Previsão da variação do preço da soja, utilizando Cadeia de Markov

Rafaela Boeira Cechin (Universidade de Caxias do Sul) rafaelacechin@hotmail.com.br
Leandro Luís Corso (Universidade de Caxias do Sul) leandro.lcorso@gmail.com

Resumo:
Considerando a importância da exportação para o mercado brasileiro, e principalmente neste caso específico, a exportação de soja, que colocou o país em uma posição de destaque por ser um dos maiores exportadores deste grão do mundo, informações como variação do preço podem ser úteis para gestores de agronegócio. Este artigo estuda a variação do preço da *commodity* soja, obtida com o uso das Cadeias de Markov. Para isso, os dados foram coletados no Centro de Estudos Avançados em Economia Aplicada, departamento que faz parte da Universidade de São Paulo, e são referentes aos dias úteis de janeiro a agosto de 2017. Foi observado que a maior probabilidade é de variação diária de 0 a 0,99% do preço, e que seu tempo de recorrência esperado é de 2,5 dias.
Palavras-chave: Cadeia de Markov, Variação de preço, Commodity soja.

Forecast the variation of the soybean's price, using Markov Chain

Abstract:
Considering the importance of exporting to the Brazilian market, and especially in this specific case, the export of soybeans, which put the country in a prominent position because is one of the largest exporters of this grain in the world, information as the price's variation can be useful for agribusiness managers. This article studies a variation of the price of the soybean commodity, obtained with the use of Markov Chains. For this, data were collected at the Center for Advanced Studies in Applied Economics, a department part of the University of São Paulo, and refer to the working days from January to August 2017. It was observed that the highest probability is daily variation is 0 to 0.99% of the price, and that its expected recurrence time is 2.5 days.
Key-words: Markov Chain, Price's variation, Soybean commodity.

1. Introdução

No atual momento econômico internacional, o Brasil é um dos maiores importadores de soja do mundo, assim como outras *commodities*. Isso deve-se ao fato da disponibilidade de recursos híbridos e terras cultiváveis, propiciando o país a ter uma posição de destaque. De acordo com o endereço eletrônico do Ministério da Agricultura, foi vendido o equivalente a mais de dois bilhões de dólares de soja em grão no mês de agosto de 2017, representando uma quantidade de quase seis milhões de toneladas. Ainda, há um aumento na quantidade de soja em grão exportada, quando comparado com os respectivos meses do ano de 2016.

A soja é um grão presente na alimentação humana e de animais, sendo fonte de vitaminas e proteínas. Além desta importância para o mercado brasileiro, ela ainda traz diversas vantagens, como a prevenção diversos cânceres, ainda auxiliando no tratamento de diabetes e obesidade, e na diminuição do colesterol.

Para analisar a variação do preço da *commodity* soja, foi explorado o conceito de Cadeias de Markov, que é estudado na Pesquisa Operacional, matéria presente nos cursos de Engenharia de Produção, e é uma análise estatística de um processo, que o estado futuro depende unicamente do seu estado atual. O objetivo deste trabalho é calcular a probabilidade de variação do preço da soja, com as Cadeias de Markov, além do tempo de recorrência esperado, a fim de ser uma informação útil para gestores de agronegócio.

2. Referencial bibliográfico

A metodologia de Cadeias de Markov (MC, do inglês *Markov Chains*) pode ser aplicada em diversos cenários, como por exemplo Andersen, Nilsen e Reinhardt (2017), que utilizaram este conceito em um problema de superlotação de enfermarias de hospitais. No trabalho dos autores, chegou-se à conclusão que o número de pacientes rejeitados na primeira chegada ao hospital pôde ser diminuído em 11,8%, com uma redistribuição de leitos já disponíveis no hospital. Já Staudt, Coelho e Gonçalves (2011) estudaram MC como ferramenta para obtenção do fator de capacidade de uma empresa, concluindo que alguns setores necessitam análise sobre investimentos em capacitação, já que ultrapassaram 90% da produção.

Outro exemplo é de Jantsch (2017), que utilizou Cadeias de Markov para simulação das condições dos clientes, a partir de uma análise de previsão da inadimplência e mensuração do risco de crédito no uso de cartão de crédito; tendo como resultado uma caracterização satisfatória do perfil dos indivíduos com maior risco de crédito. Ainda, Siltala e Granvik (2017) estudam a aplicação de Cadeias de Markov para estimar a massa de um asteroide, e concluíram que o algoritmo de MC que utilizaram, forneceu estimativas de incertezas mais realistas, comparando com os outros cálculos realizados.

De acordo com Taha (2007), a família de variáveis aleatórias $\{X_{t_n}\} = \{x_1, x_2, \dots, x_n\}$, que descreve o estado do sistema em pontos discretos no tempo t, forma um processo estatístico. Siltala e Granvik (2017) ainda comentam que tal processo é uma Cadeia de Markov se a probabilidade de ocorrência de um estado futuro depende apenas do estado presente, ou seja, se é independente dos eventos passados. Sendo assim, Taha (2007) e Andersen, Nilsen e Reinhardt (2017) apresentam a probabilidade condicional P na Equações 1 e 2.

$$p_{ij} = P\{X_{t+1} = j | X_t = i\} \tag{1}$$

$$p_{ij}^n = P\{X_{t+n} = j | X_t = i\} \tag{2}$$

Onde, X_t representa o estado do processo no tempo t, e p_{ij}^n como sendo a probabilidade de que um processo passe do estado i ao estado j em n passos no tempo t, conforme apresentado por Jantsch (2017). Por ser probabilidade condicional, Hillier e Lierberman (2005) afirmam que estes valores não podem ser negativos, e seu somatório deve ser igual a 1. Staudt, Coelho e Gonçalves (2011) e Andersen, Nilsen e Reinhardt (2017) comentam que uma maneira conveniente de apresentar estas transições é com o uso da Matriz de Transição, exibida na Equação 3.

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1m} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nm} \end{bmatrix} \tag{3}$$

Os autores demonstram que a transição dos estados ocorre do índice da linha para o da coluna, ou seja, a probabilidade p_{nm} corresponde a transição do estado n para o estado m. Taha (2007) comenta também que o somatório de cada linha da matriz deve ser igual a 1, além de que uma matriz é classificada como ergódica quando é possível ir de qualquer estado a outro qualquer em n passos de tempo.

Hillier e Lierberman (2005) então apresentam a probabilidade de estado estável π_j para uma cadeia ergódica de Markov na Equação 4 e sua propriedade na Equação 5.

$$\pi_j = \sum_{i=0}^{M} \pi_i \, p_{ij} \text{ , para } j = 0, 1, \dots, M \tag{4}$$

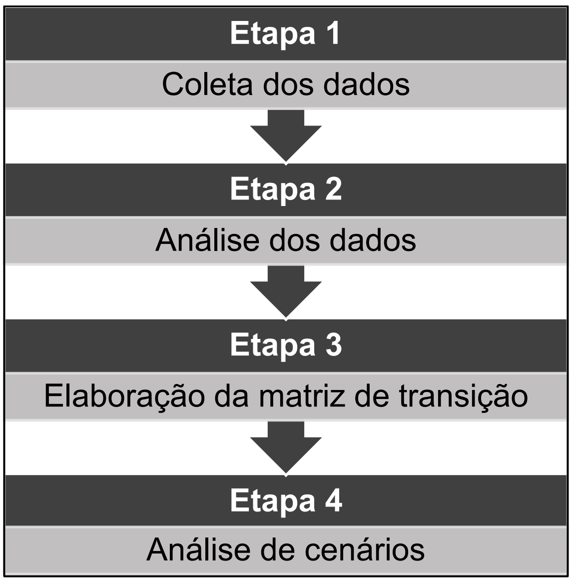
$$\sum_{j=0}^{M} \pi_j = 1 \tag{5}$$

Os autores explicam que, depois de um grande número de transições, a probabilidade de encontrar o processo em um determinado estado, por exemplo j, tende ao valor π_j, independente da distribuição de probabilidade do estado inicial. Ainda é possível analisar o tempo de recorrência esperado, representado por μ_{ii}, que é o número esperado de transições até que o processo retorne ao estado inicial i, apresentado na Equação 6.

$$\mu_{ii} = \frac{1}{\pi_i}, \text{ para } i = 0, 1, \dots, M \tag{6}$$

3. Aplicação da metodologia e resultados

Com o intuito de atingir o objetivo proposto para este estudo, a metodologia elaborada fio dividida em quatro etapas, conforme ilustradas na Figura 1.



Fonte: Elaborado pelos autores (2017)

Figura 1 – Metodologia

3.1 Coleta dos dados

Para este estudo, foram coletados os valores de preço da soja brasileira em grão a granel tipo exportação, de acordo com o endereço eletrônico do Centro de Estudos Avançados em Economia Aplicada (Cepea), que faz parte do Departamento de Economia, Administração e Sociologia da Universidade de São Paulo (USP). O intervalo de tempo dos dados é dia-a-dia comercial em que há negociação de derivativos na Bolsa de Valores de São Paulo (BM&FBovespa), e são referentes a 2 de janeiro de 2017 a 31 de agosto de 2017. A unidade de medida é real por saca de 60kg e os valores são da soja comercializada no porto de Paranaguá, em Paraná, e referentes a cotação do contrato futuro de soja na BM&FBovespa. A Figura 2 apresenta estes dados.

Fonte: Elaborado pelos autores (2017)

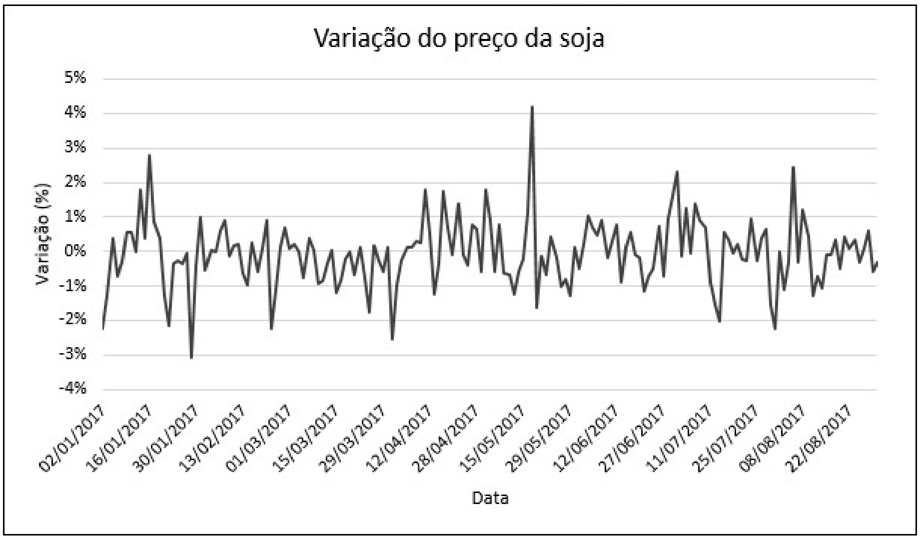
Figura 2 – Preço diário da soja

3.2 Análise dos dados

Com estes dados disponíveis, foi então calculado a sua variação percentual entre os dias, com a fórmula apresentada na Equação 7.

$$variação\ \% = \frac{P_{t+1} - P_t}{P_t} 100 \quad (7)$$

Onde, P_t é o preço no tempo t e P_{t+1} é o preço no tempo $t + 1$. Exemplificando, conforme a Figura 2, o preço da soja no dia 2 de janeiro de 2017 foi de R$77,22 e no dia 3 do mesmo mês, foi R$75,48, então a variação é de -2,25%, com o sinal de negativo representando uma queda no preço. Os demais valores são apresentados no gráfico da Figura 3.



Fonte: Elaborado pelos autores (2017)

Figura 3 – Variação do preço

3.3 Elaboração da matriz de transição

Para poder calcular a matriz de transição, foi necessário definir intervalos para a variação percentual, conforme exibido na Tabela 1.

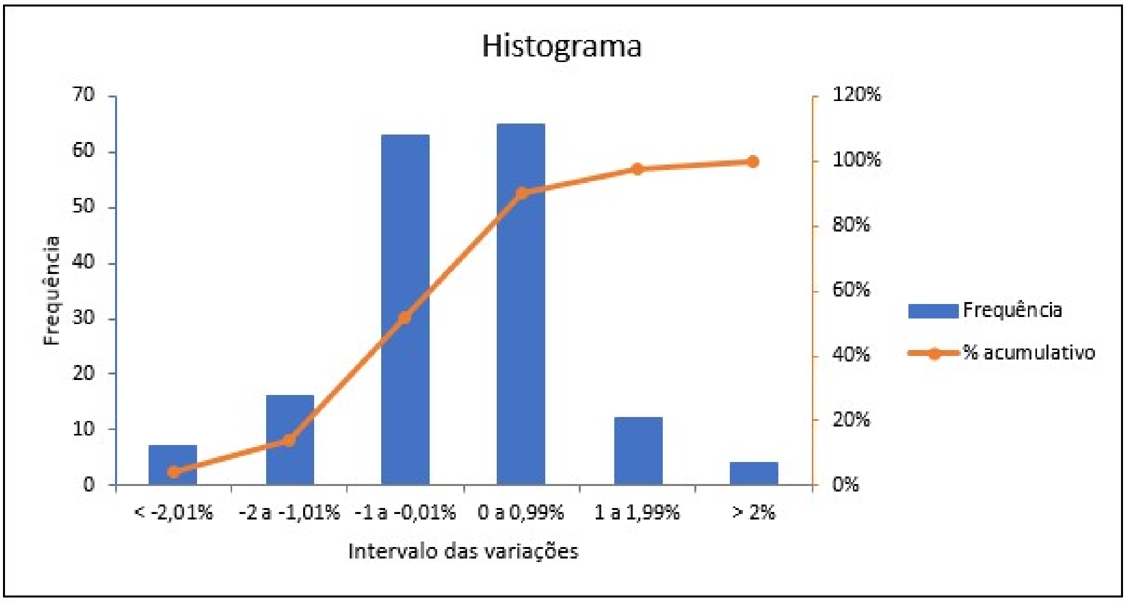
Intervalo das variações
< -2,01%
de -2 a -1,01%
de -1 a -0,01%
de 0 a 0,99%
de 1 a 1,99%
> 2%

Fonte: Elaborado pelos autores (2017)

Tabela 1 – Intervalo das variações

Então, foi possível analisar a frequência e o percentual acumulativo que ocorriam as variações em cada um destes intervalos, apresentados na Figura 4.

Observando o gráfico, percebe-se que a maior frequência de variação ocorreu no intervalo de 0 a 0,99%. Assim, utilizando estas informações, foi possível calcular a matriz com as probabilidades de transição de Markov, analisando quantas vezes a variação do preço saiu do estado i para o estado j. A matriz é apresentada na Figura 5.



Fonte: Elaborado pelos autores (2017)

Figura 4 – Histograma dos intervalos da variação do preço da soja

	de \ para	< -2,01%	-2 a -1,01%	-1 a -0,01%	0 a 0,99%	1 a 1,99%	> 2%
P =	< -2,01%	0,000	0,286	0,429	0,286	0,000	0,000
	-2 a -1,01%	0,188	0,000	0,563	0,250	0,000	0,000
	-1 a -0,01%	0,016	0,113	0,306	0,419	0,129	0,016
	0 a 0,99%	0,031	0,092	0,415	0,385	0,062	0,015
	1 a 1,99%	0,000	0,000	0,250	0,583	0,000	0,167
	> 2%	0,000	0,250	0,500	0,250	0,000	0,000

Fonte: Elaborado pelos autores (2017)

Figura 5 – Matriz de transição

3.4 Análise de cenários

A análise de cenários necessita do cálculo do estado estável, conforme explicado na Equação 4. Então, as Equações 8 e 9 apresentam o estado estável para a variação do preço da soja.

$$(\pi_0\ \pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5) = (\pi_0\ \pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5)\begin{bmatrix} 0 & 0{,}286 & 0{,}429 & 0{,}286 & 0 & 0 \\ 0{,}188 & 0 & 0{,}563 & 0{,}25 & 0 & 0 \\ 0{,}016 & 0{,}113 & 0{,}306 & 0{,}419 & 0{,}129 & 0{,}016 \\ 0{,}031 & 0{,}092 & 0{,}415 & 0{,}385 & 0{,}062 & 0{,}015 \\ 0 & 0 & 0{,}25 & 0{,}583 & 0 & 0{,}167 \\ 0 & 0{,}25 & 0{,}5 & 0{,}25 & 0 & 0 \end{bmatrix} \quad (8)$$

$$(\pi_0\ \pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5\ \pi_6) = 1 \quad (9)$$

Resultando no seguinte sistema de equações, exibido pelas Equações 10 a 16.

$$\pi_0 = 0\pi_0 + 0{,}188\pi_1 + 0{,}016\pi_2 + 0{,}031\pi_3 + 0\pi_4 + 0\pi_5 \quad (10)$$

$$\pi_1 = 0{,}286\pi_0 + 0\pi_1 + 0{,}113\pi_2 + 0{,}092\pi_3 + 0\pi_4 + 0{,}25\pi_5 \quad (11)$$

$$\pi_2 = 0{,}429\pi_0 + 0{,}563\pi_1 + 0{,}306\pi_2 + 0{,}415\pi_3 + 0{,}25\pi_4 + 0{,}5\pi_5 \quad (12)$$

$$\pi_3 = 0{,}286\pi_0 + 0{,}25\pi_1 + 0{,}419\pi_2 + 0{,}385\pi_3 + 0{,}583\pi_4 + 0{,}25\pi_5 \quad (13)$$

$$\pi_4 = 0\pi_0 + 0\pi_1 + 0{,}129\pi_2 + 0{,}062\pi_3 + 0\pi_4 + 0\pi_5 \quad (14)$$

$$\pi_5 = 0\pi_0 + 0\pi_1 + 0{,}016\pi_2 + 0{,}015\pi_3 + 0{,}167\pi_4 + 0\pi_5 \quad (15)$$

$$1 = \pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 \quad (16)$$

Por haver sete equações e seis incógnitas, é necessário desconsiderar uma das equações. Assim, resolvendo este sistema de equações, obtêm-se os seguintes resultados:

$\pi_0 = 0{,}036$ $\quad$ $\pi_1 = 0{,}095$

$\pi_2 = 0{,}378$ $\quad$ $\pi_3 = 0{,}393$

$\pi_4 = 0{,}073$ $\quad$ $\pi_5 = 0{,}024$

Estes valores representam a possibilidade de se encontrar nos determinados estados, conforme Tabela 2.

Intervalo	Probabilidade
< -2,01%	3,6%
de -2 a -1,01%	9,5%
de -1 a -0,01%	37,8%
de 0 a 0,99%	39,3%
de 1 a 1,99%	7,3%
> 2%	2,4%

Fonte: Elaborado pelos autores (2017)

Tabela 2 – Probabilidade dos estados

Com a Tabela 2, observa-se que a maior probabilidade é que o preço da soja varie de 0 a 0,99%, com uma probabilidade de 39,3%. Ainda, calculando com a Equação 6 previamente apresentada, é possível analisar o tempo de recorrência esperado para cada probabilidade de estado estável, conforme a Tabela 3.

Probabilidade de estado estável π_j	Tempo de recorrência esperado μ_{ii}
π_0	27,8 dias
π_1	10,5 dias
π_2	2,6 dias
π_3	2,5 dias
π_4	13,7 dias
π_5	41,7 dias

Fonte: Elaborado pelos autores (2017)

Tabela 3 – Tempo de recorrência esperado

Com a Tabela 3, observa-se que, quando se está analisando uma variação de 0 a 0,99%, por exemplo, o tempo de recorrência esperado para esta variação é de 2,5 dias, ou seja, é aguardado que em 2,5 dias ocorra novamente um valor que varie entre 0 a 0,99%.

4. Considerações finais

Este artigo estudou Cadeias de Markov com uma aplicação para analisar a variação do preço da soja, um grão tão importante para a economia do país e para a alimentação das pessoas. As MC, sendo um modelo estatístico de processos que pode ser utilizado em aplicações reais, fazem parte dos estudos de Pesquisa Operacional, que por sua vez, compõe a grade curricular da Engenharia de Produção.

Os cálculos deste trabalho foram realizados no *software* Excel, devido aos ganhos quanto rapidez e precisão nos resultados. Então, utilizando-se dados referentes a 2 de janeiro de 2017 a 31 de agosto de 2017, observou-se que a maior probabilidade de variação diária do preço da soja é de 0 a 0,99%, e que esta variação provavelmente ocorrerá novamente em 2,5 dias.

Informações como esta podem ser vantajosas no agronegócio da soja, pois a previsão da variação dos preços permite ajudar na tomada de decisão quanto a compra e/ou venda do *commodity*. Ainda, este conceito estudado pode ser utilizado como ferramenta de auxílio no processo de planejamento das organizações, sejam do setor público ou iniciativa privada. Assim, este trabalhou almejou instruir Cadeias de Markov e suas possíveis aplicações práticas, podendo fornecer informações para as organizações e, consequentemente, podem ter um melhor direcionamento de seus esforços e investimentos quanto a comercialização da soja, pelo monitoramento de seus preços.

Este artigo atingiu seus objetivos satisfatoriamente, por proporcionar resultados que podem ser estudados de forma analítica. Como trabalhos futuros, sugere-se o uso de Cadeias de Markov para se analisar outras *commodities*.

Referências bibliográficas

ANDERSEN, Anders Reenberg; NIELSEN, Bo Friis; REINHARDT, Line Blander. *Optimization of hospital ward resources with patient relocation using Markov chain modeling.* European Journal of Operational Research, v. 260, n. 3, p. 1152-1163, 2017.

BRASIL, Ministério da Agricultura. *Site institucional.* Disponível em: <http://www.agricultura.gov.br/>.

Cepea - Centro de Estudos Avançados em Economia Aplicada. *Site institucional*. Disponível em: <http://www.cepea.esalq.usp.br/br>.

HILLIER, Frederick; LIEBERMAN, Gerald. *Introduction to Operations Research*. 8. ed. Nova Iorque: McGraw-Hill, 2005.

JANTSCH, Leonardo. *Análise do risco de crédito no uso do cartão de crédito*. 2017, 80 f. Dissertação (Mestrado em Ciências Contábeis) – Universidade do Vale do Rio dos Sinos, São Leopoldo, 2017.

SILTALA, Lauri; GRANVIK, Mikael. *Asteroid mass estimation using Markov-chain Monte Carlo*. Icarus, v. 297, p. 149-159, 2017.

STAUDT, Francielly Hedler; COELHO, Antonio Sérgio; GONÇALVES, Mirian Buss. *Determinação da capacidade real necessária de um processo produtivo utilizando cadeia de Markov*. Production, v. 21, n. 4, p. 634-644, 2011.

TAHA, Hamdy A. *Operations research*: an introduction. 8. ed. Nova Iorque: Pearson, 2007.