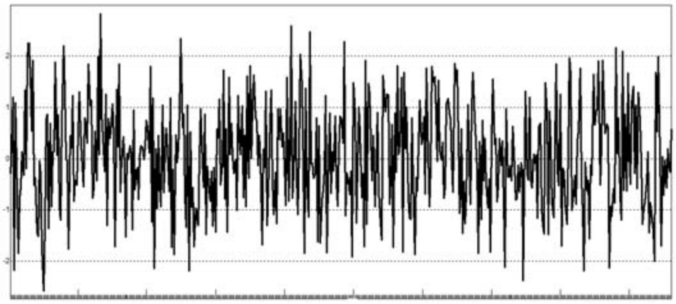
R-square 0,5423 **Adjusted R-square 0,5414**

Durbin-Watson 0,05411

O coeficiente de X é altamente significante, o R^2 é bastante alto, e a estatística F da regressão é altamente significante também. Tudo indicaria que existe uma forte relação entre as duas variáveis, mas a priori não deveria existir qualquer relação, pois as duas séries são apenas passeios aleatórios. **Como você poderia suspeitar que existe algo errado? Olhe para a Durbin-Watson – o seu valor é extremamente pequeno. Segundo Granger e Newbold, desconfie da regressão e suspeite que a regressão é espúria se $R^2 > d$, onde d é a estatística de Durbin-Watson.**

Este exemplo ilustra o fenômeno da **regressão espúria.** Se fizermos a regressão de duas séries não estacionárias, podemos em algumas situações (como a do exemplo) observar uma alta correlação entre as séries.

Se, agora, ajustamos a regressão nas primeiras diferenças (e não mais em nível), ou seja, ΔY versus ΔX, os resultados não serão significativos. O gráfico da série Y após a primeira diferença é mostrado a seguir – é um ruído branco. A série X diferenciada tem comportamento análogo.



Os resultados da regressão ΔY versus ΔX são:

Termo	Coeficiente	Erro Padrão	Estat. t	Significância
_CONST	0,038123	0,046389	0,821822	0,588821 <-
_XDIFF	0,028639	0,046145	0,620630	0,465157 <-

Ambos os coeficientes não são significantes. Também, R^2 = 0,0007729, virtualmente zero, e agora a estatística de Durbin-Watson é d = 1,958, bem próxima de 2, indicando que os resíduos deste modelo não exibem autocorrelação de lag 1.

21.6. TESTES DE ESTACIONARIEDADE

Se a estacionariedade de uma série é uma questão tão importante, como podemos descobrir se uma série é ou não estacionária?

Existem duas estratégias principais:

1) Métodos gráficos
2) Teste do correlograma

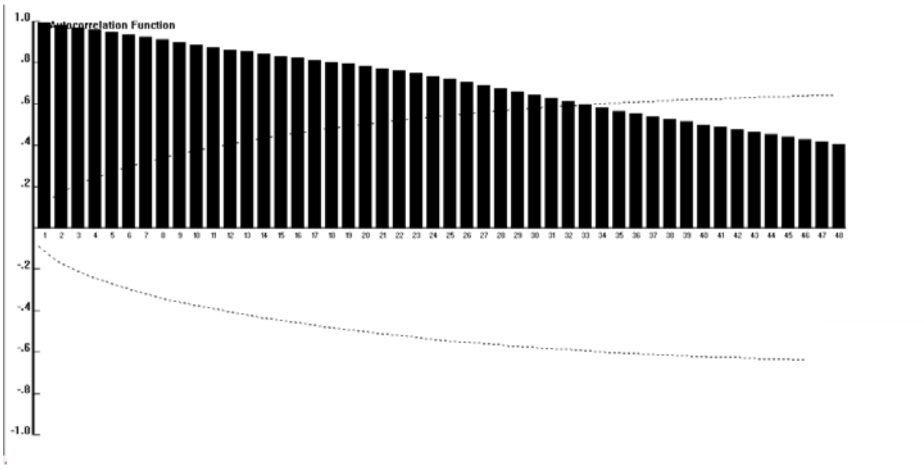
O teste de raiz unitária, que verifica formalmente se uma série é estacionária, será explicado na próxima seção.

A aplicação de métodos gráficos é bastante simples – já comentamos que o primeiro passo na análise de uma série temporal é fazer seu gráfico em relação ao tempo. Uma série **não será** estacionária se sua média e/ou variância se alterarem ao longo do tempo. Nos exemplos iniciais deste capítulo vimos que as séries econômicas apresentadas não tinham médias constantes, e portanto não eram estacionárias.

Função de autocorrelação e correlograma

A função de correlação (ACF ou FAC) e seu gráfico em relação aos "lags" (correlograma) podem ser usados para verificar se uma série é ou não estacionária.

A próxima figura mostra um exemplo típico do **correlograma de uma série não estacionária** – é a ACF da random walk simulada Y_t. Note como a ACF decai lentamente.



Ao diferenciarmos a série da random walk, obtemos uma série estacionária. O gráfico da sua ACF é mostrado a seguir:

Note que os valores das autocorrelações estão dentro do intervalo de confiança (existem alguns poucos fora do intervalo, mas "escapam" do IC por pouco, e são autocorrelações espúrias). Este gráfico indica a ACF de um processo puramente aleatório (um ruído branco). Não existem (na prática), autocorrelações significantes em todos os "lags".

Os gráficos anteriores são, naturalmente, calculados a partir das funções de autocorrelação amostrais (vide definição 21.9), ou seja, são os correlogramas amostrais.

Significância estatística dos coeficientes de correlação amostrais

Para verificar se as autocorrelações amostrais são significantes, usamos a **fórmula de Bartlett**. Lembre-se que as autocorrelações amostrais são, para **grandes amostras**, **aproximadamente** distribuídas como: $r_k \sim N\left(0, \frac{1}{n}\right)$ onde n é o tamanho da série. Logo, o IC 95% para a k-ésima autocorrelação é: $r_k \pm 1{,}96\sqrt{\frac{1}{n}}$. Se este intervalo inclui zero, não rejeitamos a hipótese de que ρ_k seja igual a zero.

Uma outra abordagem é o teste de significância conjunta das m primeiras autocorrelações. Este teste é feito através da **estatística Q de Box e Pierce**, definida como:

$$Q = n\sum_{k=1}^{m} \hat{\rho}_k^2 \qquad \textbf{(21.8.8)}$$

Onde n = tamanho da série, m = número de defasagens que se quer testar. Para grandes amostras, a estatística Q é aproximadamente Qui-Quadrado com m graus de liberdade. Rejeita-se a hipótese nula $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0$ se Q for grande, ou seja, se Q > $q_{\alpha,m}$ onde $q_{\alpha,m}$ é o percentil (1-α)% da distribuição Qui-Quadrado com m graus de liberdade.

Uma modificação da estatística de Box e Pierce é a estatística de Ljung e Box, definida como:

$$\text{LB} = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right) \sim \chi^2 m \qquad \textbf{(21.8.9)}$$

A estatística de **Ljung e Box é também assintoticamente distribuída como uma Qui-Quadrado com m graus de liberdade. Mas, a estatística LB tem melhores propriedades que a estatística de Box e Pierce.**

A estatística LB tem maior potência que a Box-Pierce, ou seja, maior probabilidade de rejeitar a hipótese nula, no caso de pequenas amostras.