Transitividade da Causalidade e da Cointegração

Investigação com o Método de Monte Carlo

Lúcia Paiva Martins de Sousa

Mestrado em Análise de Dados e Sistemas de Apoio à Decisão

Orientador: Pedro Cosme da Costa Vieira

Faculdade de Economia do Porto

2006

Transitividade da Causalidade e da Cointegração

Investigação com o Método de Monte Carlo

Lúcia Paiva Martins de Sousa

Mestrado em Análise de Dados e Sistemas de Apoio à Decisão

Orientador: Pedro Cosme da Costa Vieira

Faculdade de Economia do Porto

2006

Breve nota biográfica

A autora é natural do Brasil e tem cidadania portuguesa.

Licenciou-se em Matemática, ramo de especialização científica em Matemática Aplicada em 1990.

Foi assistente estagiária entre 1991 e 1995 e assistente convidada entre 1995 e 1999 na Faculdade de Economia da Universidade do Porto, leccionando Informática.

É equiparada a assistente do segundo triénio na Escola Superior de Tecnologia do Instituto Politécnico de Viseu.

Tem publicado um artigo na Revista Portuguesa de Economia Regional (vol. 10, pp. 5-14) e um *working paper* na FEP (vol. 193).

Resumo

Neste trabalho propomo-nos estudar a existência de transitividade entre variáveis que, em termos estatísticos, estão relacionadas. Em particular, vamos estudar a existência de transitividade entre variáveis ligadas por causalidade de Granger e entre variáveis cointegradas. Em termos intuitivos será de aceitar a existência de transitividade.

Sendo que, no contexto do problema que queríamos estudar, a modelização matemática e consequente manipulação algébrica se nos pareceu complexa e de aplicação limitada, adoptamos como metodologia a experimentação estatística que é conhecida na literatura como Método de Monte Carlo. No sentido de enquadrarmos este método, apresentamos no capítulo 2 deste trabalho, numa perspectiva histórica, um resumo dos trabalhos dos principais autores pioneiros.

Dos resultados obtidos, concluímos que, ao contrário do que parecia intuitivo, será de rejeitar a conjectura de que existe transitividade entre relações mesmo que estas sejam estatisticamente significativas.

Índice

Índice de tabelas

1. Introdução

Sendo que o Mundo é um todo muito complexo, então a sua compreensão está para além da capacidade humana. No entanto, se considerarmos o Mundo dividido em fenómenos parcelares modelizáveis como variáveis estatísticas (processo de análise), podemos compreender o Mundo como um conjunto de relações de causa efeito entre as variáveis observadas (processo de síntese). Assim sendo, podemos dizer que, em termos genéricos, a compreensão do Mundo começará pela observação parcelar dos fenómenos ou acontecimentos naturais, passará pela construção de variadas hipóteses de relacionamento entre as parcelas observadas e termina na avaliação destas hipóteses explicativas com vista à sua rejeição ou aceitação. A avaliação serve para, de entre todas as hipóteses propostas, escolher a que parece estar mais de acordo com a realidade.

O processo intelectual que permite a construção das hipóteses explicativas a partir das parcelas observadas é um exercício de indução pois parte de casos particulares e tenta explicações gerais. Pelo contrário, o processo de avaliação é um exercício de dedução em que se confronta, em novas situações, o previsto pelas hipóteses explicativas com a realidade observada.

Normalmente, as explicações contêm várias hipóteses interligadas a que, em termos genéricos, chamamos de teorias.

Uma teoria será tanto mais generalizável quanto mais distante for da explicação imediata e directa da correlação entre as parcelas observadas. Uma teoria generalizável terá que ser menos descritiva do observado e conter uma explicação para novas situações aparentemente diferentes da realidade observada: por exemplo, a descrição do movimento de Galileu explica a queda dos graves tal qual foi observado mas a lei da gravidade de *Newton* não só explica a queda da maçã como permitiu prever o movimento dos corpos celestes (excepto a trajectória de Mercúrio que apenas foi explicada pela teoria geral da gravidade de Einstein).

A modelização matemática foi o instrumento conceptual que permitiu o aparecimento e desenvolvimento da ciência. No entanto, pelo menos desde os meados do século XIX, a dedução matemática têm-se mostrado menos produtiva no aprofundamento da ciência. Esta dificuldade surge, em primeiro lugar, do facto de o progresso da ciência obrigar à construção de teoria cada vez mais generalizáveis que têm uma dificuldade de dedução e avaliação crescente e de, em segundo lugar, se constatar que na generalidade das relações de causa efeito existe um certo grau de aleatoriedade.

Sendo que, inicialmente, a aleatoriedade foi entendida como uma peculiaridade dos jogos de sorte e azar e, consequentemente, sendo dada pouca importância à teoria da probabilidade, a necessidade de teorizar conjuntos vastos de objectos inter-relacionados trouxe importância aos modelos estatísticos. Sendo que estes modelos, por poderem ser manipulados algebricamente, o facto de poder ser utilizada a experimentação estatística computacional abre novos caminhos para a ciência.

A possibilidade de aplicação de cálculo computacional é um grande passo para a ciência porque se trata de uma ferramenta criada pelo Homem que complementa a sua inteligência. E esta ferramenta ajuda a desenhar ferramentas com melhor desempenho dai se observar que o seu preço decresce rapidamente.

Sendo que a experimentação estatística tem na sua génese os jogos de sorte e azar, denomina-se genericamente esta metodologia por "Método de Monte Carlo".

Nesta dissertação, primeiro apresentamos os trabalhos pioneiros do "Método de Monte Carlo" que vão desde o Século XVII até meados do século XX. Posteriormente ilustramos a importância da aplicação da experimentação estatística na Econometria, nomeadamente no estudo da transitividade da causalidade de Granger e na Cointegração que, em termos intuitivos, será de aceitar a sua existência.

Do estudo conduzido, concluímos que, ao contrário do que parece intuitivo, será de rejeitar a conjectura de que existe transitividade entre relações de causalidade ou de cointegração mesmo que estas sejam estatisticamente significativas.

2. Os pioneiros

Neste ponto vamos apresentar os trabalhos pioneiros da experimentação estatística que começam em meados do sec. XVIII com Buffon (1777) e terminam em meados de Século XX com Metropolis e Ulam (1949).

Sendo que o tempo disponível para realizar uma tese de mestrado deve ser curto, pensamos que seria demasiado trabalhoso realizar um levantamento bibliográfico exaustivo sobre a evolução do "Método de Monte Carlo". Assim, como a construção deve começar pelos alicerces, concentramo-nos no estudo dos pioneiros com o desígnio de num trabalho posterior estender esta revisão da literatura.

2.1. Lord Buffon (Georges Louis Leclerc)

É considerado na literatura (ver, por exemplo, Ross, 1976)[1], que Buffon (1733) utiliza pela primeira vez os fundamentos do Método de Monte Carlo. Georges Louis Leclerc, Conde de Buffon, projectou uma experiência estatística para determinar o valor de π. Nessa experiência, é atirada uma agulha de forma aleatória sobre um conjunto de linhas paralelas igualmente distanciadas. Esta experiência que se encontra formalizada em Buffon (1777) ficou conhecida como "problema da agulha" sendo a primeira utilização da média de uma experiência aleatória repetida muitas vezes na obtenção da solução de um problema determinístico.

Na sua experiência, Buffon (1777) divide uma folha de papel em fatias de largura W utilizando linhas paralelas com espessura desprezível. Depois, atira aleatoriamente uma agulha de comprimento L, com $L < W$, e calcula a fracção de vezes que a agulha intersecta umas das linhas (não existe possibilidade de a agulha intersectar duas linhas em simultâneo).

Sendo que a agulha é atirada aleatoriamente para próximo do centro e a área do papel onde se realiza a experiência é tal que torna quase impossível a agulha cair fora dele, então pode ser assumido que a distância do centro da agulha à linha mais próxima segue distribuição uniforme no intervalo $[0, 0.5W]$. Por outro lado, a agulha faz um angulo θ com a direcção das linhas paralelas, em que θ também segue uma distribuição uniforme no intervalo $[0, \pi]$ (ver fig.1).

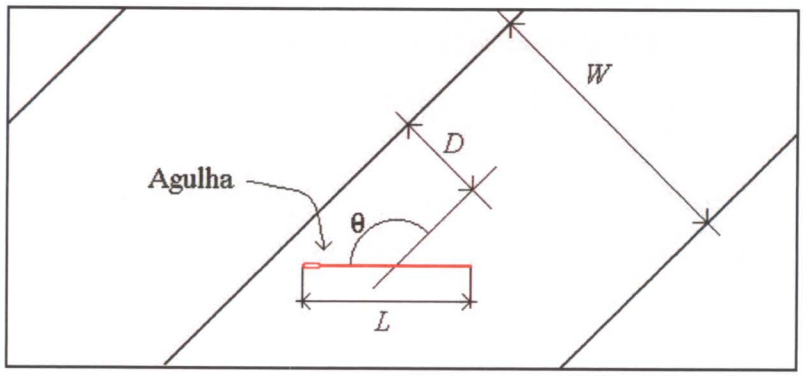

Fig. 1 - Visualização de W, L, D e θ no "problema da agulha"

A agulha intersecta uma das linhas se a distância D for menor ou igual à distância $H = 0.5\, L\, Seno(\theta)$ (ver fig. 1 e fig. 2).

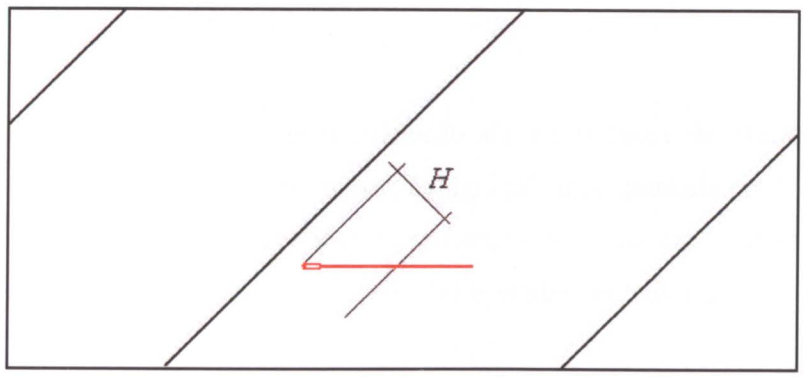

Fig. 2 - Visualização de H no "problema da agulha"

[1] Também ver http://www.user.fh-stralsund.de/~khammer/Essay.htm e http://www.efg2.com/Lab/Mathematics/Buffon.htm

Considerando primeiro como um dado o angulo θ que a agulha faz com a direcção das linhas paralelas, a agulha intersecta a linha mais próxima se $D \leq 0.5\, L\, Seno(\theta)$. Então, repetindo um número infindável de vezes o lançamento da agulha, a proporção de vezes em que a agulha intersecta uma linha vem dada por $f(\theta) = L\, Seno(\theta) / W$.

Para calcularmos, em termos esperados, a proporção de vezes que a agulha intersecta uma das linhas considerando um angulo qualquer θ que é uma extracção de uma distribuição uniforme com densidade de probabilidade $1/\pi$ no intervalo [0; π], aplicamos o valor esperado à expressão $f(\theta)$ anterior:

$$A = \int_0^x f(x)\frac{1}{\pi}dx \quad = \int_0^{\pi} \frac{L}{W} seno(\theta)\frac{1}{\pi}d\theta \quad = \frac{L}{W}\frac{2}{\pi} \tag{2.1}$$

Interpretando a proporção das vezes que a agulha intersecta uma linha como uma probabilidade, então numa experiência aleatória em que a agulha é lançada N vezes que das quais M vezes a agulha intersecta uma das linhas, obtém-se um estimador cêntrico para π invertendo a expressão anterior (ver, e.g., Kalos e Whitlock, 1986):

$$\hat{\pi} = \frac{2\,L}{W}\frac{N}{M} \tag{2.2}$$

Com o objectivo de ilustrar a evolução da experiência de Buffon (1777), apresentamos uma experiência do "problema da agulha" implementada em Matlab (ver anexo 1) em que a experiência aleatória é substituída por uma experiência pseudo - aleatória e em que comparamos as estimativas com o verdadeiro valor de π.

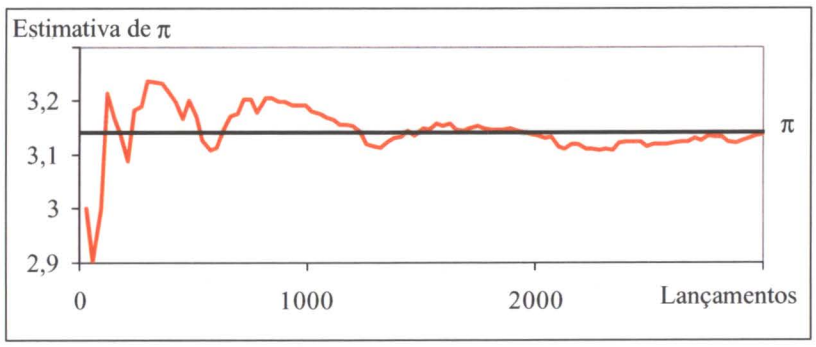

Fig. 3 - Uma experiência pseudo-aleatória[2] do "problema da agulha"

Na figura anterior verifica-se que, em tendência, o erro da estimativa diminui com o número de lançamentos realizados. No entanto, se o verdadeiro valor de π não fosse conhecido, seria necessário calcular uma estimativa para o "erro" da estimativa. Considerando que cada estimativa é uma extracção independente da função distribuição das estimativas, podemos estimar o erro da estimativa como o desvio padrão de um conjunto de experiências. Na figura seguinte, apresentamos o desvio padrão considerando 30 experiências[3]. Ajustada pelo método dos mínimos quadros uma curva iso-elástica, estima-se que, em média, o erro de cálculo de π decresce 24,5% quando se duplica o número lançamentos ($1{,}0919\ NLançamentos^{-0{,}4053}$):

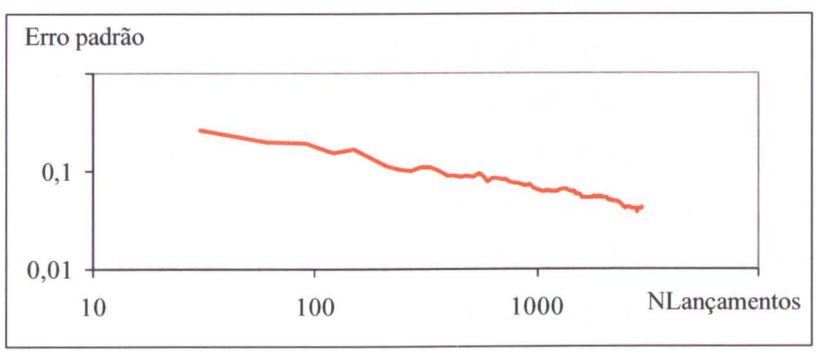

Fig. 4 - Evolução do erro com o tamanho da experiência

[2] Uma variável pseudo-aleatória comporta-se em termos estatísticos como uma aleatória. Posteriormente retomaremos este assunto.

[3] O leitor pode comparar com http://www.geocities.com/CollegePark/Quad/2435/buffon.html

Em termos estatísticos, se os lançamentos forem aleatórios e independentes, a duplicação do número de lançamentos faz diminuir o erro, em termos esperados, em $1-1/\sqrt{2} = 29{,}3\%$ (da simulação resulta 24,5%) e a distribuição do estimador de π converge para a distribuição Normal quando o aumento do número de experiências aumenta. Assim, seria de "observar" em 2/3 dos casos que a grandeza do erro verdadeiro da corrida fosse menor que o erro padrão simulado (o que não parece, em termos qualitativos, se verifique na nossa experiência):

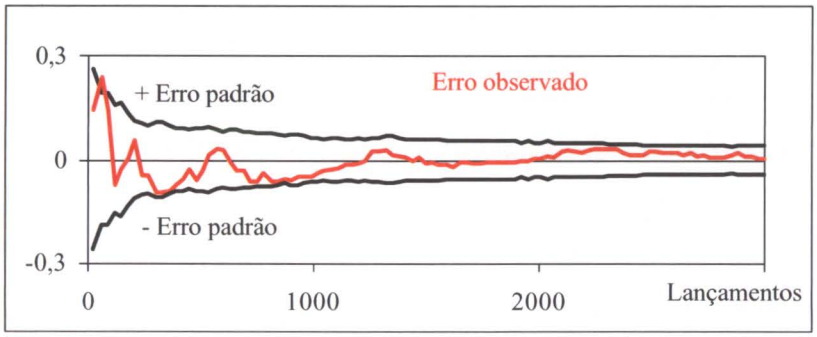

Fig. 5 - Comparação do erro padrão com o erro observado

2.2. Lord Rayleigh (John William Strutt)

Apesar de entre Buffon (1777) e Rayleigh (1899) outros autores terem usado experiências semelhantes ao "problema da agulha", é Rayleigh (1899) quem primeiro afirma que se pode usar um fenómeno físico simples mas estocástico para encontrar uma solução aproximada para outro fenómeno físico determinístico mas de difícil resolução algébrica. Assim, John William Strutt, conde de Rayleigh, demonstra que uma simples "caminhada" aleatória numa linha sem limites "absorventes" é uma solução aproximação para uma equação diferencial parabólica.

Considerando o fenómeno aleatório "número m de falhas e n de sucessos num conjunto de μ experiências em que a probabilidade de falha de cada experiência é p" com m e n suficientemente grandes para poderem ser consideradas contínuas. Este fenómeno aleatório traduz um movimento *browneano* numa linha com saltos unitários. Assim, a "partícula" salta na linha m vezes para a direita e n vezes para a esquerda. Denominemos por x a distância da partícula ao ponto inicial decorrido um determinado

número de experiências, podemos representar o movimento da partícula num plano em que a abcissa é o tempo e a ordenada é a posição:

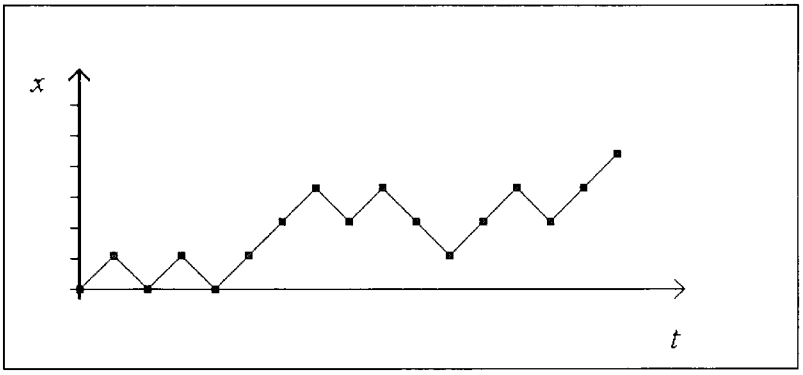

Fig. 6 - Movimento *browneano* numa linha, representado no tempo

Considerando que a probabilidade p de falha é 0.5, é igualmente provável acontecer um salto para a direita como para a esquerda pelo que a partícula não tem tendência de se afastar do ponto de origem. Considerando saltos pequenos e um número grande de experiências, podemos considerar o fenómeno como se fosse contínuo. Desta forma podemos representar a função densidade de probabilidade de a partícula se encontrar na posição x ao fim de μ experiências (que é uma medida equivalente ao tempo):

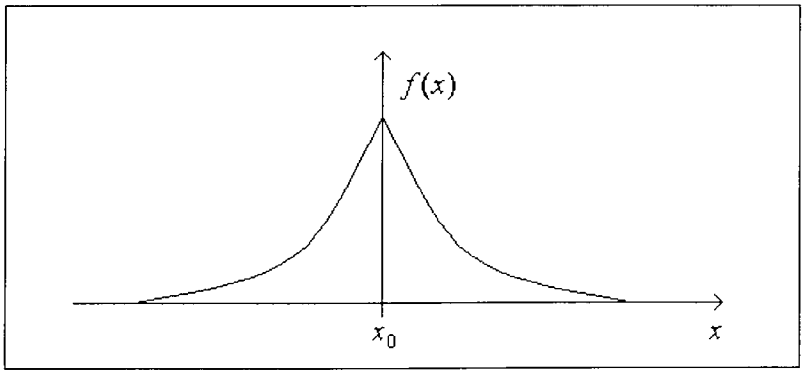

Fig. 7 - Função densidade de probabilidade de a partícula parar em x

Podemos agora determinar a densidade de probabilidade $f(x/\mu)$ da partícula ter percorrido a distância x quando são realizadas μ experiências e a probabilidade p de falha é 0.5:

$$f(x \mid \mu) = \frac{1}{\sqrt{2\pi\mu}} \exp\left(-\frac{x^2}{2\mu}\right) \tag{2.3}$$

Esta equação é a solução para a equação diferencial parabólica que é conhecida na física como "função de difusão" em que μ representa o tempo e x a distância à matéria/energia a difundir:

$$\frac{df}{d\mu} = \alpha \frac{\partial^2 f}{\partial x^2} \tag{2.4}$$

Desta forma, Rayleigh (1899) prova que se pode encontrar uma solução aproximada de problema de "difusão", que é determinístico, observando um movimento *browneano*, que é estocástico.

Assim, ao repetir-se um número grande de vezes o fenómeno "soma que resultado de μ experiências em que cada uma tem a probabilidade 0.5 de ser $+\alpha$ e 0.5 de ser $-\alpha$", a frequência relativa "amostral" é uma solução aproximada para a equação diferencial parabólica num instante de tempo fixo. Para que μ possa ser um número elevado e mesmo assim se poder determinar a equação para instantes de tempo próximos de zero, usa-se α como uma escala temporal, i.e., $t = \mu\,\alpha$.

2.3. Lord Kelvin (William Thomson)

Sendo conhecida a lei dos gases estáticos a partir da observação experimental em laboratório (lei de Boyle: $P \cdot V$ = Constante; lei de Charles: V/T = Constante; hipótese de Avogadro: V / n = Constante), deriva-se a lei dos gases estáticos. Esta lei diz que num gás estático (em que não se observa movimento), a pressão, P, vezes o volume do vaso que o contém, V, é proporcional ao número de átomos, n, vezes a temperatura, T, sendo a relação de proporcionalidade, R, a constante de Boltzmann:

$$P \cdot V = n \cdot R \cdot T \tag{2.5}$$

Com o objectivo de justificar esta lei, Maxwell (1860) propõe a teoria cinética dos gases em que é assumido que os gases são formados por moléculas que, apesar de não se observar movimento à escala macro, se movimentam livremente com uma determinada velocidade. A velocidade de uma moléculas do gás escolhida à sorte é uma variável aleatória vectorial em $\Re^3$. As moléculas que constituem o gás acabam por chocar com as paredes do vaso que o contém como se fossem balas. Assim, a pressão que se observa resultará da soma das quantidades de movimento transferidas para a parede do vaso que resultam das colisões (em número astronómico) das moléculas com a parede.

Em termos formais, a pressão vem dada pelo valor esperado da quantidade de movimento transferida no choque de uma molécula seleccionada de forma aleatória. Considerando que a norma da velocidade é v e segue a função distribuição $F(v|T)$ e que o angulo que o vector velocidade faz com a parede é θ e segue distribuição uniforme no intervalo $[-\pi, \pi]$, a pressão virá dada por:

$$P = A \cdot \int_{-\pi/2}^{\pi/2} \int_{0}^{+\infty} f(v \mid T) \cdot \cos(u(\theta)) \cdot dv \cdot d\theta \tag{2.6}$$

Sob o pressuposto de que os gases são formados por moléculas mono-atómicas (formadas por um só átomo), a sua densidade é suficientemente baixa para que nunca ocorram colisões entre as moléculas e as colisões com a parede são perfeitamente elásticas, então a função distribuição da norma da velocidade da moléculas individuais é a equação de Maxwell – Boltzmann que tem um único parâmetro que é a energia interna do gás, T, que é proporcional à temperatura absoluta do gás. Assim, a velocidade v de uma molécula escolhida à sorte tem $MB(v \mid T)$ como função distribuição, Maxwell (1860).

O gás de Maxwell denomina-se de "perfeito" porque os gases "reais" são formados por moléculas multi-átomos com elevada densidade (ocorrem colisões entre moléculas). Sendo assim, uma percentagem significativa da energia interna do gás

encontra-se sob a forma de velocidade de rotação. Esta energia de rotação tem como efeito diminuir a velocidade linear das moléculas e alterar o seu comportamento quando chocam com a parede (há transferencia não só de movimento mas também de momento angular). Kelvin (1901) investiga até que ponto estas diferenças serão importantes (será de prever que os gases reais tenham uma pressão menor que a prevista pela lei dos gases perfeitos).

Sendo que o problema é analiticamente intratável, William Thomson, Barão de Kelvin, propõe que a pressão causada pelo gás pode ser aproximada utilizando o que hoje é conhecido como o método de Monte-Carlo. Assim, é o primeiro autor que usa uma experiência aleatória no estudo de um problema estocástico.

Kelvin (1901) considera um pequeno número de moléculas em que, primeiro, atribuímos uma certa percentagem da energia interna à rotação e, segundo, extraímos vectores velocidade $\vec{V} = (v, \theta)$ com norma v segundo a distribuição de Maxwell – Boltzmann (com a energia interna diminuída da percentagem afecta à rotação) e angulo θ segundo a distribuição uniforme, então, a pressão virá dada por:

$$P = A \cdot \frac{1}{n} \sum_{i=1}^{n} v_i \cdot \cos(\theta_i) \tag{2.7}$$

Kelvin (1901), realiza várias experiências com densidades e temperaturas diferentes e gera aleatoriamente as velocidades de um conjunto relativamente pequeno de moléculas (entre 100 e 900) procurando ajustar os resultados à lei dos gases perfeitos. Na realização da experiência aleatória, Kelvin (1901) utilizou baralhos de cartas numeradas e o lançamento ao ar de moedas. Apesar de utilizar uma experiência com poucas "moléculas", das simulações realizadas Kelvin(1901) pode concluir que, nos gases reais, a forma funcional da distribuição das velocidades é significativamente diferente da equação de Maxwell – Boltzmann.

2.4. Student (William Sealey Gosset)

Student (1908a, 1908b), utiliza pela primeira vez o Método de Monte Carlo em problemas estatísticos. Com o objectivo de estudar a função distribuição *a priori* de

uma estatística (o coeficiente de correlação linear e a média) para amostras pequenas que apenas eram conhecidas para amostras grandes.

Teoricamente, sendo dado o coeficiente de correlação da população, R, então o coeficiente de correlação amostral é um valor contido no intervalo [– 1,1] cuja função de distribuição *a priori* já era conhecida para amostras grandes, Pearson (1896). O problema de Student (1908a) seria fazer um estudo exploratório, por experimentação, da forma da função de distribuição *a priori* para o caso de amostras de pequena dimensão.

Student (1908a) construiu uma população artificial com 3000 indivíduos caracterizados por duas medidas não correlacionadas (em que o coeficiente de correlação linear era 0). Com o objectivo de estudar como a função distribuição do estimador do coeficiente de correlação linear evolui com o tamanho da amostra, Student (1908a) realizou uma experiência em que considera amostras com 4 indivíduos, 8 indivíduos e 30 indivíduos. Na experiência Student (1908a) partiu aleatoriamente a população artificial em 750 amostras com 4 indivíduos em cada uma delas, calculou os 750 coeficientes de correlação amostral e desenhou a função distribuição experimental. Depois, Student (1908a) combinou as amostra de dimensão 4 duas a duas, obtendo 750 amostras com 8 indivíduos (que considerou suficientemente independentes) e desenhou a função distribuição experimental. Finalmente extraiu 100 amostras com 30 indivíduos cada e desenhou a função distribuição experimental. Do estudo da evolução da função distribuição experimental com o tamanho da amostra, Student (1908a) conjecturou que a função distribuição do estimador do coeficiente de correlação linear, r, deveria ter a seguinte forma funcional (em que n traduz o tamanho da amostra):

$$f(r) = p_0 \cdot \left(1 - r^2\right)^{\frac{n-4}{2}} \tag{2.8}$$

Esta forma funcional tem como propriedade convergir para o resultado analítico de Pearson (1896).

Student (1908a) não conseguiu intuir nenhuma forma funcional para a função densidade de probabilidade do estimador do coeficiente de correlação linear quando o seu valor é diferente de zero.

Em Student (1904b) a questão vira-se para o estudo do comportamento da média amostral quando o valor médio e a variância da população são desconhecidos. Sendo a

amostra grande, resulta pelo teorema limite central que a função distribuição *a priori* da média amostral é a lei Normal em que se assume como valor médio a média amostral e como variância a variância amostral. No entanto, para amostras pequenas, a função distribuição *a priori* é dependente do tamanho da amostra e da função distribuição da população. Como o resultado depende da função distribuição da população, Student (1904b) assume o caso particular em que a população segue uma distribuição Normal com valor médio e variância desconhecidos.

Antes de tentar resolver o problema analiticamente, Student (1904b) desenhou uma experiência semelhante à utilizada no estudo do coeficiente de correlação linear: criou uma população artificial com distribuição Normal composta por 3000 indivíduos com valor médio e variância conhecidas. Depois, considerou várias experiências em que variou o tamanho das amostras (tamanho igual a 4, 5, 6, 7, 8, 9 e 10) e representou os histogramas de frequências da média amostral. Reparou que o histograma se assemelhava na forma à distribuição Normal mas que o desconhecimento da média fazia com que a dispersão fosse ligeiramente maior e que a diferença era tanto mais pequena quanto maior fosse a amostra.

Baseado nesta experiência, Student (1904b) acrescentou à forma funcional da distribuição Normal um parâmetro dependente do tamanho da amostra (que se denomina de graus de liberdade). Assumindo ainda que o desconhecimento do valor médio induz um aumento da dispersão da distribuição da média amostral, Student (1904b) derivou a curva analítica função distribuição *a priori* das médias amostrais quando são desconhecidos o valor médio e a variância da população e que hoje é conhecida por distribuição *t* de Student.

2.5. Richard Courant, Kurt O. Friedrichs e Hans Lewy

Courant *et al.* (1928) estudam problemas de fronteira nas equações diferenciais parciais (EDP). Apesar de serem problemas de resolução analítica muito difícil, são importantes porque permitem modelizar o meio contínuo. Um exemplo muito conhecido de meio contínuo é o modelo de difusão em que a velocidade de variação da variável em estudo num dado ponto é função do gradiente espacial dessa mesma variável.

Na figura 8 representa-se uma peça metálica de geometria recortada e com furos, soldada entre uma zona quente (por exemplo, a 500ºC) e uma zona fria (por exemplo, a 100ºC) e pretende-se determinar a temperatura num qualquer ponto genérico dentro do meio contínuo que a chapa metálica representa. A temperatura vai evoluindo no tempo até que estabiliza quando todas as derivadas espaciais de segunda ordem forem zero.

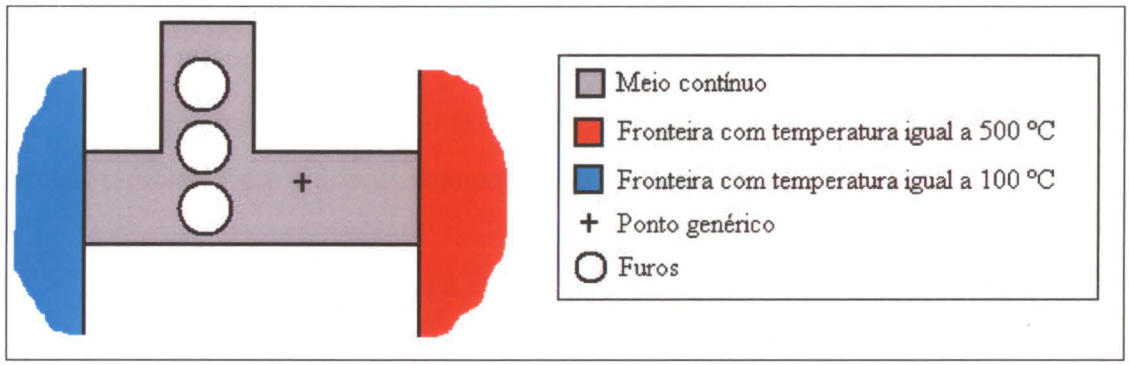

Fig. 8 – Esquema de difusão de calor num meio contínuo irregular

Sendo que os problemas do meio contínuo podem ser resolvidos para casos em que a geometria é simples, os casos com geometria complicada, de que a figura 8 é um pequeno exemplo, são irresolúveis. Uma forma de encontrar uma solução aproximada considera-se o meio contínuo reduzido a apenas um conjunto limitado de pontos que estão dispostos segundo uma malha rectilínea. Desta forma, as equações diferenciais transformam-se em equações às diferenças.

Courant *et al.* (1928), imaginando que o problema de difusão é um caminho aleatório ao longo do meio contínuo, propõem uma imitação numérica desse processo aleatório. Este trabalho é muito importante porque prova que o comportamento aleatório dos à escala microscópica (dos átomos) se traduz num comportamento determinístico à escala dos corpos sólidos (macroscópica). E é esta ideia que vai dar directamente ao algoritmo do Monte Carlo. Assim, Courant *et al.* (1928), partem de um conjunto de "átomos" (por exemplo, com 1000 átomos) e, um a um, sorteiam o local do seu aparecimento e um caminho que parte desse ponto e desaparece num ponto da fronteira (movimento browniano). Depois passam a outro dos "átomos" até não haver mais.

Ao longo do caminho de cada "átomo", estendem-se as equações às diferenças para os pontos vizinhos desse. Desta forma, com o aumento do número de "átomos", a solução vai-se aproximando da solução do problema pretendido.

Em termos formais, consideremos $X = G$ um conjunto de pontos que formam um meio contínuo fechado em $\Re^n$. Deste conjunto são escolhidos os pontos G_h dispostos segundo uma malha regular de largura h em que G_f são os pontos da malha que formam a fronteira: $G_f \subset G_h \subset G$.

Quando estamos no ponto interior $X = (x_1,..., x_n)$, para $i \in \{1,..., n)$, as derivadas parciais espaciais V em ordem às ordenadas vêm aproximadas por

$$V_i(X) \approx \frac{[u(..., x_i + h,...) - u(X)]}{h} \text{ e } V_i(X) \approx \frac{[u(X) - u(..., x_i - h,...)]}{h} \qquad (2.9)$$

Mas nós conhecemos as derivadas parciais e queremos determinar o valor da função real $u(X)$ de variável em $\Re^n$. Então, estando o "átomo" em X, pela inversão da expressão 2.8 podemos calcular os valores de $u(\ )$ nos 2 n pontos que pertencem à malha e são vizinhos de X.

$$u(..., x_i + h,...) \approx u(X) + h\, V_i(X) \text{ e } u(..., x_i - h,...) \approx u(X) - h\, V_i(X) \qquad (2.10)$$

Como esses pontos têm uma proposta de solução da vez anterior que um átomo passou na sua vizinhança, $u^a(\ )$ vamos repartir as diferenças entre o valor anteriormente calculado e o de agora entre o ponto X e os seus vizinhos. Assim, o valor $u(X)^d$ que aproxima $u(X)$ depois de o "átomo" ter passado por X ficará:

$$u(X)^d = \alpha\, u(X)^a + (1 - \alpha)\, \frac{\sum (u_i{}^a - u^d{}_i)}{2\, n} \qquad (2.11)$$

O valor de $u(\ )$no ponto genérico Y vizinho de X virá actualizado para:

$$u(Y)^d \approx \beta\, u(X)^d + (1 - \beta)\, u(Y)^d \qquad (2.12)$$

Para que nos aproximemos da solução correcta, os parâmetros $\propto$ e β têm que pertencer ao conjunto fechado [0, 1] e o produto de $\propto$ por β tem que pertencer ao conjunto aberto] 0, 1 [.

Se um dos pontos do caminho ou da vizinhança for um ponto da fronteira onde é dado o valor de u, então este valor não é actualizado. Se o ponto do caminho for um ponto da fronteira onde é dado o valor de u, o "átomo" desaparece.

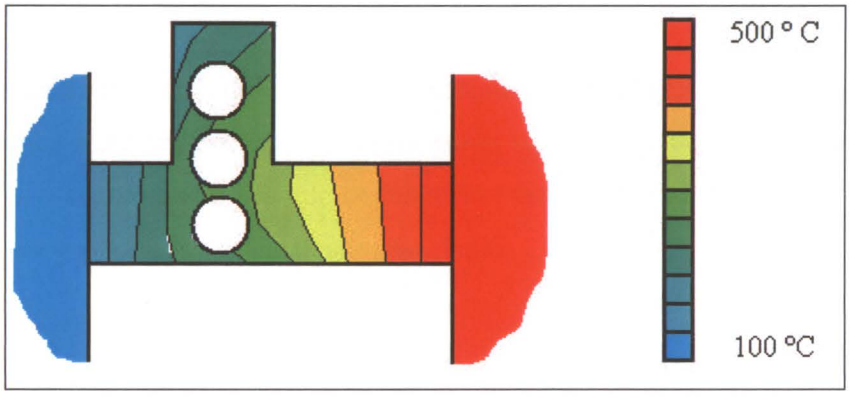

Fig. 9 – Esquema ilustrativo da temperatura num meio contínuo irregular

Se as derivadas parciais forem de ordem m, então transformam-se num sistema com m derivadas parciais de primeira ordem e a função $u(\ )$ será uma função vectorial em $\Re^m$ de variável em $\Re^n$.

2.6. Andrey Nikolaevich Kolmogorov

Kolmogorov (1931) formaliza o que hoje é conhecido como o processo estocástico markoviano de primeira ordem, em tempo contínuo. É um trabalho importante porque confirma que existe uma equivalência entre um sistema cujo o conhecimento do seu estado é dado por uma função distribuição de probabilidades e uma população de indivíduos em que cada um deles segue uma trajectória aleatória (*random walk*) e em que a probabilidade se transforma numa frequência relativa.

Vamos imaginar que o sistema Y pode estar num de vários estados sendo que os estados possíveis formam o conjunto U. Sendo assim, no instante t, o sistema Y está no estado y_t. Sendo U um conjunto discreto teremos:

$$y_t = u_i \in U, i \in \{1, ..., n\} \text{ e } t \in [t_0, t_1] \tag{2.13}$$

A questão que se põe agora é saber em que estado estará o sistema no instante $t + h$. O processo será de primeira ordem se o estado do sistema em $t + h$ depender do estado do sistema em t:

$$y_{t+h} = f_t(y_t, h) \tag{2.14}$$

No mundo que nos rodeia há muitos exemplo de "processos de primeira ordem" que, em termos matemáticos, se modelizam como sistemas de equações diferenciais. Por exemplo, o angulo que o pêndulo de tamanho l faz com a vertical resolve a equação seguinte (e que $g \approx 9{,}81$ é a aceleração da gravidade):

$$\frac{d^2 y_t}{dt^2} = -\frac{g}{l} sen(y_t) \tag{2.15}$$

Se a transição de estado for estocástica, o processo diz-se markoviano. Por exemplo, atirando um dado equilibrado ao ar, a probabilidade de cair a face "1" virada para cima é de 1/6 (e independente do estado anterior – é de ordem zero). Kolmogorov (1931) denomina estes processos de "stochastically definite process", i.e. processos definidos de forma estocástica.

Um processo que reuna estas duas características (que dependa do estado anterior e que seja estocástico) será um processo markoviano de primeira ordem.

Por exemplo, vamos supor que o estado do tempo é uma de quatro possibilidades: U = {"sol", "nublado com abertas", "nublado", "muito nublado"). Sendo que hoje está "sol", então o estado do tempo amanhã será "sol" com 65% de probabilidade, "nublado com abertas" com 20% de probabilidade, "nublado" com 10% de probabilidade e "muito nublado" com 5% de probabilidade (somando 100%). Considerando um sistema linear e aditivo, poderíamos, em termos de álgebra matricial, dizer que o estado do tempo amanha será dado pelo estado do tempo hoje mais uma matriz de transição (que contém as probabilidades de transição entre estados):

$$y_{t+1} = \begin{bmatrix} .65 & .15 & .05 & .00 \\ .20 & .60 & .15 & .10 \\ .10 & .15 & .60 & .40 \\ .05 & 10 & .20 & .50 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.16}$$

Com este sistema assim formalizado, poderemos prever o estado do tempo (em termos de probabilidades) num futuro distante. Por exemplo, o estar hoje "sol" ou "muito nublado", em probabilidade, não tem influência no estado do tempo daqui a 10 dias (multiplica-se o estado do tempo 10 vezes pela matriz de transição de 16):

Daqui a 10 dias \ Hoje	"sol"	"muito nublado"
"sol"	17,5%	16,0%
"nublado com abertas"	27,2%	26,6%
"nublado"	34,7%	35,9%
"muito nublado"	20,7%	21,5%

Tabela 1 – Previsão do estado futuro em função do estado presente

Kolmogorov (1931) intuiu que, apesar de haver apenas um sistema, as probabilidades podem ser vistas como frequências relativas de vários "sistemas" que evoluem em simultâneo. Assim, pode-se seguir as histórias individuais de cada "sistema" sorteando em que estado o sistema cai, e reconstituindo a todo o tempo as probabilidades pelo cálculo das frequências relativas dos "sistemas". O movimento de cada "sistema" será uma caminhada aleatória (*random walk*) entre os estados possíveis dados por *U*. Na figura 10 representa-se uma hipotética história do estado do tempo em que os nomes dos estados foram substituídos, sem perda, por números.

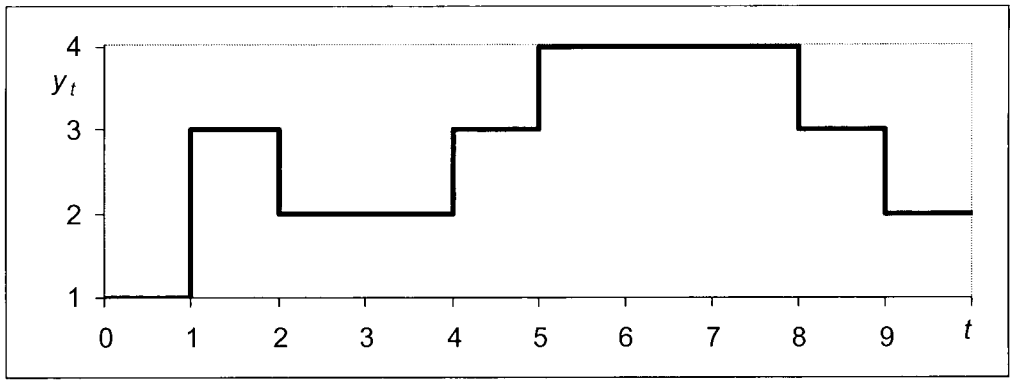

Fig. 10 – Caminhada aleatória de um dos "sistemas"

O processo estocástico markoviano que modeliza cada "sistema" individual pode ser seguido em tempo contínuo se a mudança de estado for modelizada como um "processo de chegada" tipo processo de Poisson. Neste processo, o sistema mantém-se no mesmo estado até que é "perturbado" por uma chegada, altura em que "salta" para outro estado. O ser uma trajectória aos saltos traduz que estamos em presença de um movimento *browneano*.

A transformação das probabilidades em frequências relativas e a modelização de cada "sistema" como um "processo de chegada" veio permitir a aplicação do Método de Monte Carlo na resolução de qualquer sistema dinâmico.

Apesar de Kolmogorov (1931) ter considerado que havia apenas um sistema, veremos que a metodologia pode ser estendida a uma população de indivíduos (sistemas) a evoluir em simultâneo, podendo haver transformação de uns noutros (processo de morte e nascimento).

2.7. Enrico Fermi

Em 1933, Fermi, numa tentativa de obter átomos radioactivos por transmutação, bombardeou substâncias com neutrões. Como não obteve resultados imediatos (as substancias não se tornavam radioactivas), conduziu um estudo estatístico para avaliar qual deveria ser a probabilidade de um neutrão chocar com um átomo da placa bombardeada. Sendo que determinou que essa probabilidade era muito pequena, aumentou a potência da fonte emissora de neutrões e obteve resultados imediatos para o Alumínio, Fermi (1934). Apesar de não ter publicado nada sobre a experiência

estatística, a literatura considera-o como o pioneiro do uso do Método de Monte Carlo no estudo da difusão dos neutrões (Segre, 1962 e Metropolis, 1987). Nas suas experiências, Fermi usou números aleatórios verdadeiros (usou dados ou cartas baralhadas). Posteriormente, Fermi construiu um "computador analógico" que, aplicado sobre um desenho do "reactor" a estudar, implementava o Método de Monte Carlo, o Fermiac (ver ponto 2.8).

2.8. John Von Neumann, Robert Davis Richtmyer e Stalisnaw Ulam

Quando foi iniciado o projecto Manhattan cujo objectivo era a criação da bomba atómica já se sabia que ao bombardear átomos com neutrões se obtinham outros elementos que eram radioactivos. Esta descoberta levou à atribuição do prémio Nobel da Física a Fermi (1938).

O principio da bomba atómica seria uma reacção em cadeia. Primeiro, um núcleo de um átomo era atingido por um neutrão e transformava-se num núcleo diferente que era instável. Depois, este explodia com libertação de energia e emissão de vários neutrões. Se partes desses neutrões atingissem outros núcleos de átomos, então criava-se uma reacção em cadeia.

Sendo que os átomos estão colocados numa grelha tridimensional e os núcleos são muito pequenos relativamente ao espaço ocupado pelos átomos (1/10000 do raio atómico), a probabilidade de um neutrão emitido num determinado local atingir um núcleo de outro átomo é muito pequena (ver fig. 11).

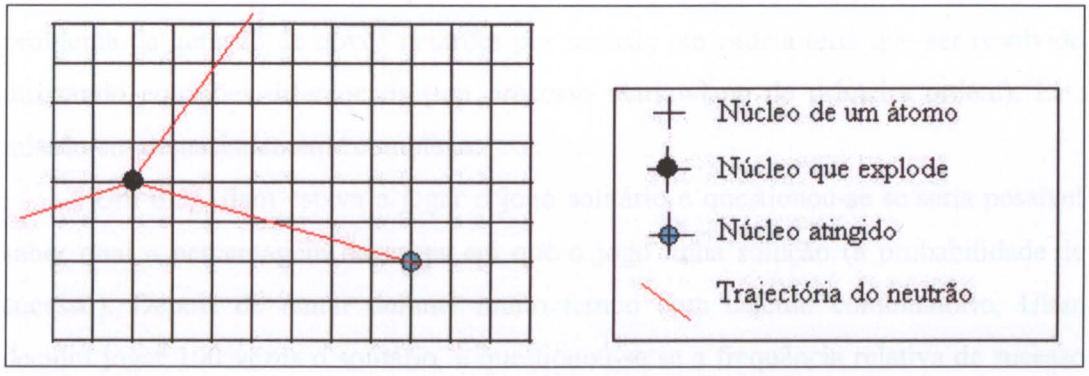

Fig. 11 – Ilustração do problema da difusão dos neutrões

Como, em termos de reacção do Urânio 235, na fissão de um núcleo são emitidos cerca de 2,5 neutrões, a bomba atómica precisaria de uma dimensão que garantisse que a probabilidade de um neutrão chocar com um núcleo fosse superior a 40%.

Comparando com Courant *et al.* (1928), estamos em presença de um problema de difusão em que a fronteira "absorvente" dos átomos está localizada dentro do meio contínuo seguido de um problema de transmutação atómica e multiplicação de neutrões. Fermi (1934) já tinha abordado a primeira parte desta questão (a difusão de neutrões).
O aparecimento do computador digital em 1945, o ENIAC, veio abrir novas possibilidades na utilização da amostragem estatísticas. Isto porque a necessidade de realizar muitos cálculos, tornava muito dispendiosa a utilização do Método de Monte Carlo (usando uma máquina de calcular mecânica).

Criado o computador digital que permitia a realização rápida de cálculos, ainda subsistia o problema da geração dos números aleatórios. Ulam, Richtmyer e Von Neumann (1947) propõem a substituição dos números aleatórios por séries caóticas que denominam por números pseudo-aleatórios.

O projecto Manhattan era secreto pelo que os trabalhos aí desenvolvidos apenas foram publicados depois do material ter sido desclassificado. Sendo que os conceitos necessários para operacionalizar o Método de Monte Carlo de forma a que fossem utilizados computadores *multi-purpose* foram desenvolvidos durante o projecto, os colegas e os historiadores atribuíram a Stanislaw Marcin Ulam, Robert Davis Richtmyer e John Von Neumann o mérito pelo seu desenvolvimento, Eckhardt (1987).

Em termos clássicos, o problema da difusão dos neutrões juntamente com o problema da geração de novos neutrões por reacção em cadeia teria que ser resolvido utilizando equações diferenciais (um processo markoviano de primeira ordem). Esta missão era demasiadamente complexa.

Um dia, Ulam estava a jogar o jogo solitário e questionou-se se seria possível saber qual a percentagem de vezes em que o jogo tinha solução (a probabilidade de sucesso). Depois de tentar durante muito tempo com cálculo combinatório, Ulam decidiu jogar 100 vezes o solitário, e questionou-se se a frequência relativa de sucesso observada não seria uma aproximação válida para o resultado pretendido. Recordando

que Courant *et al.* (1928) e Fermi (1934) tinham usado alguns átomos individuais para simular problemas idênticos ao da difusão e geração de neutrões, Ulam pensou que, se fosse possível utilizar computadores rápidos, este problema poderia ser resolvido verificando o percurso de um conjunto limitado de neutrões e de átomos.

Ulam descreveu a ideia a John Von Neumann que começou a procurar substitutos determinísticos para os números aleatórios, Neumann e Ulam (1946)[4]. As séries caóticas pareceram um substituto perfeito por serem fáceis de criar e terem boas características estatísticas.

O algoritmo proposto denomina-se de "dígitos do meio do quadrado": i) Pensando que queremos um número no intervalo [0; 1], por exemplo, com 6 casas decimais. ii) Parte-se de um número inteiro com 6 dígitos, iii) acha-se o seu quadrado (que é um número com 12 dígitos) e iv) aproveitam-se os seis dígitos do meio. v) Este novo número é usado para calcular o seguinte, etc. vi) Obtêm-se os números pseudo-aleatórios dividindo a série obtida por um milhão (no caso de querermos 6 casas decimais).

Sendo que o conjunto dos números pseudo-aleatórios obtidos com os "dígitos do meio do quadrado" têm comportamento parecido com uma função distribuição uniforme, consegue-se transformar estes números noutros que seguem uma qualquer outra função distribuição aplicando-lhes a função inversa da função distribuição pretendida.

Ulam descreveu a ideia a Robert Richtmyer que começou a estudar até que ponto a aproximação estatística é adequada para resolver problemas de difusão e multiplicação de neutrões. Concluiu que, mesmo utilizando números pseudo-aleatórios, o teorema limite central era aplicável.

Em termos de resolução do problema da difusão, transmutação e criação de neutrões, foram considerados vários passos.

Primeiro, estimaram experimentalmente quantos neutrões são emitidos quando um átomo explode (por exemplo, 3 neutrões).

Segundo, assumiram que o átomo que explode está num determinado ponto de uma massa tridimensional de material de densidade e geometria conhecidos. Registavam que tinha explodido mais um átomo.

---

[4] Documento não publicado mas referido em Ulam (1982)

Terceiro, sorteavam a direcção em que um dos neutrões era emitido e a sua velocidade.

Quarto, seguiam a trajectória do neutrão e verificaram se colidia com um núcleo ou se saía da massa sem colidir. Se saísse, registavam que tinha saído mais um neutrão e simulavam a emissão de novo neutrão (voltavam ao ponto terceiro).

Quinto, no caso do neutrão colidir com outro átomo, sorteavam, em função da velocidade, se o neutrão se juntava ao núcleo (fundia) ou se escapava com velocidade menor (sorteavam a perda de velocidade). Se escapasse, continuavam a seguir a trajectória (voltavam ao ponto quarto). Se fundisse, voltavam ao ponto segundo, agora com as coordenadas do novo átomo (podiam-se formar átomos não activos, o que, sem perda de generalidade, se abstrai deste esquema).

Se, decorrido um determinado tempo de simulação, tivessem fugido N neutrões e tivessem explodido M átomos, a reacção em cadeia manter-se-ia se o número de neutrões "perdidos" por átomo fosse menor que os gerados por cada átomo (o que tinha sido determinado experimentalmente):

$$N/M+1<3 \tag{2.17}$$

No exemplo da figura 12, a reacção não "explode" porque 8/4+1 não verifica a inequação 2.17.

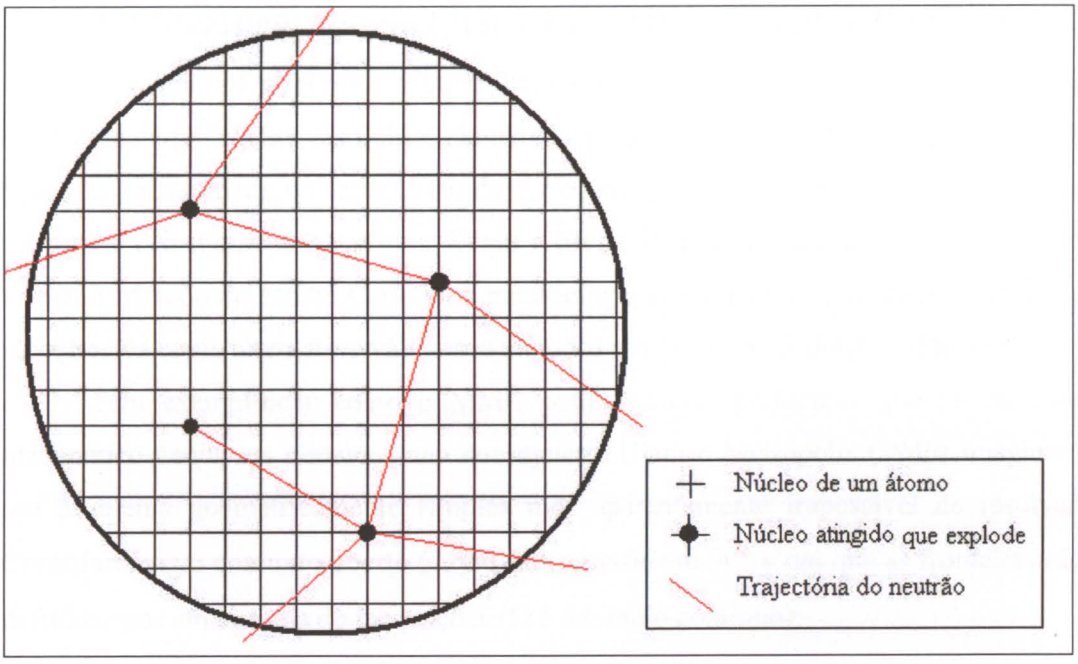

Fig. 12 – Ilustração do problema da difusão dos neutrões

Como a bomba não podia explodir durante a construção, experimentaram mecanismos de detonação, como seja dividir a esfera em várias partes que, no momento da explosão, se juntavam e aumentar a densidade (simulando o uso de explosivos). Experimentaram ainda substâncias absorventes de neutrões e substâncias que alteram a velocidade dos neutrões (moderadores ou reflectores).

A geometria variável e a alteração da densidade ainda é o mecanismo de detonação das bombas atómicas actuais.

Posteriormente à Segunda Guerra Mundial, em 1948, durante um período em que o ENIAC esteve inactivo, Fermi replicou o trabalho de Ulam, Richtmyer e Von Neumann (1947) com uma máquina analógica, o Fermiac, que circulava sobre um desenho bidimensional da massa de material a estudar (Metropolis, 1987). Esta máquina veio a perder toda a importância como tecnologia porque aconteceu um desenvolvimento exponencial dos computadores digitais.

2.9. Stanislaw Marcin Ulam e Nicholas Constantine Metropolis

Pegando nos princípios contidos em Courant *et al.* (1928) de que os problemas à macro-escala podem ser resolvidos simulando interacções aleatórias à micro-escala, e a proposta de Neumann e Ulam (1946) de que os números pseudo-aleatórios são um bom substituto dos números aleatórios, Ulam e Metropolis (1949) fazem uma síntese que tornou o Método de Monte Carlo uma metodologia aceite pela comunidade cientifica e reconhecida como alternativa válida aos métodos matemático - dedutivos clássicos.

Para exemplificar como o MMC pode resolver problemas que os métodos matemático-dedutivos clássicos não conseguem, Ulam e Metropolis (1949) imaginam um problema geometricamente simples mas aparentemente impossível de resolver. Consideremos o conjunto aberto Ω que está contido em $\Re^{20}$ e em que as fronteiras são definidas por um sistema de inequações (Ω é um meio contínuo):

$$\Omega = \{X \in R^{20} : f_1(X) < 0; f_2(X) < 0; ...\} \tag{2.18}$$

A questão mais simples que se pode colocar é como avaliar o volume de Ω.

Para resolver esta questão, primeiro temos que identificar os valores i e s que permitem definir o hiper-cubo Ψ que contém Ω:

$$\Psi = \{X \in R^{20} : i < x_1 < s; ...; i < x_{20} < s\}, \Omega \subset \Psi \tag{2.19}$$

Em termos determinísticos (clássicos), consideramos o meio discretizado numa malha de lado h como aproximação ao meio contínuo Ψ. Resultarão n^{20} pontos em que $n = (s - i) / h$.

Sendo que contamos quantos dos pontos da malha verificam as inequações, obtemos o volume de Ω multiplicando o número de pontos por h^{20}. O problema é que se, por exemplo, discretizada cada dimensão em dez pontos (o que não é uma avaliação muito detalhada) se forem avaliados mil milhões de pontos por segundo, serão precisos 3170 anos para calcular o volume.

Na aplicação do MMC, primeiro sabe-se que o volume de Ψ é igual a $(s-i)^{20}$. Segundo, obtém-se o ponto $Xi = (x_1, \ldots, x_{20})$ sorteando 20 números de forma

independente e distribuição uniforme no intervalo $]i, s[$. Agora verifica-se se esse ponto está dentro ou fora de Ω. Terceiro, repete-se o sorteio de novos pontos e mede-se a proporção Q de pontos que cai dentro de Ω. Quarto, o volume de Ω será aproximado por $Q(s-i)^{20}$. O teorema limite central garante que se forem sorteados N pontos, então o erro terá distribuição normal e será proporcional a $1/\sqrt{N}$.

Como, por questões de economia, se pretende utilizar um processador digital *multi-purpose*, ter-se-á que substituir os números aleatórios por outros que pareçam aleatórios mas que são gerados uns a partir dos outros segundo uma série bem determinada. As séries usadas para gerar estes números pseudo-aleatórios são caóticas como, por exemplo, a função quadrática $x_t = 3{,}95\ (1-x_{t-1})\ x_{t-1}$.

Este exemplo mostra que se pode encontrar uma solução aproximada para os problemas complexos pela repetição pseudo-aleatória de fenómenos simples. A solução encontrada tem um erro que se rege pelas leis estatísticas.

Como o MMC teve um grande desenvolvimento no contexto do projecto Manhattan que levou à construção da bomba atómica americana, Ulam e Metropolis (1949) mostram como se aplica o MMC ao processo de "morte e nascimento" que acontece numa reacção nuclear em cadeia. Esta aplicação, por ser fundamental no desenvolvimento da bomba atómica, apenas foi desclassificada depois da *URSS* ter explodido a sua primeira bomba atómica.

3. Transitividade da causalidade de Granger

Neste ponto vamos investigar se existe transitividade quando há relações de causa efeito significativas entre variáveis. A noção de relação de causa efeito que adoptamos é a causalidade de Granger.

3.1. Causalidade de Granger

Quando se teoriza que uma variável x induz um efeito noutra variável y, em termos científicos, essa hipotética relação causa-efeito tem que ser avaliada por comparação com o que é observado. A causalidade de Granger usa séries temporais na avaliação da hipótese conjecturada. Assim, Granger (1969) propõe que, conhecidas séries temporais de x e y, x causa y se:

$$y_t = f(x_{t-1}, z) + \varepsilon_t \tag{3.1}$$

Nesta relação, z representa outras variáveis onde se pode incluir a variável y desfasada e ε_t é a parte não previsível do modelo.

Sendo que entre x e y existe uma relação estatisticamente significativa, então fica reforçada a hipotética relação causa-efeito.

Segundo Granger (1969), considera-se uma relação linear do tipo

$$y_t = \sum_{j=1}^{q} \beta_j y_{t-j} + \sum_{j=1}^{k} \gamma_j x_{t-j} + \varepsilon \tag{3.2}$$

Na avaliação da causalidade, consideram-se as seguintes hipóteses:

$$H_0 : \forall j, \gamma_j = 0 \quad versus \quad H_1 : \exists j, \gamma_j \neq 0 \tag{3.3}$$

Sendo que, sob H_0, se calcula a soma dos quadrados dos erros (RRSS – sem tratamento), e, sob H_1, se calcula também a soma dos quadrados dos erros (URSS – com tratamento), então, sob H_0, a estatística que mede o ganho médio por cada tratamento (estatística do tipo *F* de *Wald*) segue distribuição *F*(*k*,*T*–*k*) (distribuição *F* de *Snedecor* com *k* e *T*–*k* graus de liberdade):

$$F = \frac{(RRSS - URSS)/k}{URSS/(T-k)} \tag{3.4}$$

3.2. Uso da causalidade na previsão

Sendo que se conhecem os valores das séries temporais *x* e *y* até ao período *t* e se temos a convicção de que *x* causa *y* (se rejeita a hipótese H_0), então, o valor de y_{t+1} pode-se prever, em parte, a partir dos valores conhecidos de *x*.

Supondo o modelo de causalidade simplificado,

$$y_t = \beta\, y_{t-1} + \gamma\, x_{t-1} + \varepsilon_t \tag{3.5}$$

Como conhecemos y_t e x_t e uma estimativa para os parâmetros, podemos avançar um período para "fora da amostra":

$$E[y_{t+1}] = \hat{\beta}\, y_t + \hat{\gamma}\, x_t \tag{3.6}$$

A questão que se coloca é avaliar qual o erro de previsão para o período fora da amostra quando se rejeita H_0 para o nível de significância α. Chao, Corradi e Swanson (2001) consideram que o método de Monte Carlo é a ferramenta adequada a estudar esta questão.

A aplicação do método de Monte Carlo ao estudo da causalidade de Granger parte da construção de duas séries temporais com características estatísticas pré-determinadas, a amostra. No caso concreto são geradas as séries *y* e *x* dadas por:

$$\begin{cases} x_t = a_1 + a_2 x_{t-1} + \varepsilon_{1,t} \\ y_t = \pi_1 + \pi_2 y_{t-1} + \pi_3 x_{t-1} + \varepsilon_{2,t} \end{cases} \tag{3.7}$$

O valor dos parâmetros π_1, π_2, π_3, a_1, a_2 são conhecidos e $\varepsilon_{1,t}$ e $\varepsilon_{2,t}$ são extraídos aleatoriamente de uma função distribuição conhecida (por exemplo, a distribuição normal com média zero e variância σ^2 conhecida). Além disso, conhecem-se os valores iniciais da série, i.e. y_1 e x_1.

Assumindo agora que não se conhece como foi gerada a amostra, estimam-se os parâmetros do modelo sem tratamento e calcula-se a soma dos quadrados dos erros, *RRSS*:

$$e_t = y_t - (\hat{\pi}_1 + \hat{\pi}_2 y_{t-1}), \quad RRSS = \sum_1^N \left(e_t^2\right) \tag{3.8}$$

Estimam-se também os parâmetros do modelo com tratamento e calcula-se a soma dos quadrados dos erros, *URSS*:

$$e_t = y_t - (\hat{\pi}_1 + \hat{\pi}_2 y_{t-1} + \hat{\pi}_3 x_{t-1}), \quad URSS = \sum_1^N \left(e_t^2\right) \tag{3.9}$$

No teste de significância será usada a estatística que mede o ganho médio por cada tratamento (estatística do tipo *F* de *Wald*):

$$F = \frac{(RRSS - URSS)/k}{URSS/(T-k)} \tag{3.10}$$

Sob H_0 e na avaliação dentro da amostra, *F* segue a distribuição χ_k^2.

Este processo que compreende a geração da amostra, a estimação dos parâmetros e o testar a significância dos parâmetros é uma simulação.

Se inicialmente impusermos H_0, repetindo a simulação muitas vezes, a frequência relativa com que se rejeita H_0 é igual ao nível de significância α.

Podemos avaliar o uso da causalidade na previsão para fora da amostra comparando a significância com a frequência relativa com que se rejeita H_0 quando F é calculado com observações fora da amostra.

Como normalmente estima-se o modelo e a previsão é feita para o "período seguinte". E no "período seguinte", re-estima-se o modelo e aplica-se outra vez para prever mais um período para a frente, então na nossa simulação vamos utilizar uma estratégia idêntica:

1 - Gera-se uma série temporal com T períodos

2 - Consideramos os primeiros $N = (1-p)\,T$ períodos, em que $0 < p < 1$.

3 – Para $M = \{N, N+1, ..., T-1\}$ estima-se o modelo com os primeiros M períodos e calcula-se o erro sem tratamento, $e_{M+1} = y_{M+1} - (\hat{\pi}_1 + \hat{\pi}_2 y_M)$ e com tratamento, $f_{M+1} = y_{M+1} - (\hat{\pi}_1 + \hat{\pi}_2 y_M + \hat{\pi}_3 x_M)$. Notar que, no geral, as estimativas sem tratamento são diferentes das estimativas com tratamento.

4 – Calculam-se os erros $RRSS = \sum_{N+1}^{T} (e_t^2)$ $URSS = \sum_{N+1}^{T} (e_t^2)$ e com estes, calcula-se a estatística $F = \dfrac{(RRSS - URSS)}{URSS/(T-N-2)}$.

Notar que $k = 1$ e o número de termos usados no cálculo de F são $T - (N + 1)$.

5 – Rejeita-se H_0 se F for maior que o valor crítico correcto quando o F é calculado dentro da amostra.

Podemos agora comparar os valores críticos da distribuição χ^2 com um grau de liberdade (2,706, 3,841 e 6,635 para uma significância de 10%, 5% e 1%, respectivamente) com os valores críticos calculados com os valores previstos fora da amostra (ver tabela 2).

$\alpha \setminus p$	0,1	0,3	0,5
1%	2,027	3,722	5,100
5%	1,111	2,125	3,026
10%	0,746	1,496	2,197

Tabela 2 – Valores críticos "fora da amostra"

O facto dos valores o valor crítico quando a estatística F é calculada fora da amostra serem mais pequenos que quando calculados dentro da amostra traduz que o tratamento induz menor descida no erro. Sob H_0, isso é bom porque não existe

causalidade. No entanto, caso seja de rejeitar H_0, a previsão fora da amostra é pior que a previsão dentro da amostra. O erro de previsão será tanto maior quanto menor for a percentagem de casos usados no cálculo da estatística, (ver anexo 2, tabela 17).

Estes resultados estão de acordo com Chao, Corradi e Swanson (2001: Table 3).

3.3. Estudo da transitividade da causalidade de Granger

Neste ponto vamos investigar se sendo estatisticamente significativo que x causa z e que z causa y, até que ponto podemos dizer que será estatisticamente significativo que x causa y. Por exemplo, sabido que fumar aumenta de forma significativa o teor de monóxido de carbono no sangue e que o aumento de monóxido de carbono no sangue aumenta de forma significativa a probabilidade de ocorrência de acidentes vasculares, será que podemos afirmar que fumar aumenta de forma significativa a probabilidade de ocorrência de acidentes vasculares?

Em termos formais, a transitividade resulta da seguinte transformação algébrica:

$$\begin{cases} x_t = \beta_{1,0} + \beta_{1,1} x_{t-1} + \varepsilon_{1,t} \\ z_t = \beta_{2,0} + \beta_{2,1} z_{t-1} + \beta_{2,2} x_{t-1} + \varepsilon_{2,t} \\ y_t = \beta_{3,0} + \beta_{3,1} y_{t-1} + \beta_{3,2} z_{t-1} + \varepsilon_{3,t} \end{cases} \qquad (3.11)$$

$$\Rightarrow y_t = \delta_0 + \beta_{3,1} y_{t-1} + \delta_2 z_{t-2} + \delta_3 x_{t-2} + \xi_t$$

Invertendo $x_t = \beta_{1,0} + \beta_{1,1} x_{t-1} + \varepsilon_{1,t}$, termos então uma relação entre y_t e x_{t-1}:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 z_{t-2} + \varphi_3 x_{t-1} + \gamma_t \qquad (3.12)$$

Em termos empíricos, a questão que se põe é se, com base numa amostra, se se rejeitar a hipótese $H_0 : (\beta_{2,2} = 0 \vee \beta_{3,2} = 0)$ contra $H_1 : (\beta_{2,2} \neq 0 \wedge \beta_{3,2} \neq 0)$, automaticamente podemos rejeitar a hipótese $H_0' : \varphi_3 = 0$ contra $H_1' : \varphi_3 \neq 0$.

Fica claro que a transitividade engloba outros casos. Representando S que o parâmetro é significativo e N que não o é, haverá 8 casos a considerar (ver tabela 3).

$\beta_{2,2}$	S	S	N	N	S	S	N	N
$\beta_{3,2}$	S	N	S	N	S	N	S	N
φ_3	S	N	N	N	N	S	S	S
Transitividade?	Sim	Sim	Sim	Sim	Não	Não	Não	Não

Tabela 3 – Casos possíveis no estudo da transitividade

A aplicação do método de Monte Carlo à questão da transitividade da causalidade de Granger começa pela construção de três séries temporais com características estatísticas pré-determinadas, a amostra. No estudo, geramos as séries temporais assumindo que $(\beta_{2,2} = 0 \wedge \beta_{3,2} = 0)$, que todos os outros parâmetros são iguais e conhecidos, e que $\varepsilon_{1,t}$, $\varepsilon_{2,t}$ e $\varepsilon_{3,t}$ são extraídos da distribuição normal de média nula e variâncias unitária:

$$\begin{cases} x_t = 1 + \pi\, x_{t-1} + \varepsilon_{1,t} \\ z_t = 1 + \pi\, z_{t-1} + \varepsilon_{2,t} \\ y_t = 1 + \pi\, y_{t-1} + \varepsilon_{3,t} \end{cases} \tag{3.13}$$

Conduzimos simulações para todos os valores π do conjunto $\{0{,}1;\ 0{,}3;\ 0{,}5;\ 0{,}7;\ 0{,}9\}$, considerando amostras de tamanho $T = \{250, 500, 1000\}$ (um total de 15 casos). O valor de π traduz a "variabilidade" da amostra. Assim, para efeitos de estimação, quanto maior for (mas menor que 1), melhor será a qualidade da amostra.

Utilizando a metodologia de Granger (ver ponto 3.1.), estimamos cada um dos modelos (restritos e não restritos) e calculamos a estatísticas F de *Wald* e avaliamos a significância dos parâmetros para 1%, 2,5%, 5% e 10% de nível de significância.

Finalmente, verificamos a percentagem de casos em que se observava a não existência de transitividade.

Replicamos este processo 20000 vezes (ver os resultados na tabela 4).

De forma a ser possível avaliar o erro do cálculo, dividimos as 20000 réplicas em dois conjuntos de 10000 réplicas cada e determinamos o erro como $r = |s_1 - s_2|/\sqrt{2}$.

No caso concreto da tabela 4, estimamos que o erro médio é de 0,0927 pontos percentuais e o erro máximo é 0,3323 pontos percentuais.

Em termos de legenda das tabelas, temos (ver as expressões 3.11 e 3.12):

$$\begin{cases} eq.1 \;\; : z_t = \beta_{2,0} + \beta_{2,1} z_{t-1} + \beta_{2,2} x_{t-1} + \varepsilon_{2,t} \\ eq.2 \;\; : y_t = \beta_{3,0} + \beta_{3,1} y_{t-1} + \beta_{3,2} z_{t-1} + \varepsilon_{3,t} \\ eq.3a : y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 z_{t-2} + \varphi_3 x_{t-1} + \gamma_t \\ eq.3b : y_t = \delta_0 + \beta_{3,1} y_{t-1} + \delta_2 z_{t-2} + \delta_3 x_{t-2} + \xi_t \end{cases} \tag{3.14}$$

Podemos verificar na tabela 4 que, não existindo causalidade de Granger (sob H_0), a frequência relativa em que não se verifica a existência de transitividade é ligeiramente superior (cerca de 10%) ao nível de significância com que se testa a existência de causalidade. Assim sendo, os resultados reforça a conjectura de que existe transitividade na causalidade de Granger.

π	$\alpha \setminus T$	250		500		1000	
		eq. 3a	*eq. 3b*	*eq. 3a*	*eq. 3b*	*eq. 3a*	*eq. 3b*
0,1	1%	1,09%	1,16%	0,89%	1,07%	1,03%	1,00%
	2,5%	2,71%	2,87%	2,54%	2,66%	2,67%	2,48%
	5%	5,46%	5,70%	5,27%	5,35%	5,36%	4,92%
	10%	11,16%	11,29%	10,81%	11,23%	10,69%	10,39%
0,3	1%	1,15%	1,09%	1,01%	1,15%	1,03%	1,08%
	2,5%	2,74%	2,81%	2,74%	2,99%	2,69%	2,65%
	5%	5,53%	5,72%	5,46%	5,76%	5,29%	5,54%
	10%	11,06%	11,03%	11,03%	11,42%	10,73%	11,13%
0,5	1%	1,04%	1,22%	0,93%	1,08%	1,15%	1,01%
	2,5%	2,84%	2,88%	2,50%	2,55%	2,50%	2,63%
	5%	5,49%	5,54%	5,40%	5,26%	5,24%	5,31%
	10%	11,35%	11,34%	11,3%	10,72%	10,70%	10,96%
0,7	1%	1,22%	1,24%	1,19%	1,15%	0,96%	1,01%
	2,5%	2,82%	3,16%	2,92%	2,96%	2,50%	2,56%
	5%	5,90%	6,13%	5,53%	5,80%	5,20%	5,24%
	10%	11,96%	12,20%	11,14%	11,29%	10,80%	11,08%
0,9	1%	1,60%	1,66%	1,38%	1,32%	1,14%	1,19%
	2,5%	3,34%	3,89%	3,04%	3,35%	3,01%	3,04%
	5%	6,73%	7,57%	6,08%	6,42%	5,97%	6,07%
	10%	13,40%	14,18%	12,02%	12,78%	11,78%	12,24%

Tabela 4 – Percentagem de "não transitividade" em função de π, α e T (sob H_0)

Mas a transitividade resulta de a maior parte dos casos serem "não rejeita H_0, não rejeita H_0, não rejeita H_0". Sendo que o mais importante é a situação em que se rejeita H_0, vamos agora avaliar a frequência relativa em que se rejeita H_0 para os três modelos (mas ainda sob H_0). Quer isto dizer, x causa de forma significativa z, z causa de forma significativa y e x causa de forma significativa y. Podemos ver pelos resultados, (ver tabela 5), que a percentagem de casos "positivos" é muito baixa (cerca de 100 vezes menos frequente que o nível de significância).

No caso da tabela 5, estimamos que o erro médio do cálculo é de 0,0078 pontos percentuais e o erro máximo é 0,0495 pontos percentuais.

Os resultados não apresentam diferenças quando consideramos a *eq. 3a* ou a *eq. 3b* (ver expressão 3.14). Quer isto dizer que a transitividade, a existir, não aumenta o desfasamento do modelo.

Dado que nas tabelas 4 e 5 fizemos o estudo sob o pressuposto de se verificar H_0, então os casos "positivos" são falsos positivos. Em particular, quando se observa que "significativo + significativo $\Rightarrow$ significativo" estamos em presença de conclusões erradas (erro tipo I).

π	$\alpha \setminus T$	250		500		1000	
		eq. 3a	*eq. 3b*	*eq. 3a*	*eq. 3b*	*eq. 3a*	*eq. 3b*
0,1	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	5%	0,01%	0,01%	0,03%	0,02%	0,02%	0,04%
	10%	0,11%	0,07%	0,14%	0,13%	0,09%	0,13%
0,3	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,00%	0,00%	0,00%	0,00%	0,01%	0,01%
	5%	0,01%	0,02%	0,00%	0,01%	0,02%	0,02%
	10%	0,10%	0,12%	0,09%	0,12%	0,11%	0,11%
0,5	1%	0,00%	0,01%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,01%	0,01%	0,00%	0,00%	0,00%	0,00%
	5%	0,02%	0,03%	0,02%	0,03%	0,01%	0,01%
	10%	0,14%	0,13%	0,17%	0,14%	0,10%	0,12%
0,7	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,00%	0,00%	0,00%	0,01%	0,00%	0,00%
	5%	0,02%	0,01%	0,02%	0,01%	0,01%	0,01%
	10%	0,15%	0,08%	0,12%	0,12%	0,09%	0,08%
0,9	1%	0,01%	0,00%	0,00%	0,00%	0,01%	0,00%
	2,5%	0,01%	0,02%	0,01%	0,00%	0,01%	0,00%
	5%	0,02%	0,04%	0,04%	0,02%	0,01%	0,03%
	10%	0,22%	0,21%	0,18%	0,19%	0,08%	0,14%

Tabela 5 – Percentagem de "falsos positivos" em função de π, α e T

Dado que o estudo cujos resultados apresentamos nas tabelas 4 e 5 são feitos sob o pressuposto de se verificar H_0, então os casos "positivos" em que se observa "significativo + significativo $\Rightarrow$ significativo" são falsos positivos. Como nos interessa saber o que se passa quando de facto existe causalidade entre x e z e entre z e y, vamos então agora estudar a situação em que $\beta_{2,2} = 0{,}1$ e $\beta_{3,2} = 0{,}1$ (ver as expressões 3.11 e 3.12). O estudo conduzido quando não se verifica H_0, trata-se de uma avaliação da "potência do teste".

Repetindo o procedimento utilizado no caso em que H_0 era verdade (tabelas 4 e 5), obtivemos os a percentagem de vezes em que se rejeita H_0 em cada equação da expressão 3.14 (e que são verdadeiros "positivos").

π	$\alpha \setminus T$	250				500			
		eq. 1	*eq.* 2	*eq.* 3*a*	*eq.* 3*b*	*eq.* 1	*eq.* 2	*eq.* 3*a*	*eq.* 3*b*
0,1	1%	16,34%	16,63%	1,04%	1,18%	37,18%	37,12%	1,01%	0,98%
	2,5%	25,80%	25,95%	2,64%	2,75%	50,19%	50,37%	2,67%	2,47%
	5%	35,69%	35,82%	5,29%	5,44%	61,46%	61,79%	5,03%	5,25%
	10%	47,87%	47,99%	10,52%	10,61%	72,54%	72,77%	10,05%	10,19%
0,3	1%	18,12%	18,13%	1,03%	1,06%	40,38%	40,89%	1,09%	1,12%
	2,5%	27,70%	28,01%	2,71%	2,73%	53,96%	53,61%	2,61%	2,65%
	5%	37,66%	38,58%	5,12%	5,59%	64,64%	64,51%	5,21%	5,43%
	10%	49,92%	50,75%	10,42%	10,64%	75,23%	75,33%	10,27%	10,50%
0,5	1%	22,77%	23,10%	1,25%	1,29%	50,46%	49,72%	1,00%	1,10%
	2,5%	33,80%	34,11%	2,75%	2,95%	63,58%	62,55%	2,42%	2,54%
	5%	44,64%	44,53%	5,52%	5,73%	73,24%	72,68%	5,17%	5,15%
	10%	56,85%	57,04%	10,53%	10,73%	82,29%	82,08%	10,16%	10,26%
0,7	1%	35,18%	36,21%	1,28%	1,32%	70,12%	70,32%	1,18%	1,07%
	2,5%	47,71%	48,21%	2,84%	2,99%	80,15%	80,74%	2,88%	2,59%
	5%	58,61%	58,75%	5,56%	5,62%	86,91%	87,26%	5,26%	5,36%
	10%	70,22%	70,30%	10,69%	10,96%	92,24%	92,52%	10,40%	10,27%
0,9	1%	80,90%	81,12%	1,87%	1,78%	98,63%	98,69%	1,38%	1,38%
	2,5%	87,40%	87,43%	3,72%	3,72%	99,29%	99,36%	3,22%	3,09%
	5%	91,74%	91,57%	6,86%	6,96%	99,66%	99,67%	6,00%	6,06%
	10%	94,87%	94,94%	12,55%	12,90%	99,88%	99,84%	11,33%	11,95%

Tabela 6 – Potência do teste de *Wald* em função de π, α e T

Os resultados, ver tabela 6, mostram que a percentagem de aceitação das equações *eq.* 1 e *eq.* 2 aumenta com a extensão da série, com o nível de significância e com o valor de π. Por outro lado, a percentagem de aceitação das equações *eq.* 3*a* e *eq.* 3*b* são idênticas entre si e iguais ao nível de significância (como se estivéssemos sob H_0). Isto traduz que, em termos estatísticos, não existe relação causalidade entre x e y,

independentemente do desfasamento considerado para a variável x ser 1 período ou 2 períodos (ver expressões 3.11 e 3.12).

No caso da tabela 6, o erro médio do cálculo estimado é de 0,1537 pontos percentuais.

π	$\alpha \setminus T$	1000			
		eq. 1	*eq.* 2	*eq.* 3*a*	*eq.* 3*b*
0,1	1%	72,20%	72,53%	0,97%	1,00%
	2,5%	82,31%	82,42%	2,52%	2,57%
	5%	88,79%	88,46%	5,02%	5,10%
	10%	93,68%	93,62%	9,96%	9,98%
0,3	1%	76,66%	77,15%	1,03%	1,11%
	2,5%	85,29%	85,77%	2,57%	2,72%
	5%	90,65%	91,18%	5,26%	5,20%
	10%	95,02%	95,38%	10,42%	10,05%
0,5	1%	85,28%	85,57%	1,02%	0,96%
	2,5%	91,72%	91,72%	2,58%	2,42%
	5%	95,11%	95,03%	5,18%	4,94%
	10%	97,42%	97,53%	10,31%	9,97%
0,7	1%	96,17%	96,64%	1,05%	1,10%
	2,5%	98,14%	98,50%	2,71%	2,64%
	5%	99,10%	99,23%	5,11%	5,05%
	10%	99,65%	99,68%	9,87%	10,19%
0,9	1%	100,00%	100,00%	1,10%	1,19%
	2,5%	100,00%	100,00%	2,79%	2,86%
	5%	100,00%	100,00%	5,46%	5,62%
	10%	100,00%	100,00%	10,60%	10,87%

Tabela 6 (continuação) – Potência do teste de *Wald* em função de π, α e T

Sob H_1, a frequência com que se rejeita a existência de transitividade é crescente com a extensão da série, com a significância e com π. Em termos numéricos, a frequência de rejeição da existência de causalidade é bastante elevada, havendo casos em que é superior a 98%(ver tabela 7). Referente à tabela 7 o erro médio do cálculo estimado é 0,1937 pontos percentuais.

Podemos ainda avaliar a transitividade nos casos "verdadeiros positivos" (ver tabela 8). Apesar de ser muito frequente a aceitação (ver tabela 6), é bastante pequena a percentagem de casos em que se "rejeita H_0" na equação *eq.* 1 e "rejeita H_0" na equação *eq.* 2 implica que se "rejeita H_0" na equação *eq.* 3*a* ou eq. 3*b*. estes resultados enfraquecem a conjectura de que existe transitividade na causalidade de Granger. Para os resultados da tabela 8 estimamos o erro médio do cálculo em 0,0972 pp.

π	$\alpha \setminus T$	250		500		1000	
		eq. 3a	*eq. 3b*	*eq. 3a*	*eq. 3b*	*eq. 3a*	*eq. 3b*
0,1	1%	3,87%	3,60%	13,99%	13,79%	53,02%	52,18%
	2,5%	9,23%	9,00%	25,79%	26,03%	67,36%	66,92%
	5%	16,84%	16,63%	37,90%	38,87%	76,31%	75,79%
	10%	28,58%	28,25%	51,16%	52,27%	80,46%	80,31%
0,3	1%	4,34%	4,46%	17,53%	17,12%	58,63%	58,20%
	2,5%	10,45%	10,27%	30,49%	29,69%	71,86%	71,58%
	5%	18,27%	18,64%	43,04%	42,12%	79,49%	78,87%
	10%	30,49%	30,76%	55,78%	55,33%	81,88%	82,13%
0,5	1%	6,06%	6,41%	25,08%	25,28%	72,64%	72,66%
	2,5%	12,95%	13,63%	40,00%	39,99%	82,45%	82,42%
	5%	22,19%	22,93%	53,15%	52,95%	86,16%	86,16%
	10%	35,61%	38,82%	64,30%	64,21%	86,00%	86,17%
0,7	1%	13,84%	13,41%	48,79%	48,66%	92,10%	92,34%
	2,5%	25,26%	24,33%	63,34%	63,82%	94,18%	94,24%
	5%	36,77%	36,13%	72,70%	73,21%	93,29%	93,16%
	10%	48,91%	49,26%	77,77%	78,24%	89,30%	89,27%
0,9	1%	64,75%	64,42%	95,86%	95,97%	98,80%	98,76%
	2,5%	73,81%	74,16%	95,66%	95,63%	97,11%	97,13%
	5%	78,56%	78,44%	93,36%	93,19%	94,48%	94,55%
	10%	79,55%	78,82%	88,60%	88,04%	89,46%	89,33%

Tabela 7 – Percentagem de "não transitividade" em função de π, α e T (sob H_1)

π	$\alpha \setminus T$	250		500		1000	
		eq. 3a	*eq. 3b*	*eq. 3a*	*eq. 3b*	*eq. 3a*	*eq. 3b*
0,1	1%	0,02%	0,04%	0,13%	0,12%	0,57%	0,56%
	2,5%	0,18%	0,16%	0,74%	0,66%	1,75%	1,74%
	5%	0,73%	0,65%	1,96%	1,94%	4,09%	3,96%
	10%	2,48%	2,23%	5,36%	5,34%	9,03%	9,10%
0,3	1%	0,05%	0,04%	0,16%	0,19%	0,67%	0,62%
	2,5%	0,20%	0,23%	0,77%	0,81%	1,82%	1,97%
	5%	0,81%	0,71%	2,23%	2,18%	4,15%	4,27%
	10%	2,77%	2,44%	5,77%	5,82%	9,18%	9,06%
0,5	1%	0,09%	0,06%	0,28%	0,26%	0,80%	0,70%
	2,5%	0,39%	0,35%	1,05%	0,83%	2,04%	1,97%
	5%	1,20%	1,04%	2,79%	2,49%	4,50%	4,44%
	10%	3,65%	3,30%	6,69%	6,55%	9,46%	9,27%
0,7	1%	0,14%	0,13%	0,57%	0,58%	1,03%	0,95%
	2,5%	0,77%	0,70%	1,75%	1,84%	2,52%	2,52%
	5%	2,03%	2,09%	4,17%	4,11%	5,02%	5,24%
	10%	5,61%	5,72%	9,16%	9,18%	9,92%	10,49%
0,9	1%	1,19%	1,26%	1,56%	1,34%	1,13%	1,17%
	2,5%	3,06%	2,90%	3,46%	3,04%	2,88%	2,97%
	5%	6,17%	5,78%	6,30%	6,20%	5,55%	5,51%
	10%	11,94%	11,59%	11,69%	11,55%	10,61%	10,61%

Tabela 8 – Percentagem de "verdadeiros positivos" em função de π, α e T

3.4. Conclusão

Sendo que parece intuitivo que quando é estatisticamente significativo que x causa z e que z causa y se pode concluir que é estatisticamente significativo que x causa y, esta intuição tem que ser avaliada usando métodos rigorosos. Sendo que não é praticável o uso da metodologia matemático-dedutiva, optamos pela experimentação estatística conhecida como Método de Monte Carlo.

Os resultados a que chegamos permitem enfraquecer o que parece intuitivo: não existe transitividade na causalidade de Granger. Assim sendo, quando é estatisticamente significativo que x causa z e que z causa y, nada se pode dizer quanto à significância de x causar y.

Sendo que não existe transitividade entre duas relações de causalidade, por indução fica provado que não existirá transitividade entre $n > 2$ relações de causalidade.

4. Transitividade entre séries cointegradas

Nas séries económicas, normalmente verifica-se uma relação de dupla causalidade. Isto é, a variável x tem um efeito na variável y e a variável y também tem um efeito na variável x. Sendo assim, não se consegue identificar qual a variável que é causa e qual a que é efeito. Para ultrapassar esta questão surgiu o conceito de variáveis cointegradas. De forma semelhante ao estudo que apresentamos relativamente à transitividade da causalidade de Granger, neste ponto, vamos estudar a conjectura de que existe transitividade entre variáveis cointegradas. I.e., vamos estudar se quando é estatisticamente significativo que x e z são cointegradas e que z e y também são cointegradas, se pode concluir que é estatisticamente significativo que x e y são cointegradas.

4.1. Cointegração

Quando se teoriza que há uma relação entre a variável x e a variável y, mas não se consegue identificar em que sentido é a hipotética relação causa-efeito, i.e., o sentido da causalidade de Granger, podemos mesmo assim propor a existência de uma relação em que as variáveis se influenciam mutuamente. Em termos estatísticos, partindo das séries temporais de x e y, teremos o seguinte modelo de interacção mútua entre as variáveis:

$$\begin{cases} x_t = f(x_{t-1}, y_{t-1}, w_t) + \varepsilon_t \\ y_t = f(x_{t-1}, y_{t-1}, w_t) + \xi_t \end{cases} \tag{4.1}$$

Nesta relação, w_t representa outras variáveis e ε_t e ξ_t representam as partes não previsíveis do modelo (termos aleatórios).

Assumindo, sem perda de generalidade, que a relação é linear, que o vector w_t é o vector nulo (não há outras variáveis explicativas que não x e y desfasadas), e que o desfasamento é de um período, teremos o seguinte modelo:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \beta_{1,0} \\ \beta_{2,0} \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \xi_t \end{bmatrix} \tag{4.2}$$

Sendo que as séries temporais x e y não mostram tendência ao longo "tempo" (são estacionárias) e os parâmetros $\beta_{1,1}$ e $\beta_{2,2}$ são menores que a unidade (não têm raízes unitárias), então é possível testar a existência de cointegração entre as variáveis, Johansen (1988).

Sendo que se conjectura que as variáveis estão cointegradas, então, utilizando dados empíricos, podemos testar se os parâmetros do modelo 4.2 são estatísticamente significativos e, desta forma, reforçar ou enfraquecer a conjectura, Granger (1981).

A avaliação empírica da significância da cointegração não é trivial, baseando-se em "teoremas limite" (só para amostras de grande dimensão) que têm sofrido melhoramentos ao longo do tempo. No nosso estudo, sem perda de generalidade, vamos avaliar a existência de cointegração pela conjunção dos modelos individuais. Desta forma e dado que, sob H_0, os modelos são independentes, podemos avaliar a existência de cointegração pela combinação de duas estatísticas de *Wald*. A hipótese nula da cointegração a avaliar será

$$H_0 : \beta_{1,2} = 0 \vee \beta_{2,1} = 0 \quad versus \quad H_1 : \beta_{1,2} \neq 0 \wedge \beta_{2,1} \neq 0 \tag{4.3}$$

Como as séries são independentes, o nível de significância com que é testada a hipótese nula da expressão 4.3 é o produto dos níveis de significância com que são testadas as hipóteses nulas individuais. Por exemplo, se pretendermos testar a cointegração para um nível de significância de 5%, as séries individuais serão testadas com um nível de significância de 15.81%. Na tabela 9 mostramos os níveis de significância e os valores críticos adoptados nas experiências.

No ponto 3.1 pode ser vista a explicação da aplicação do teste de *Wald*.

α conjunto	α individual	Valor crítico
1%	10,00%	2,72
2,5%	15,81%	2,00
5%	22,36%	1,49
10%	31,62%	1,01

Tabela 9 – Níveis de significância e valores críticos adoptados

No sentido de avaliar a aplicabilidade do teste de *Wald* com os valores críticos da tabela 9, conduzimos uma experiência em que para $\pi = 0.1$, $N = 250$ e 60000 réplicas cujos resultados mostramos na tabela 10. Os resultados experimentais obtidos são próximos das percentagens conjecturadas pelo que será de aceitar a validade da aplicação do teste de *Wald* no estudo da cointegração.

α individual	Perc. rejeição de H_0 da cointegração
10,00%	1,14%
15,81%	2,67%
22,36%	5,12%
31,62%	9,93%

Tabela 10 – Níveis de significância "experimental"

Como nota, no caso das séries terem tendência da mesma ordem, em termos práticos é possível testar a cointegração diferenciando as séries m vezes até se obterem séries estacionárias. Neste caso, sendo as séries das diferenças cointegradas, diz-se que as variáveis são cointegradas de ordem m.

4.2. Estudo da transitividade da cointegração

Neste ponto vamos investigar se sendo estatisticamente significativo que x e z são cointegradas e que z e y também são cointegradas, até que ponto podemos dizer que será estatisticamente significativo que x e y são cointegradas. Por exemplo, sabido que, em termos estatísticos, por um lado, a altura e a quantidade de comida consumida são variáveis cointegradas e, por outro lado, o rendimento e a quantidade de comida consumida também são variáveis cointegradas, será que podemos afirmar que a altura e o rendimento são variáveis cointegradas?

Em termos formais, sendo que existem duas relações de cointegração com uma variável em comum,

$$\begin{cases} \begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} \beta_{1,0} \\ \beta_{2,0} \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \\ \begin{bmatrix} z_t \\ y_t \end{bmatrix} = \begin{bmatrix} \beta_{3,0} \\ \beta_{4,0} \end{bmatrix} + \begin{bmatrix} \beta_{3,1} & \beta_{3,2} \\ \beta_{4,1} & \beta_{4,2} \end{bmatrix} \begin{bmatrix} z_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{3,t} \\ \varepsilon_{4,t} \end{bmatrix} \end{cases} \tag{4.4}$$

A transitividade resulta da transformação algébrica do sistema de relações:

$$\begin{cases} x_t = \beta_{1,0} + \beta_{1,1} x_{t-1} + \beta_{1,2}\left[\beta_{3,0} + \beta_{3,1} z_{t-2} + \beta_{3,2} y_{t-2} + \varepsilon_{3,t}\right] + \varepsilon_{1,t} \\ y_t = \beta_{4,0} + \beta_{4,1}\left[\beta_{2,0} + \beta_{2,1} x_{t-2} + \beta_{2,2} z_{t-2} + \varepsilon_{2,t}\right] + \beta_{4,2} y_{t-1} + \varepsilon_{4,t} \end{cases} \tag{4.5}$$

Simplificando e invertendo as equações de forma a substituir y_{t-2} e x_{t-2} por y_{t-1} e x_{t-1},respectivamente, obtemos a relação de cointegração entre as variáveis x e y:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varphi_{1,0} \\ \varphi_{2,0} \end{bmatrix} + \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} \\ \varphi_{2,1} & \varphi_{2,2} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varphi_{1,3} \\ \varphi_{1,3} \end{bmatrix} \begin{bmatrix} z_{t-2} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{bmatrix} \tag{4.6}$$

Em termos empíricos, a questão que se põe é se, com base numa amostra, se se rejeitar a hipótese $H_0 : (\beta_{1,2} = 0 \vee \beta_{2,1} = 0 \vee \beta_{3,2} = 0 \vee \beta_{4,1} = 0)$ contra $H_1 : (\beta_{1,2} \neq 0 \wedge \beta_{2,1} \neq 0 \wedge \beta_{3,2} \neq 0 \wedge \beta_{4,1} \neq 0)$, automaticamente podemos rejeitar a hipótese $H_0' : \varphi_{1,2} = 0 \vee \varphi_{2,1} = 0$ contra $H_1' : \varphi_{1,2} \neq 0 \wedge \varphi_{2,1} \neq 0$.

Fica claro que a transitividade engloba outros casos em que pelo menos uma das relações de cointegração é não significativa. Representando S que os parâmetros são significativos e N que não o são, haverá 8 casos a considerar (ver tabela 11).

x e z cointegradas	S	S	N	N	S	S	N	N
z e y cointegradas	S	N	S	N	S	N	S	N
x e y cointegradas	S	N	N	N	N	S	S	S
Transitividade?	Sim	Sim	Sim	Sim	Não	Não	Não	Não

Tabela 11 – Casos possíveis no estudo da transitividade da cointegração

A aplicação do método de Monte Carlo à questão da transitividade da cointegração começa pela construção de três séries temporais com características estatísticas pré-determinadas, a amostra. No estudo, primeiro geramos as séries temporais assumindo que não existe cointegração (sob H_0). Em particular, assumimos que $\beta_{1,2} = 0 \wedge \beta_{2,1} = 0 \wedge \beta_{3,2} = 0 \wedge \beta_{4,1} = 0$. Assumimos ainda que todos os restantes parâmetros são iguais e conhecidos, e que $\varepsilon_{1,t}$, $\varepsilon_{2,t}$ e $\varepsilon_{3,t}$ são extraídos da distribuição normal de média nula e variâncias unitária de forma independente:

$$\begin{cases} x_t = 1 + \pi\ x_{t-1} + \varepsilon_{1,t} \\ z_t = 1 + \pi\ z_{t-1} + \varepsilon_{2,t} \\ y_t = 1 + \pi\ y_{t-1} + \varepsilon_{3,t} \end{cases} \tag{4.7}$$

Conduzimos simulações para todos os valores π do conjunto $\{0{,}1;\ 0{,}3;\ 0{,}5;\ 0{,}7;\ 0{,}9\}$, considerando amostras de tamanho $T = \{250, 500, 1000\}$ (um total de 15 casos). O valor de π traduz a "variabilidade" da amostra. Assim, para efeitos de estimação, quanto maior for (mas menor que 1), melhor será a qualidade da amostra.

Utilizando o método dos mínimos quadrados estimamos os parâmetros dos modelos (restrito e não restrito) e calculamos a estatísticas F de *Wald* e avaliamos a significância dos parâmetros para 1%, 2,5%, 5% e 10% de nível de significância.

Finalmente, verificamos a percentagem de casos em que se observa a não existência de transitividade usando o teste de *Wald* (ver o ponto 4.1.).

Replicamos este processo 20000 vezes (ver tabela 12). De forma a ser possível avaliar o erro do cálculo, dividimos as 20000 réplicas em dois conjuntos de 10000 réplicas cada e determinamos o erro como $r = |s_1 - s_2|/\sqrt{2}$. O erro médio é de 0,1193 pp.

Em termos de legenda das tabelas 12 e 13, "*rel. 3a*" representa a expressão 4.6 e "*rel. 3b*" representa a expressão 4.5.

π	$\alpha \setminus T$	250		500		1000	
		rel. 3a	*rel. 3b*	*rel. 3a*	*rel. 3b*	*rel. 3a*	*rel. 3b*
0,1	1%	1,10%	1,12%	1,08%	1,05%	0,95%	1,05%
	2,5%	2,76%	2,73%	2,62%	2,67%	2,49%	2,54%
	5%	5,49%	5,40%	5,14%	5,12%	5,03%	5,02%
	10%	11,21%	11,08%	10,64%	10,59%	10,51%	10,42%
0,3	1%	1,03%	1,06%	1,11%	0,96%	0,87%	0,96%
	2,5%	2,69%	2,76%	2,75%	2,52%	2,30%	2,42%
	5%	5,35%	5,57%	5,45%	5,28%	5,01%	5,01%
	10%	10,89%	11,20%	11,39%	10,78%	10,77%	10,71%
0,5	1%	1,13%	1,22%	0,93%	0,88%	1,06%	0,91%
	2,5%	2,70%	2,96%	2,47%	2,48%	2,78%	2,60%
	5%	5,31%	5,72%	5,16%	5,13%	5,24%	5,28%
	10%	11,12%	11,67%	10,93%	10,76%	10,78%	11,21%
0,7	1%	1,18%	1,12%	1,09%	1,13%	1,06%	1,04%
	2,5%	2,82%	2,82%	2,90%	2,70%	2,68%	2,50%
	5%	5,86%	5,77%	5,73%	5,11%	5,54%	5,13%
	10%	11,61%	11,74%	11,57%	10,75%	11,04%	11,02%
0,9	1%	1,78%	1,61%	1,34%	1,25%	1,02%	1,13%
	2,5%	3,97%	3,76%	3,19%	3,11%	2,81%	2,84%
	5%	7,38%	7,26%	6,32%	6,30%	5,67%	5,73%
	10%	13,64%	13,81%	12,49%	12,34%	11,64%	11,78%

Tabela 12 – Percentagem de "não transitividade" em função de π, α e T (sob H_0)

Como os resultados da tabela 12 traduzem que os casos de não transitividade da cointegração são da ordem de grandeza do nível de significância, a experiência realizada permite reforçar a conjectura de que existe transitividade na cointegração.

No entanto, a transitividade da experiência representada na tabela 12 resulta de a maior parte dos casos serem "não rejeita H_0, não rejeita H_0, não rejeita H_0". Como o mais importante é a situação em que se rejeita H_0 (casos "positivos"), vamos agora avaliar a frequência relativa em que se rejeita H_0 para as três relações de cointegração (mas ainda sob H_0). Quer isto dizer, se x está cointegrado de forma significativa com z, z está cointegrado de forma significativa com y e x está cointegrado de forma significativa com y.

Dado que nas tabelas 12 e 13 fizemos o estudo sob o pressuposto de se verificar H_0, então os casos "positivos" são falsos positivos. Em particular, quando se observa

que "significativo + significativo ⇒ significativo" estamos em presença de conclusões erradas (erro tipo I).

π	$\alpha \setminus T$	250		500		1000	
		Rel. 3a	*rel. 3b*	*rel. 3a*	*rel. 3b*	*rel. 3a*	*rel. 3b*
0,1	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	5%	0,03%	0,01%	0,02%	0,00%	0,01%	0,01%
	10%	0,08%	0,11%	0,12%	0,07%	0,10%	0,11%
0,3	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,01%	0,01%	0,00%	0,00%	0,00%	0,00%
	5%	0,03%	0,02%	0,00%	0,01%	0,02%	0,01%
	10%	0,10%	0,11%	0,10%	0,10%	0,12%	0,08%
0,5	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,00%	0,00%	0,00%	0,01%	0,00%	0,00%
	5%	0,02%	0,00%	0,03%	0,02%	0,02%	0,01%
	10%	0,12%	0,10%	0,10%	0,10%	0,09%	0,09%
0,7	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,00%	0,01%	0,00%	0,01%	0,01%	0,00%
	5%	0,01%	0,02%	0,01%	0,03%	0,02%	0,01%
	10%	0,12%	0,14%	0,15%	0,14%	0,09%	0,05%
0,9	1%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	2,5%	0,00%	0,01%	0,00%	0,01%	0,00%	0,01%
	5%	0,02%	0,03%	0,01%	0,04%	0,02%	0,02%
	10%	0,17%	0,17%	0,16%	0,16%	0,15%	0,09%

Tabela 13 – Percentagem de "falsos positivos" em função de π, α e T

Podemos ver pelos resultados, que a percentagem de casos "positivos" é muito baixa (cerca de 100 vezes menos frequente que o nível de significância), (ver tabela 13).

Os resultados não apresentam diferenças quando consideramos a *rel. 3a* ou a *rel. 3b*. Quer isto dizer que a transitividade, a existir, não aumenta o desfasamento do modelo. O erro médio do cálculo estimado é 0,0074 pp.

Como nos interessa saber o que se passa quando de facto existe cointegração entre as séries x e z e entre as séries z e y, vamos então agora estudar a situação em que $\beta_{1,2} = 0,1$; $\beta_{2,1} = 0,1$; $\beta_{3,2} = 0,1$ e $\beta_{4,1} = 0,1$ (ver a expressão 4.4). O estudo conduzido quando não se verifica H_0, trata-se de uma avaliação da "potência do teste".

Repetindo o procedimento utilizado no caso em que H_0 era verdade (tabelas 12 e 13), obtivemos os a percentagem de vezes em que se rejeita H_0 em cada relação da expressão 4.4 (*rel.* 1 e *rel.* 2) e na expressão 4.5 (*rel. 3b*) e 4.6 (*rel. 3a*) e que são verdadeiros "positivos".

π	$\alpha \setminus T$	250				500			
		rel. 1	*rel.* 2	*rel.* 3*a*	*rel.* 3*b*	*rel.* 1	*rel.* 2	*rel.* 3*a*	*rel.* 3*b*
0,1	1%	16,62%	16,52%	1,01%	1,14%	37,09%	36,76%	1,16%	1,01%
	2,5%	26,32%	26,14%	2,52%	2,59%	49,84%	49,71%	2,78%	2,55%
	5%	35,56%	35,73%	5,25%	5,13%	61,12%	61,08%	5,20%	5,00%
	10%	47,56%	47,73%	10,25%	10,39%	72,60%	72,38%	10,17%	10,44%
0,3	1%	18,06%	18,06%	1,01%	1,15%	40,29%	40,73%	1,10%	0,98%
	2,5%	28,20%	27,81%	2,56%	2,75%	53,37%	53,31%	2,54%	2,46%
	5%	37,97%	37,76%	5,19%	5,50%	64,24%	64,10%	5,28%	5,01%
	10%	50,03%	49,71%	10,42%	10,74%	74,92%	75,00%	10,57%	10,08%
0,5	1%	22,41%	22,88%	1,13%	1,11%	49,70%	50,32%	1,13%	1,05%
	2,5%	33,50%	33,52%	2,83%	2,69%	62,69%	63,19%	2,72%	2,53%
	5%	44,21%	44,26%	5,43%	5,59%	72,73%	73,54%	5,48%	5,19%
	10%	56,84%	56,42%	10,46%	10,71%	82,05%	82,66%	10,77%	10,13%
0,7	1%	35,43%	35,97%	1,22%	1,31%	70,40%	69,93%	1,31%	1,11%
	2,5%	47,99%	48,38%	2,79%	2,91%	80,87%	80,30%	3,02%	2,84%
	5%	58,47%	59,11%	5,66%	5,82%	87,40%	86,79%	5,52%	5,54%
	10%	69,33%	70,29%	11,11%	11,24%	92,69%	92,51%	10,51%	10,73%
0,9	1%	80,74%	81,00%	1,71%	1,71%	98,86%	98,62%	1,44%	1,35%
	2,5%	87,19%	87,37%	3,95%	3,71%	99,49%	99,31%	3,26%	3,18%
	5%	91,28%	91,16%	7,14%	6,99%	99,74%	99,62%	6,05%	6,05%
	10%	94,67%	94,60%	13,14%	12,73%	99,88%	99,84%	11,52%	11,59%

Tabela 14 – Potência do teste de *Wald* em função de π, α e T

π	$\alpha \setminus T$	1000			
		rel. 1	*rel.* 2	*rel.* 3*a*	*rel.* 3*b*
0,1	1%	72,49%	72,34%	1,15%	0,94%
	2,5%	82,30%	82,11%	2,64%	2,55%
	5%	88,58%	88,43%	5,08%	5,00%
	10%	93,39%	93,57%	10,03%	10,23%
0,3	1%	76,80%	76,93%	1,07%	0,99%
	2,5%	85,48%	85,65%	2,45%	2,75%
	5%	91,06%	90,96%	4,93%	5,25%
	10%	95,18%	95,19%	9,89%	10,05%
0,5	1%	85,29%	85,92%	1,12%	0,99%
	2,5%	91,52%	92,20%	2,85%	2,53%
	5%	95,08%	95,51%	5,29%	5,02%
	10%	97,65%	97,68%	10,59%	10,28%
0,7	1%	96,22%	96,31%	1,08%	1,11%
	2,5%	98,19%	98,30%	2,61%	2,62%
	5%	99,11%	99,23%	5,32%	5,11%
	10%	99,66%	99,68%	11,02%	10,12%
0,9	1%	100,00%	100,00%	1,15%	1,21%
	2,5%	100,00%	100,00%	2,75%	2,74%
	5%	100,00%	100,00%	5,56%	5,56%
	10%	100,00%	100,00%	10,97%	10,76%

Tabela 14 (continuação) – Potência do teste de *Wald* em função de π, α e T

Os resultados, ver tabela 14, mostram que a percentagem de aceitação das relações de cointegração *rel.* 1 e *rel.* 2 aumentam com a extensão da série, com o nível de significância e com π. Por outro lado, a percentagem de aceitação das equações *eq.* 3*a* e *eq.* 3*b* são idênticas entre si e iguais ao nível de significância (como se estivéssemos sob H_0). Isto traduz que, em termos estatísticos, não existe relação de cointegração das variáveis *x* e *y*, independentemente do desfasamento considerado para a variável *x* e y ser 1 período ou 2 períodos (ver equações 4.5 e 4.6). Referente à tabela 14 estimamos o erro médio do cálculo em 0,1481 pp.

Sob H_1, a frequência com que se rejeita a existência de transitividade aumenta com a extensão da série, com o nível de significância e com π (ver tabela 15). Para a tabela 15 estimamos o erro médio em 0,272 pp.

π	$\alpha \setminus T$	250		500		1000	
		rel. 3*a*	*rel.* 3*b*	*rel.* 3*a*	*rel.* 3*b*	*rel.* 3*a*	*rel.* 3*b*
0,1	1%	5,69%	5,79%	27,44%	27,59%	76,23%	76,19%
	2,5%	12,12%	12,16%	39,77%	40,15%	83,47%	83,70%
	5%	20,30%	20,41%	50,91%	51,19%	86,05%	86,41%
	10%	32,06%	31,62%	60,18%	60,50%	85,15%	85,42%
0,3	1%	6,75%	6,97%	31,36%	32,28%	81,39%	80,74%
	2,5%	14,16%	14,14%	44,79%	45,48%	87,04%	86,71%
	5%	22,46%	22,69%	55,18%	55,56%	88,81%	88,48%
	10%	33,80%	34,10%	63,78%	64,38%	86,67%	86,85%
0,5	1%	10,88%	10,85%	43,63%	44,64%	89,68%	89,55%
	2,5%	20,07%	19,41%	56,62%	57,55%	92,12%	92,11%
	5%	29,39%	29,53%	65,84%	65,88%	91,99%	91,76%
	10%	41,02%	41,36%	71,81%	71,89%	88,55%	88,27%
0,7	1%	24,59%	24,43%	71,91%	71,57%	97,63%	97,53%
	2,5%	36,26%	36,06%	80,05%	79,69%	96,74%	96,91%
	5%	46,80%	46,07%	83,62%	83,04%	94,32%	94,67%
	10%	56,75%	55,71%	83,57%	82,87%	89,45%	89,52%
0,9	1%	80,02%	79,90%	98,09%	98,27%	98,81%	98,87%
	2,5%	84,34%	84,08%	96,50%	96,72%	97,25%	97,21%
	5%	84,95%	84,88%	93,93%	94,10%	94,67%	94,53%
	10%	82,84%	82,73%	88,63%	88,88%	89,24%	89,39%

Tabela 15 – Percentagem de "não transitividade" em função de π, α e T (sob H_1)

Podemos agora avaliar a transitividade nos casos "verdadeiros positivos". Apesar de ser muito frequente a aceitação (ver tabela 14), é bastante pequena a percentagem de casos em que se "rejeita H_0" na equação *rel.* 1 e "rejeita H_0" na equação *rel.* 2 implica que se "rejeita H_0" na equação *rel.* 3*a* ou rel. 3*b*. estes resultados

enfraquecem a conjectura de que existe transitividade na cointegração. Referente à tabela 16, estimamos o erro médio em 0,0832 pp.

π	$\alpha \setminus T$	250		500		1000	
		rel. 3a	*rel. 3b*	*rel. 3a*	*rel. 3b*	*rel. 3a*	*rel. 3b*
0,1	1%	0,05%	0,05%	0,33%	0,33%	0,78%	0,81%
	2,5%	0,25%	0,25%	1,16%	1,00%	2,14%	2,07%
	5%	0,81%	0,85%	2,69%	2,51%	4,45%	4,28%
	10%	2,75%	2,78%	6,42%	6,39%	9,48%	9,10%
0,3	1%	0,07%	0,10%	0,31%	0,31%	0,85%	0,83%
	2,5%	0,27%	0,35%	1,18%	1,12%	2,36%	2,18%
	5%	0,87%	1,07%	2,78%	2,94%	4,61%	4,52%
	10%	2,94%	3,20%	6,78%	7,07%	9,47%	9,63%
0,5	1%	0,15%	0,12%	0,45%	0,46%	0,96%	1,04%
	2,5%	0,51%	0,49%	1,56%	1,51%	2,56%	2,63%
	5%	1,33%	1,36%	3,59%	3,43%	5,05%	5,17%
	10%	3,93%	3,98%	8,01%	7,80%	9,98%	10,16%
0,7	1%	0,28%	0,27%	0,80%	0,71%	1,15%	1,06%
	2,5%	1,01%	0,98%	2,18%	2,06%	2,63%	2,63%
	5%	2,52%	2,56%	4,74%	4,39%	5,12%	5,11%
	10%	6,36%	6,00%	9,56%	9,67%	10,07%	9,75%
0,9	1%	1,46%	1,40%	1,27%	1,27%	1,21%	1,19%
	2,5%	3,27%	3,26%	3,07%	3,16%	2,73%	2,74%
	5%	6,29%	6,07%	5,90%	5,97%	5,35%	5,31%
	10%	12,00%	11,63%	11,27%	11,30%	10,59%	10,45%

Tabela 16 – Percentagem de "verdadeiros positivos" em função de π, α e T

4.3. Conclusão

Sendo que parece intuitivo que quando é estatisticamente significativo que x está cointegrado com z e que z está cointegrado com y se pode concluir que é estatisticamente significativo que x está cointegrado com y, esta intuição tem que ser avaliada usando métodos rigorosos. Sendo que não é praticável o uso da metodologia matemático-dedutiva, optamos pela experimentação estatística conhecida como Método de Monte Carlo.

Os resultados a que chegamos permitem enfraquecer o que parece intuitivo: não existe transitividade na cointegração. Assim sendo, quando é estatisticamente significativo que x está cointegrado com z e que z está cointegrado com y, nada se pode dizer quanto à significância de x está cointegrado com y.

Sendo que não existe transitividade entre duas relações cointegradas, por indução fica provado que não existirá transitividade entre $n > 2$ relações cointegradas.

5. Conclusão

Neste trabalho propusemo-nos estudar a existência de transitividade entre variáveis que, em termos estatísticos, estão relacionadas. Em particular, pensamos em estudar a existência de transitividade entre variáveis ligadas por causalidade de Granger e entre variáveis cointegradas. Em termos intuitivos será de aceitar a existência de transitividade.

Sendo que, no contexto do problema que queríamos estudar, a modelização matemática e consequente manipulação algébrica se nos pareceu complexa e de aplicação limitada, adoptamos como metodologia a experimentação estatística que é conhecida na literatura como Método de Monte Carlo. No sentido de enquadrarmos este método, apresentamos no capítulo 2 deste trabalho, numa perspectiva histórica, um resumo dos trabalhos dos principais autores pioneiros.

O Método de Monte Carlo tem grande potencial de aplicação porque usa cálculo computacional cujo custo tem diminuído rapidamente.

No estudo da transitividade da causalidade de Granger, considerando três variáveis estatísticas que traduzem séries temporais, x, z e y, em que x causa de forma significativa z e z causa de forma significativa y, nada podemos dizer quanto a x causar de forma significativa y. No estudo da transitividade da cointegração, considerando da mesma forma três variáveis estatísticas que traduzem séries temporais, x, z e y, em que x e z estão cointegradas de forma significativa e z e y estão cointegradas de forma significativa, nada podemos dizer quanto a x e y estarem cointegradas de forma significativa.

Dados os resultados, concluímos que, ao contrário do que parece intuitivo, será de rejeitar a conjectura de que existe transitividade entre relações mesmo que estas sejam estatisticamente significativas.

Referências bibliográficas

Buffon, G, (1733), Nota do Editor, *Histoire de l'Academie Royal des Sciences*, pp. 43-45 (Sobre uma palestra apresentada na Academia Real de Ciências de Paris).

Buffon, G. (1777), "Essai d'arithmétique morale", *Histoire naturelle, générale et particulière*, Sup. 4, pp. 46-123.

Courant, R., K. Friedrichs e H. Levy (1967), "On Partial Difference Equations of Mathematical Physics", *IBM journal*, 11, pp. 215-234.

Ulam, S. e J. Von Neumann (1946), documento não publicado

Ulam, S. (1982), "John Von Neumann, 1903-1957", *Annals of the History of Computing*, 4, pp. 157-181.

Eckhardt, R. (1987), "Stan Ulam, John Von Neumann, and 'The Monte Carlo Method'", *Los Alamos Science*, Special Issue, 15, pp. 131-137.

Fermi, E. (1938), "Artificial Radioactivity Produced by Neutron Bombardment", *Nobel Lectures - Physics*, pp. 1-8.

Fermi, E. (1934), "Radioattività indotta da bombardamento di neutroni", *La Ricerca Scientifica*, 5, p. 238.

Kalos, M. H. e P.A. Whitlock (1986), *Monte Carlo Methods: Volume I: Basics*, John Wiley & Sons: New York.

Kelvin (1901), "Nineteenth century clouds over the dynamical theory of heat and light", *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 2, pp. 1-40.

Kolmogorov, A.N. (1931- em Russo), "On analytical methods in probability theory" in Shiryayev, A.N., ed, (1992), *Selected works of A, N, Kolmogorov*, II, Kluwer Academic Publications, pp. 62-108.

Maxwell, J.C. (1860), "Illustrations of the dynamical theory of gases", *Philosophical Magazine*, 20, pp. 21-37.

Metropolis, N. (1987), "The Beginning of the Monte Carlo Method", *Los Alamos Science Issue*, pp. 125-130.

Metropolis, N. e S. Ulam (1949), "The Monte Carlo method", *Journal of the American Statistical Association*, 44, pp. 335-341.

Pearson, K. (1896), "Mathematical Contributions to the Theory of Evolution III, Regression, Heredity, and Panmixia", *Philosophical Transactions of the Royal Society of London*, 187, pp. 253-318.

Rayleigh (1899), "On James Bernoulli's Theorem in Probabilities", *Philosophical Magazine*, 47, pp. 246-251.

Ross, S.M. (1976), *A First Course in Probability*, 2nd Ed,, Macmillan: New York.

Segre, E., ed. (1962). *The Collected Works of Enrico Fermi*, I e II, The University of Chicago Press: Chicago.

Student, T. (1908a), "Probable error of a correlation coefficient", *Biometrika*, 6, pp. 302-310.

Student, T, (1908b), "The probable error of a mean", *Biometrika*, 6, pp. 1-25.

Ulam, S., R.D. Richtmeyer e J. Von Neumann (1947), "Statistical methods in neutron diffusion", *Los Alamos Scientific Laboratory Report*, LAMS – 551.

Granger, C.W.J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification", *Journal of Econometrics*, 16, pp.121-130.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, No. 2-3, p231-254.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross Spectral Methods", *Econometrica*, 37, pp. 424-438.

Chao, J., V. Corradi e N. Swanson (2001), "An Out of Sample Test for Granger Causality", *Macroeconomic Dynamics*, 5, pp. 598-620.

Eells, E. e E. Sober (1983), "Probabilistic Causality and the Question of Transitivity", *Philosophy of Science*, 50: 35–57.

Engle, R. F. e C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, Vol. 55, pp251-276.

Anexo 1 – Programas de MatLab

Programa 1 – Simulação do "problema da agulha"

```
%Programa em Matlab, que faz varias simulações de atirar a agulha varias vezes,
%na 1º simulação a agulha é atirada 10 vezes, na 2º simulação a agulha é atirada 20
%vezes, na 3º , 30 vezes, etc na 30ª simulação , é atirada 300 vezes. O programa dá ao
%fim de todas as simulções o valor encontrado para o pi e o erro padrão.

%L(agulha)=1 e W(distancia entre as linhas)=1
clear
for N=1:30; %nº de simulações
   m=0;%contador das vezes que intersecta
for i=1:N*10; %nº de vezes
   distance=rand; % nº aleatorio com distribuiçao U(0,1)
   if distance>0.5
      distance=1-distance;
   end
   theta=rand;
   theta=theta*pi;

   if distance<sin(theta)/2
      m=m+1;
   end
end
M(N)=2*N*10/m;
end
plot(1:30,M)
media=mean(M);
desvio_pad=std(M);
disp(['media do valor de pi:',num2str(media),',desvio
padrão:',num2str(desvio_pad)])
```

Programa 2 – Simulação dos valores críticos "fora da amostra"

```
clear
p2=[0.1 0.3 0.5 0.7 0.9];
T=[250 500 1000];
R=input('Periodo fora da amostra 0.1, 0.3, 0.5');
for i=1:3;
    n=R*T(i) ;
    m=(2*T(i)-n-1)/2;
    for ii=1:5;
        for Numsim=1:10000;
            %gerar os dados
            x(1)=1;
            y(1)=1;
            for a=2:T(i);
                x(a)=1+p2(ii)*x(a-1)+randn;
                y(a)=1+p2(ii)*y(a-1)+randn;
            end
            erro1=0;
            erro2=0;
            %para o modelo u:
            for j=T(i)-n:T(i)-1;
                clear X
                X(1,:)=y(1:j);
                X(2,:)=x(1:j);
                y1=y(2:j+1);
                a2=zeros(3,3);
                a2(1,1)=j;
                for k=2:3;
                    a2(1,k)=sum(X(k-1,:));
                    a2(k,1)=a2(1,k);
                    for l=2:3;
                        a2(l,k)=sum(X(l-1,:).*X(k-1,:));
                    end
                end
                b(1)=sum(y1);
                for k=2:3;
                    b(k)=sum(X(k-1,:).*y1);
                end
                u=a2^(-1)*b';
                Prevyu=u(1)+u(2)*y(j)+u(3)*x(j);
                erro1=erro1+(Prevyu-y(j+1))^2;
                %para o modelo r:
                nu=polyfit(X(1,:),y1,1);
                Prevyr=nu(2)+nu(1)*y(j);
                erro2=erro2+(Prevyr-y(j+1))^2;
            end
            F(Numsim)=(erro2-erro1)/erro1*(n-2);
        end
        F=sort(F);
        disp(['obs  ', num2str(T(i)),'p2i
',num2str(p2(ii)),num2str(F(9000)),num2str(F(9500)),num2str(F(9900))])
    end
end
```

Programa 3 – Gera os dados, estima os parâmetros dos modelos, calcula os erros e obtém uma estimativa da estatística de *Wald*, para o estudo da transitividade na causalidade.

```
clear
p2=[0.1 0.3 0.5 0.7 0.9];
T=input('Nº de observaçoes');
for i=1:5;
    for ii=1:2;
        r1=0;
        r2=0;
        r3=0;
        r4=0;
        for Numsim=1:10000;
            erro1u=0;
            erro1r=0;
            erro2u=0;
            erro2r=0;
            erro3u=0;
            erro3r=0;
           %gerar os dados
           x(1)=1;
           y(1)=1;
           z(1)=1;
           for a=2:T;
               x(a)=1+p2(i)*x(a-1)+randn;
               z(a)=1+p2(i)*z(a-1)+randn;
               y(a)=1+p2(i)*y(a-1)+randn;
           end
          %para o modelo 1u:
           clear X
           X(1,:)=z(1:T-1);
           X(2,:)=x(1:T-1);
           y1=z(2:T);
           a1=zeros(3,3);
           a1(1,1)=T-1;
           for k=2:3;
               a1(1,k)=sum(X(k-1,:));
               a1(k,1)=a1(1,k);
               for l=2:3;
                   a1(l,k)=sum(X(l-1,:).*X(k-1,:));
               end
           end
           b(1)=sum(y1);
           for k=2:3;
               b(k)=sum(X(k-1,:).*y1);
           end
           u=a1^(-1)*b';
           %para o modelo 1r:
           nu=polyfit(X(1,:),y1,1);
           for j=1:T-1;
               Prevyu=u(1)+u(2)*z(j)+u(3)*x(j);
               erro1u=erro1u+(Prevyu-z(j+1))^2;
               Prevyr=nu(2)+nu(1)*z(j);
               erro1r=erro1r+(Prevyr-z(j+1))^2;
           end
           %para o modelo 2u:
```

```
    clear X
    X(1,:)=y(1:T-1);
    X(2,:)=z(1:T-1);
    y1=y(2:T);
    a1=zeros(3,3);
    a1(1,1)=T-1;
    for k=2:3;
     a1(1,k)=sum(X(k-1,:));
     a1(k,1)=a1(1,k);
     for l=2:3;
         a1(l,k)=sum(X(l-1,:).*X(k-1,:));
     end
end
b(1)=sum(y1);
for k=2:3;
    b(k)=sum(X(k-1,:).*y1);
end
u=a1^(-1)*b';
%para o modelo 2r:
nu=polyfit(X(1,:),y1,1);
for j=1:T-1;
    Prevyu=u(1)+u(2)*y(j)+u(3)*z(j);
    erro2u=erro2u+(Prevyu-y(j+1))^2;
    Prevyr=nu(2)+nu(1)*y(j);
    erro2r=erro2r+(Prevyr-y(j+1))^2;
end
%para o modelo 3u:
clear X
X(1,:)=y(2:T-1);
X(2,:)=z(1:T-2);
X(3,:)=x(2:T-1);
y1=y(3:T);
a1=zeros(4,4);
a1(1,1)=T-2;
for k=2:4;
    a1(1,k)=sum(X(k-1,:));
    a1(k,1)=a1(1,k);
    for l=2:4;
        a1(l,k)=sum(X(l-1,:).*X(k-1,:));
    end
end
b1(1)=sum(y1);
for k=2:4;
    b1(k)=sum(X(k-1,:).*y1);
end
u=a1^(-1)*b1';
%para o modelo 3r:
clear X
X(1,:)=y(2:T-1);
X(2,:)=z(1:T-2);
y1=y(3:T);
a11=zeros(3,3);
a11(1,1)=T-2;
for k=2:3;
    a11(1,k)=sum(X(k-1,:));
    a11(k,1)=a11(1,k);
    for l=2:3;
        a11(l,k)=sum(X(l-1,:).*X(k-1,:));
    end
```

```
        end
        b11(1)=sum(y1);
        for k=2:3;
            b11(k)=sum(X(k-1,:).*y1);
        end
        nu=a11^(-1)*b11';
        for j=2:T-1;
            Prevyu=u(1)+u(2)*y(j)+u(3)*z(j-1)+u(4)*x(j);
            erro3u=erro3u+(Prevyu-y(j+1))^2;
            Prevyr=nu(1)+nu(2)*y(j)+nu(3)*z(j-1);
            erro3r=erro3r+(Prevyr-y(j+1))^2;
        end
        F1=(erro1r-erro1u)/erro1u*(T-1);
        F2=(erro2r-erro2u)/erro2u*(T-1);
       F3=(erro3r-erro3u)/erro3u*(T-2);
...
```

Programa 4 – Calcula o nº de casos em que não há transitividade na causalidade.

```
...
if (F1>2.7055 & F2>2.7055 & F3<=2.7055)|(F1>2.7055 & F2<=2.7055 &
F3>2.7055)|(F1<=2.7055 & F2>2.7055 & F3>2.7055)|(F1<=2.7055 &
F2<=2.7055 & F3>2.7055);
            r1=r1+1;
        end
        if (F1>3.8415 & F2>3.8415 & F3<=3.8415)|(F1>3.8415 &
F2<=3.8415 & F3>3.8415)|(F1<=3.8415 & F2>3.8415 &
F3>3.8415)|(F1<=3.8415 & F2<=3.8415 & F3>3.8415);
             r2=r2+1;
        end
        if (F1>5.0239 & F2>5.0239 & F3<=5.0239)|(F1>5.0239 &
F2<=5.0239 & F3>5.0239)|(F1<=5.0239 & F2>5.0239 &
F3>5.0239)|(F1<=5.0239 & F2<=5.0239 & F3>5.0239);
                 r3=r3+1;
        end
        if (F1>6.6349 & F2>6.6349 & F3<=6.6349)|(F1>6.6349 &
F2<=6.6349 & F3>6.6349)|(F1<=6.6349 & F2>6.6349 &
F3>6.6349)|(F1<=6.6349 & F2<=6.6349 & F3>6.6349) ;
                 r4=r4+1;
        end
    end
    disp('10%, 5%, 2,5%, 1%')
    disp([r1,r2,r3,r4])
  end
end
```

Programa 5 – Calcula o nº de casos de significativos da transitividade na causalidade.

```
  ...
  if (F1>2.7055 & F2>2.7055 & F3>2.7055);
            r1=r1+1;
   end
   if (F1>3.8415 & F2>3.8415 & F3>3.8415);
             r2=r2+1;
   end
   if (F1>5.0239 & F2>5.0239 & F3>5.0239);
```

```
                r3=r3+1;
    end
    if (F1>6.6349 & F2>6.6349 & F3>6.6349);
                r4=r4+1;
    end
    end
    disp('10%, 5%, 2,5%, 1%')
    disp([r1,r2,r3,r4])
  end
end
```

Programa 6 – Gera os dados, estima os parâmetros dos modelos, calcula os erros e obtém uma estimativa da estatística de *Wald*, para o estudo da potência dos testes efectuados acerca da transitividade da causalidade.

```
clear
p2=[0.1 0.3 0.5 0.7 0.9];
T=input('Nº de observaçoes');
for i=1:5;
      for ii=1:2;
          r1=0;
          r2=0;
          r3=0;
          r11=0;
          r22=0;
          r33=0;
          r111=0;
          r222=0;
          r333=0;
          r1111=0;
          r2222=0;
          r3333=0;
          for Numsim=1:10000;
              erro1u=0;
              erro1r=0;
              erro2u=0;
              erro2r=0;
              erro3u=0;
              erro3r=0;
              %gerar os dados
              x(1)=1;
              y(1)=1;
              z(1)=1;
              for a=2:T;
                  x(a)=1+p2(i)*x(a-1)+randn;
                  z(a)=1+p2(i)*z(a-1)+randn;
                  y(a)=1+p2(i)*y(a-1)+randn;
             end
             %para o modelo 1u:
             clear X
             X(1,:)=z(1:T-1);
             X(2,:)=x(1:T-1);
             y1=z(2:T);
             a1=zeros(3,3);
             a1(1,1)=T-1;
             for k=2:3;
```

```
        a1(1,k)=sum(X(k-1,:));
        a1(k,1)=a1(1,k);
        for l=2:3;
            a1(l,k)=sum(X(l-1,:).*X(k-1,:));
        end
    end
    b1(1)=sum(y1);
    for k=2:3;
        b1(k)=sum(X(k-1,:).*y1);
    end
    u1=a1^(-1)*b1';
    %para o modelo 1r:
    y11=y1-0.1*X(2,:);
    nu1=polyfit(X(1,:),y11,1);
    for j=1:T-1
        Prevyu=u1(1)+u1(2)*z(j)+u1(3)*x(j);
        erro1u=erro1u+(Prevyu-z(j+1))^2;
        Prevyr=nu1(2)+nu1(1)*z(j)+0.1*x(j);
        erro1r=erro1r+(Prevyr-z(j+1))^2;
    end
    %para o modelo 2u:
    clear X
    X(1,:)=y(1:T-1);
    X(2,:)=z(1:T-1);
    y2=y(2:T);
    a2=zeros(3,3);
    a2(1,1)=T-1;
    for k=2:3;
        a2(1,k)=sum(X(k-1,:));
        a2(k,1)=a2(1,k);
        for l=2:3;
            a2(l,k)=sum(X(l-1,:).*X(k-1,:));
        end
    end
    b2(1)=sum(y2);
    for k=2:3;
        b2(k)=sum(X(k-1,:).*y2);
    end
    u2=a2^(-1)*b2';
    %para o modelo 2r:
    y12=y2-0.1*X(2,:);
    nu2=polyfit(X(1,:),y12,1);
    for j=1:T-1;
        Prevyu=u2(1)+u2(2)*y(j)+u2(3)*z(j);
        erro2u=erro2u+(Prevyu-y(j+1))^2;
        Prevyr=nu2(2)+nu2(1)*y(j)+0.1*z(j);
        erro2r=erro2r+(Prevyr-y(j+1))^2;
    end
    %para o modelo 3u:
    clear X
    X(1,:)=y(2:T-1);
    X(2,:)=z(1:T-2);
    X(3,:)=x(1:T-2);
    y3=y(3:T);
    a3=zeros(4,4);
    a3(1,1)=T-2;
    for k=2:4;
        a3(1,k)=sum(X(k-1,:));
        a3(k,1)=a3(1,k);
```

```
                for l=2:4;
                    a3(l,k)=sum(X(l-1,:).*X(k-1,:));
                end
            end
            b3(1)=sum(y3);
            for k=2:4;
                b3(k)=sum(X(k-1,:).*y3);
            end
            u3=a3^(-1)*b3';
            %para o modelo 3r:
            a31=zeros(3,3);
            a31(1,1)=T-2;
            for k=2:3;
                a31(1,k)=sum(X(k-1,:));
                a31(k,1)=a31(1,k);
                for l=2:3;
                    a31(l,k)=sum(X(l-1,:).*X(k-1,:));
                end
            end
            b31(1)=sum(y3);
            for k=2:3;
                b31(k)=sum(X(k-1,:).*y3);
            end
            nu3=a31^(-1)*b31';
            for j=2:T-1;
                Prevyu=u3(1)+u3(2)*y(j)+u3(3)*z(j-1)+u3(4)*x(j-1);
                erro3u=erro3u+(Prevyu-y(j+1))^2;
                Prevyr=nu3(1)+nu3(2)*y(j)+nu3(3)*z(j-1);
                erro3r=erro3r+(Prevyr-y(j+1))^2;
            end
            F1=(erro1r-erro1u)/erro1u*(T-1);
            F2=(erro2r-erro2u)/erro2u*(T-1);
            F3=(erro3r-erro3u)/erro3u*(T-2);
...
```

Programa 7 – Calcula, para cada equação a taxa de rejeição, para saber a potência.

```
...
if F1>2.7055 ;
            r1=r1+1;
end
if F2>2.7055 ;
            r2=r2+1;
end
if F3>2.7055 ;
            r3=r3+1;
end
if F1>3.8415 ;
             r11=r11+1;
end
if F2>3.8415 ;
             r22=r22+1;
end
if F3>3.8415 ;
             r33=r33+1;
end
if F1>5.0239;
```

```
            r111=r111+1;
 end
 if F2>5.0239;
            r222=r222+1;
 end
 if F3>5.0239;
            r333=r333+1;
 end
 if F1>6.6349;
            r1111=r1111+1;
 end
 if F2>6.6349;
            r2222=r2222+1;
 end
 if F3>6.6349;
            r3333=r3333+1;
 end
 end
 disp('equaçao 1  10% 5% 2.5%  1%')
 disp([r1,r11,r111,r1111])
 disp('equaçao 2  10% 5% 2.5%  1%')
 disp([r2,r22,r222,r2222])
 disp('equaçao 3  10% 5% 2.5%  1%')
 disp([r3,r33,r333,r3333])
 end
end
...
```

Programa 8 – Gera os dados, estima os parâmetros dos modelos, calcula os erros e obtém uma estimativa da estatística de *Wald*, para o estudo da transitividade da cointegração.

```
clear
p2=[0.1 0.3 0.5 0.7 0.9];
T=input('Nº de observaçoes');
for i=1:5;
    for ii=1:2;
        rr1=0;
        rr2=0;
        rr3=0;
        rr4=0;
        for Numsim=1:10000;
           %gerar os dados
           x(1)=1;
           y(1)=1;
           z(1)=1;
          for a=2:T;
            x(a)=1+p2(i)*x(a-1)+randn;
            z(a)=1+p2(i)*z(a-1)+randn;
            y(a)=1+p2(i)*y(a-1)+randn;
          end
   %para o modelo u1:
   clear X
   X(1,:)=x(1:T-1);
   X(2,:)=z(1:T-1);
   y1=x(2:T);
```

```
a1=zeros(3,3);
a1(1,1)=T-1;
for k=2:3;
   a1(1,k)=sum(X(k-1,:));
   a1(k,1)=a1(1,k);
   for l=2:3;
      a1(l,k)=sum(X(l-1,:).*X(k-1,:));
   end
end
b1(1)=sum(y1);
for k=2:3;
   b1(k)=sum(X(k-1,:).*y1);
end
u1=a1^(-1)*b1';
%para o modelo r1:
r1=polyfit(X(1,:),y1,1);
errou1=0;
error1=0;
for j=2:T;
Prevyu1(j)=u1(1)+u1(2)*x(j-1)+u1(3)*z(j-1);
errou1=errou1+(Prevyu1(j)-x(j))^2;
Prevyr1(j)=r1(2)+r1(1)*x(j-1);
error1=error1+(Prevyr1(j)-x(j))^2;
end
F1=(error1-errou1)/errou1*(T-2);
%para o modelo u2:
clear X
X(1,:)=x(1:T-1);
X(2,:)=z(1:T-1);
y2=z(2:T);
a2=zeros(3,3);
a2(1,1)=T-1;
for k=2:3;
   a2(1,k)=sum(X(k-1,:));
   a2(k,1)=a2(1,k);
   for l=2:3;
     a2(l,k)=sum(X(l-1,:).*X(k-1,:));
   end
end
b2(1)=sum(y2);
for k=2:3;
   b2(k)=sum(X(k-1,:).*y2);
end
u2=a2^(-1)*b2';
%para o modelo r2:
r2=polyfit(X(2,:),y2,1);
errou2=0;
error2=0;
for j=2:T;
   Prevyu2(j)=u2(1)+u2(2)*x(j-1)+u2(3)*z(j-1);
   errou2=errou2+(Prevyu2(j)-z(j))^2;
   Prevyr2(j)=r2(2)+r2(1)*z(j-1);
   error2=error2+(Prevyr2(j)-z(j))^2;
end
F2=(error2-errou2)/errou2*(T-2);
%para o modelo u3:
clear X
X(1,:)=z(1:T-1);
X(2,:)=y(1:T-1);
```

```
y3=z(2:T);
a3=zeros(3,3);
a3(1,1)=T-1;
for k=2:3;
   a3(1,k)=sum(X(k-1,:));
   a3(k,1)=a3(1,k);
   for l=2:3;
      a3(l,k)=sum(X(l-1,:).*X(k-1,:));
   end
end
b3(1)=sum(y3);
for k=2:3;
   b3(k)=sum(X(k-1,:).*y3);
end
u3=a3^(-1)*b3';
%para o modelo r3:
r3=polyfit(X(1,:),y3,1);
errou3=0;
error3=0;
for j=2:T;
   Prevyu3(j)=u3(1)+u3(2)*z(j-1)+u3(3)*y(j-1);
   errou3=errou3+(Prevyu3(j)-z(j))^2;
   Prevyr3(j)=r3(2)+r3(1)*z(j-1);
   error3=error3+(Prevyr3(j)-z(j))^2;
end
F3=(error3-errou3)/errou3*(T-2);
%para o modelo u4:
clear X
X(1,:)=z(1:T-1);
X(2,:)=y(1:T-1);
y4=y(2:T);
a4=zeros(3,3);
a4(1,1)=T-1;
for k=2:3;
   a4(1,k)=sum(X(k-1,:));
   a4(k,1)=a4(1,k);
   for l=2:3;
      a4(l,k)=sum(X(l-1,:).*X(k-1,:));
   end
end
b4(1)=sum(y4);
for k=2:3;
   b4(k)=sum(X(k-1,:).*y4);
end
u4=a4^(-1)*b4';
%para o modelo r4:
r4=polyfit(X(2,:),y4,1);
errou4=0;
error4=0;
for j=2:T;
   Prevyu4(j)=u4(1)+u4(2)*z(j-1)+u4(3)*y(j-1);
   errou4=errou4+(Prevyu4(j)-y(j))^2;
   Prevyr4(j)=r4(2)+r4(1)*y(j-1);
   error4=error4+(Prevyr4(j)-y(j))^2;
end
F4=(error4-errou4)/errou4*(T-2);
%para o modelo u5:
clear X
X(1,:)=x(2:T-1);
```

```
    X(2,:)=y(1:T-2);  %mudar: y(2:T-1)<->y(1:T-2)
    X(3,:)=z(1:T-2);
    y5=x(3:T);
    a5=zeros(4,4);
    a5(1,1)=T-2;
    for k=2:4;
       a5(1,k)=sum(X(k-1,:));
       a5(k,1)=a5(1,k);
       for l=2:4;
          a5(l,k)=sum(X(l-1,:).*X(k-1,:));
       end
    end
    b5(1)=sum(y5);
    for k=2:4;
       b5(k)=sum(X(k-1,:).*y5);
    end
    u5=a5^(-1)*b5';
    %para o modelo r5:
    clear X
    X(1,:)=x(2:T-1);
    X(2,:)=z(1:T-2);
    yr5=x(3:T);
    ar5=zeros(3,3);
    ar5(1,1)=T-2;
    for k=2:3;
       ar5(1,k)=sum(X(k-1,:));
       ar5(k,1)=ar5(1,k);
       for l=2:3;
          ar5(l,k)=sum(X(l-1,:).*X(k-1,:));
       end
    end
    br5(1)=sum(yr5);
    for k=2:3;
       br5(k)=sum(X(k-1,:).*yr5);
    end
    r5=ar5^(-1)*br5';
    errou5=0;
    error5=0;
    for j=3:T;
       Prevyu5(j)=u5(1)+u5(2)*x(j-1)+u5(3)*y(j-2)+u5(4)*z(j-2); %mudar
y(j-1)<->y (j-2)
       errou5=errou5+(Prevyu5(j)-x(j))^2;
       Prevyr5(j)=r5(1)+r5(2)*x(j-1)+r5(3)*z(j-2);
       error5=error5+(Prevyr5(j)-x(j))^2;
    end
    F5=(error5-errou5)/errou5*(T-3);
    %para o modelo u6:
    clear X
    X(1,:)=x(1:T-2); %mudar x(2:T-1)<->x(1:T-2)
    X(2,:)=y(2:T-1);
    X(3,:)=z(1:T-2);
    y6=y(3:T);
    a6=zeros(4,4);
    a6(1,1)=T-2;
    for k=2:4;
       a6(1,k)=sum(X(k-1,:));
       a6(k,1)=a6(1,k);
       for l=2:4;
          a6(l,k)=sum(X(l-1,:).*X(k-1,:));
```

```
      end
   end
   b6(1)=sum(y6);
   for k=2:4;
      b6(k)=sum(X(k-1,:).*y6);
   end
   u6=a6^(-1)*b6';
   %para o modelo r6:
   clear X
   X(1,:)=y(2:T-1);
   X(2,:)=z(1:T-2);
   yr6=y(3:T);
   ar6=zeros(3,3);
   ar6(1,1)=T-2;
   for k=2:3;
      ar6(1,k)=sum(X(k-1,:));
      ar6(k,1)=ar6(1,k);
      for l=2:3;
         ar6(l,k)=sum(X(l-1,:).*X(k-1,:));
      end
   end
   br6(1)=sum(yr6);
   for k=2:3;
      br6(k)=sum(X(k-1,:).*yr6);
   end
   r6=ar6^(-1)*br6';
   errou6=0;
   error6=0;
   for j=3:T;
      Prevyu6(j)=u6(1)+u6(2)*x(j-2)+u6(3)*y(j-1)+u6(4)*z(j-2); %mudar
x(j-1)<->x(j-2)
      errou6=errou6+(Prevyu6(j)-y(j))^2;
      Prevyr6(j)=r6(1)+r6(2)*y(j-1)+r6(3)*z(j-2);
      error6=error6+(Prevyr6(j)-y(j))^2;
   end
   F6=(error6-errou6)/errou6*(T-3);
...
```

Programa 9 – Calcula o nº de casos em que não há transitividade na cointegração.

```
...
if ((F1>1.01 & F2>1.01) & (F3>1.01 &F4>1.01) &(F5<=1.01 |
F6<=1.01))|((F1>1.01 & F2>1.01) & (F3<=1.01 | F4<=1.01)&(F5>1.01
&F6>1.01))|((F1<=1.01 | F2<=1.01)& (F3>1.01 & F4>1.01) & (F5>1.01
&F6>1.01))|((F1<=1.01 |F2<=1.01)&(F3<=1.01 |F4<=1.01)& (F5>1.01
&F6>1.01));
       rr1=rr1+1;
   end
   if (F1>1.49 & F2>1.49 & F3>1.49 &F4>1.49 &(F5<=1.49 |
F6<=1.49))|(F1>1.49 & F2>1.49 & (F3<=1.49 | F4<=1.49)&F5>1.49
&F6>1.49)|((F1<=1.49 | F2<=1.49)& F3>1.49 & F4>1.49 & F5>1.49
&F6>1.49)|((F1<=1.49 |F2<=1.49)&(F3<=1.49 |F4<=1.49)& F5>1.49
&F6>1.49);
       rr2=rr2+1;
   end
   if (F1>2 & F2>2 & F3>2 &F4>2 &(F5<=2 | F6<=2))|(F1>2 & F2>2 &
(F3<=2 | F4<=2)&F5>2 &F6>2)|((F1<=2 | F2<=2)& F3>2 & F4>2 & F5>2
```

```
&F6>2)|((F1<=2 |F2<=2)&(F3<=2 |F4<=2)& F5>2 &F6>2);
        rr3=rr3+1;
    end
    if (F1>2.72 & F2>2.72 & F3>2.72 &F4>2.72 &(F5<=2.72 |
F6<=2.72))|(F1>2.72 & F2>2.72 & (F3<=2.72 | F4<=2.72)&F5>2.72
&F6>2.72)|((F1<=2.72 | F2<=2.72)& F3>2.72 & F4>2.72 & F5>2.72
&F6>2.72)|((F1<=2.72 |F2<=2.72)&(F3<=2.72 |F4<=2.72)& F5>2.72
&F6>2.72);
        rr4=rr4+1;
    end
end
disp('10%, 5%, 2,5%, 1%')
disp([rr1,rr2,rr3,rr4])
end
end
```

Programa 10 – Calcula o nº de casos significativos, para a transitividade da cointegração.

```
...
if F1>1.01 & F2>1.01 & F3>1.01 &F4>1.01 &F5>1.01 &F6>1.01;
    rr1=rr1+1;
end
if F1>1.49 & F2>1.49 & F3>1.49 &F4>1.49 &F5>1.49 &F6>1.49;
    rr2=rr2+1;
end
if F1>2 & F2>2 & F3>2 &F4>2 &F5>2 &F6>2;
    rr3=rr3+1;
end
if F1>2.72 & F2>2.72 & F3>2.72 &F4>2.72 &F5>2.72 &F6>2.72;
    rr4=rr4+1;
end

disp('10%, 5%, 2,5%, 1%')
disp([rr1,rr2,rr3,rr4])
end
end
```

Programa 11 – Gera os dados, estima os parâmetros dos modelos, calcula os erros e obtém uma estimativa da estatística de *Wald*, para o estudo da potência dos testes efectuados acerca da transitividade da cointegração.

```
clear
p2=[0.1 0.3 0.5 0.7 0.9];
T=input('Nº de observaçoes');
for i=1:5;
    for ii=1:2;
       for Numsim=1:10000;
          %gerar os dados
          x(1)=1;
          y(1)=1;
          z(1)=1;
          for a=2:T;
            x(a)=1+p2(i)*x(a-1)+randn;
```

```
            z(a)=1+p2(i)*z(a-1)+randn;
            y(a)=1+p2(i)*y(a-1)+randn;
         end
         %para o modelo u1:
         clear X
         X(1,:)=x(1:T-1);
         X(2,:)=z(1:T-1);
         y1=x(2:T);
         a1=zeros(3,3);
         a1(1,1)=T-1;
         for k=2:3;
           a1(1,k)=sum(X(k-1,:));
           a1(k,1)=a1(1,k);
           for l=2:3;
              a1(l,k)=sum(X(l-1,:).*X(k-1,:));
           end
        end
        b1(1)=sum(y1);
        for k=2:3;
          b1(k)=sum(X(k-1,:).*y1);
        end
        u1=a1^(-1)*b1';
        %para o modelo r1:
        y11=y1-0.1*X(2,:);
        r1=polyfit(X(1,:),y11,1);
        errou1=0;
        error1=0;
        for j=2:T;
          Prevyu1(j)=u1(1)+u1(2)*x(j-1)+u1(3)*z(j-1);
          errou1=errou1+(Prevyu1(j)-x(j))^2;
          Prevyr1(j)=r1(2)+r1(1)*x(j-1)+0.1*z(j-1);
          error1=error1+(Prevyr1(j)-x(j))^2;
       end
       F1(Numsim)=(error1-errou1)/errou1*(T-2);
      %para o modelo u2:
      clear X
      X(1,:)=x(1:T-1);
      X(2,:)=z(1:T-1);
      y2=z(2:T);
      a2=zeros(3,3);
      a2(1,1)=T-1;
      for k=2:3;
         a2(1,k)=sum(X(k-1,:));
         a2(k,1)=a2(1,k);
         for l=2:3;
            a2(l,k)=sum(X(l-1,:).*X(k-1,:));
         end
      end
      b2(1)=sum(y2);
      for k=2:3;
         b2(k)=sum(X(k-1,:).*y2);
      end
      u2=a2^(-1)*b2';
      %para o modelo r2:
      y22=y2-0.1*X(1,:);
      r2=polyfit(X(2,:),y22,1);
       errou2=0;
      error2=0;
      for j=2:T;
```

```
        Prevyu2(j)=u2(1)+u2(2)*x(j-1)+u2(3)*z(j-1);
        errou2=errou2+(Prevyu2(j)-z(j))^2;
        Prevyr2(j)=r2(2)+r2(1)*z(j-1)+0.1*x(j-1);
        error2=error2+(Prevyr2(j)-z(j))^2;
    end
    F2(Numsim)=(error2-errou2)/errou2*(T-2);
    %para o modelo u3:
    clear X
     X(1,:)=z(1:T-1);
     X(2,:)=y(1:T-1);
     y3=z(2:T);
     a3=zeros(3,3);
     a3(1,1)=T-1;
     for k=2:3;
        a3(1,k)=sum(X(k-1,:));
        a3(k,1)=a3(1,k);
        for l=2:3;
           a3(l,k)=sum(X(l-1,:).*X(k-1,:));
        end
     end
     b3(1)=sum(y3);
     for k=2:3;
        b3(k)=sum(X(k-1,:).*y3);
     end
     u3=a3^(-1)*b3';
     %para o modelo r3:
     y33=y3-0.1*X(2,:);
     r3=polyfit(X(1,:),y33,1);
     errou3=0;
     error3=0;
     for j=2:T;
        Prevyu3(j)=u3(1)+u3(2)*z(j-1)+u3(3)*y(j-1);
        errou3=errou3+(Prevyu3(j)-z(j))^2;
        Prevyr3(j)=r3(2)+r3(1)*z(j-1)+0.1*y(j-1);
        error3=error3+(Prevyr3(j)-z(j))^2;
     end
     F3(Numsim)=(error3-errou3)/errou3*(T-2);
    %para o modelo u4:
    clear X
     X(1,:)=z(1:T-1);
     X(2,:)=y(1:T-1);
     y4=y(2:T);
     a4=zeros(3,3);
     a4(1,1)=T-1;
     for k=2:3;
        a4(1,k)=sum(X(k-1,:));
        a4(k,1)=a4(1,k);
        for l=2:3;
           a4(l,k)=sum(X(l-1,:).*X(k-1,:));
        end
     end
     b4(1)=sum(y4);
     for k=2:3;
        b4(k)=sum(X(k-1,:).*y4);
     end
     u4=a4^(-1)*b4';
     %para o modelo r4:
     y44=y4-0.1*X(1,:);
     r4=polyfit(X(2,:),y44,1);
```

```
      errou4=0;
      error4=0;
      for j=2:T;
         Prevyu4(j)=u4(1)+u4(2)*z(j-1)+u4(3)*y(j-1);
         errou4=errou4+(Prevyu4(j)-y(j))^2;
         Prevyr4(j)=r4(2)+r4(1)*y(j-1)+0.1*z(j-1);
         error4=error4+(Prevyr4(j)-y(j))^2;
      end
      F4(Numsim)=(error4-errou4)/errou4*(T-2);
      %para o modelo u5:
      clear X
      X(1,:)=x(2:T-1);
      X(2,:)=y(1:T-2);%mudar y(2:T-1)<->y(1:T-2)
      X(3,:)=z(1:T-2);
      y5=x(3:T);
      a5=zeros(4,4);
      a5(1,1)=T-2;
      for k=2:4;
         a5(1,k)=sum(X(k-1,:));
         a5(k,1)=a5(1,k);
         for l=2:4;
            a5(l,k)=sum(X(l-1,:).*X(k-1,:));
         end
      end
      b5(1)=sum(y5);
      for k=2:4;
         b5(k)=sum(X(k-1,:).*y5);
      end
      u5=a5^(-1)*b5';
      %para o modelo r5:
      clear X
      X(1,:)=x(2:T-1);
      X(2,:)=z(1:T-2);
      yr5=x(3:T);
      ar5=zeros(3,3);
      ar5(1,1)=T-2;
      for k=2:3;
         ar5(1,k)=sum(X(k-1,:));
         ar5(k,1)=ar5(1,k);
         for l=2:3;
            ar5(l,k)=sum(X(l-1,:).*X(k-1,:));
         end
      end
      br5(1)=sum(yr5);
      for k=2:3;
         br5(k)=sum(X(k-1,:).*yr5);
      end
      r5=ar5^(-1)*br5';
       errou5=0;
      error5=0;
      for j=3:T;
         Prevyu5(j)=u5(1)+u5(2)*x(j-1)+u5(3)*y(j-2)+u5(4)*z(j-2);%
mudar y(j-1)<-> y(j-2)
         errou5=errou5+(Prevyu5(j)-x(j))^2;
         Prevyr5(j)=r5(1)+r5(2)*x(j-1)+r5(3)*z(j-2);
         error5=error5+(Prevyr5(j)-x(j))^2;
      end
      F5(Numsim)=(error5-errou5)/errou5*(T-3);
      %para o modelo u6:
```

```
      clear X
      X(1,:)=x(1:T-2);%mudar x(2:T-1)<-> x(1:T-2)
      X(2,:)=y(2:T-1);
      X(3,:)=z(1:T-2);
      y6=y(3:T);
      a6=zeros(4,4);
      a6(1,1)=T-2;
      for k=2:4;
         a6(1,k)=sum(X(k-1,:));
         a6(k,1)=a6(1,k);
         for l=2:4;
            a6(l,k)=sum(X(l-1,:).*X(k-1,:));
         end
      end
      b6(1)=sum(y6);
      for k=2:4;
         b6(k)=sum(X(k-1,:).*y6);
      end
      u6=a6^(-1)*b6';
      %para o modelo r6:
      clear X
      X(1,:)=y(2:T-1);
      X(2,:)=z(1:T-2);
      yr6=y(3:T);
      ar6=zeros(3,3);
      ar6(1,1)=T-2;
      for k=2:3;
         ar6(1,k)=sum(X(k-1,:));
         ar6(k,1)=ar6(1,k);
         for l=2:3;
            ar6(l,k)=sum(X(l-1,:).*X(k-1,:));
         end
      end
      br6(1)=sum(yr6);
      for k=2:3;
         br6(k)=sum(X(k-1,:).*yr6);
      end
      r6=ar6^(-1)*br6';
       errou6=0;
      error6=0;
      for j=3:T;
         Prevyu6(j)=u6(1)+u6(2)*x(j-2)+u6(3)*y(j-1)+u6(4)*z(j-
2);%mudar x(j-1)<->x(j-2)
         errou6=errou6+(Prevyu6(j)-y(j))^2;
         Prevyr6(j)=r6(1)+r6(2)*y(j-1)+r6(3)*z(j-2);
         error6=error6+(Prevyr6(j)-y(j))^2;
      end
      F6(Numsim)=(error6-errou6)/errou6*(T-3);
       End
...
```

Programa 12 – Calcula, para cada relação a taxa de rejeição, para saber a potência.

```
...
vc=[2.7055 3.8415 5.0239 6.6349];
  for i2=1:4;
     rr1=zeros(10000,3);
```

```
      for Numsim=1:10000
         k=1;
         if F1(Numsim)& F2(Numsim)>vc(i2) ;
             rr1(Numsim,k)=1;
         end
         if F3(Numsim)& F4(Numsim)>vc(i2) ;
             rr1(Numsim,k+1)=1;
         end
         if F5(Numsim)& F6(Numsim)>vc(i2) ;
             rr1(Numsim,k+2)=1;
         end
      end
      for k=1:3;
         resul(i2,k)=sum(rr1(:,k));
      end
   end
   disp('equaçao 1 e 2 - 10% 5% 2.5%  1%')
   disp(resul(:,1)')
   disp('equaçao 3 e 4 - 10% 5% 2.5%  1%')
   disp(resul(:,2)')
   disp('equaçao 5 e 6 - 10% 5% 2.5%  1%')
   disp(resul(:,3)')
   end
end
```

Programa 13 – Programa que calcula percentagens dos casos e média dos erros dos cálculos.

```
clear
x=input('vector x');
y=input('vector y');
for i=1:4;
    valor(i)=(x(i)+y(i))/20000*100;
    media(i)=(x(i)+y(i))/2;
    erro(i)=sqrt((x(i)-media(i))^2+(y(i)-
media(i))^2)/(sqrt(2)*10000)*100;
end
disp(valor)
disp(erro)
```

Programa 14 – Programa que calcula os níveis de significância "experimental" para o estudo da transitividade da cointegração.

```
clear
p2=input('introduza p2');
T=input('Introduza nº de observações');
for ii=1:2
rejeita1=0;
rejeita2emeio=0;
rejeita5=0;
rejeita10=0;
for Numsim=1:10000;
   %gerar os dados
   x(1)=1;
```

```
   y(1)=1;
   z(1)=1;
   for a=2:T;
      x(a)=1+p2*x(a-1)+randn;
      z(a)=1+p2*z(a-1)+randn;
      y(a)=1+p2*y(a-1)+randn;
   end
   %para o modelo nao restrito (URSS)1:
   clear X
      X(1,:)=x(1:T-1);
      X(2,:)=y(1:T-1);
      y1=x(2:T);
      a1=zeros(3,3);
      a1(1,1)=T-1;
      for k=2:3;
         a1(1,k)=sum(X(k-1,:));
         a1(k,1)=a1(1,k);
         for l=2:3;
            a1(l,k)=sum(X(l-1,:).*X(k-1,:));
         end
      end
      b1(1)=sum(y1);
      for k=2:3;
         b1(k)=sum(X(k-1,:).*y1);
      end
      u1x=a1^(-1)*b1';
      %para o modelo restrito (RRSS)1:
      nu1x=polyfit(X(1,:),y1,1);
       erro1x=0;
      erro2x=0;
      for j=2:T;
         Prevyux(j)=u1x(1)+u1x(2)*x(j-1)+u1x(3)*y(j-1);
         erro1x=erro1x+(Prevyux(j)-x(j))^2;      % URSS soma: (x-
previsao x)^2
         Prevyrx(j)=nu1x(2)+nu1x(1)*x(j-1);
         erro2x=erro2x+(Prevyrx(j)-x(j))^2; %RRSS soma:(x-previsao
x)^2
    end

 %para o modelo nao restrito URSS 2:
   clear X
      X(1,:)=y(1:T-1);
      X(2,:)=x(1:T-1);
      y2=y(2:T);
      a2=zeros(3,3);
      a2(1,1)=T-1;
      for k=2:3;
         a2(1,k)=sum(X(k-1,:));
         a2(k,1)=a2(1,k);
         for l=2:3;
            a2(l,k)=sum(X(l-1,:).*X(k-1,:));
         end
      end
      b2(1)=sum(y2);
      for k=2:3;
         b2(k)=sum(X(k-1,:).*y2);
      end
      u2y=a2^(-1)*b2';
      %para o modelo r2:
```

```
      nu2y=polyfit(X(1,:),y2,1);
       erro1y=0;
      erro2y=0;
      for j=2:T;
         Prevyuy(j)=u2y(1)+u2y(2)*y(j-1)+u2y(3)*x(j-1);
         erro1y=erro1y+(Prevyuy(j)-y(j))^2;  % URSS soma: (y-previsao
y)^2
         Prevyry(j)=nu2y(2)+nu2y(1)*y(j-1);
         erro2y=erro2y+(Prevyry(j)-y(j))^2; % RRSS soma: (y-previsao
y)^2
    end
    ErroU=erro1x+erro1y;
    ErroR=erro2x+erro2y;
    F(Numsim)=(ErroR-ErroU)*(T-2)/(2*ErroU);
    if T==250
        vc=[4.695 3.745 3.035 2.325];
    else
        vc=[4.65 3.72 3.01 2.31];
    end

 if F(Numsim)> vc(1)   %distribuiçao F (2,T-2) 1%
     rejeita1=rejeita1+1;
 end
 if F(Numsim)> vc(2)   %distribuiçao F (2,T-2) 2,5%
     rejeita2emeio=rejeita2emeio+1;
 end
 if F(Numsim)> vc(3)   %distribuiçao F (2,T-2) 5%
     rejeita5=rejeita5+1;
 end
 if F(Numsim)> vc(4)   %distribuiçao F (2,T-2) 10%
     rejeita10=rejeita10+1;
 end

end
disp('1% 2.5% 5% 10%')
disp(['Rejeita H0 ',num2str(rejeita1),'  ',num2str(rejeita2emeio),'
',num2str(rejeita5),'  ',num2str(rejeita10)])
end
```

Anexo 2 - Valores críticos "fora da amostra"

		p = 10%			p = 30%			p = 50%		
π_2	$\alpha \setminus T$	250	500	1000	250	500	1000	250	500	1000
0,1	1%	2,100	1,987	2,017	3,784	3,612	3,640	4,982	5,170	4,986
	5%	1,092	1,091	1,124	2,131	2,111	2,080	2,981	3,034	2,932
	10%	0,717	0,745	0,754	1,514	1,524	1,467	2,153	2,202	2,164
0,3	1%	1,997	2,030	2,082	3,680	3,486	3,697	5,132	5,001	4,695
	5%	1,106	1,129	1,114	2,144	2,065	2,139	3,136	2,982	2,884
	10%	0,728	0,762	0,778	1,497	1,488	1,506	2,252	2,184	2,115
0,5	1%	1,963	2,014	1,962	3,857	3,784	3,547	5,027	5,179	5,025
	5%	1,064	1,148	1,094	2,140	2,187	2,035	2,992	3,001	3,032
	10%	0,715	0,777	0,734	1,450	1,528	1,464	2,181	2,157	2,216
0,7	1%	2,073	1,894	2,056	4,006	3,537	3,783	5,501	5,076	5,135
	5%	1,129	1,094	1,121	2,204	2,110	2,161	3,208	2,995	2,991
	10%	0,740	0,735	0,731	1,531	1,483	1,494	2,267	2,147	2,198
0,9	1%	2,079	2,155	2,003	3,847	3,744	3,830	5,248	5,229	5,112
	5%	1,108	1,154	1,097	2,179	2,111	2,076	3,166	3,069	2,993
	10%	0,739	0,780	0,753	1,527	1,499	1,467	2,240	2,256	2,230

Tabela 17 – Evolução dos valores críticos "fora da amostra"