



Uso de Modelos de Decomposição em Componentes Não Observáveis para predição de séries temporais não estacionárias de commodities agrícolas.

Lucas Valle Mielke¹

ICMC - USP

Paulino Ribeiro Villas Boas²

Embrapa Instrumentação

ICMC - USP

1 Introdução

A predição de séries temporais é frequentemente feita por Modelos Autoregressivos de Médias Móveis (ARMA, Autoregressive Moving Average em inglês) sendo a principal desvantagem desses modelos o requisito de que as séries temporais estudadas sejam estacionárias, o que frequentemente não ocorre [1]. Para contornar esse problema, geralmente são feitos processos de diferenciação de dados, que podem ser feitos diretamente no conjunto de dados ao se calcular a diferença do valor da variável no passo t com o valor do passo anterior $t - 1$ [4], ou configurando os modelos para diferenciar sucessivamente a série analisada até ela se tornar estacionária, como no caso dos Modelos Autorregressivos Integrados de Médias Móveis (ARIMA, Autoregressive Integrated Moving Average em inglês) e suas variações [3].

Apesar da possibilidade de diferenciação, os resultados dos modelos ARIMA continuam sendo sensíveis a dados perturbados e que não se desenvolvem ao redor de uma média constante [3]. Diante disso, os Modelos de Decomposição em Componentes Não Observáveis (UCM, Unobserved Component Model em Inglês) surgem como uma alternativa promissora a esses modelos por não assumirem estacionariedade dos dados, além disso, os modelos UCM podem ser compreendidos em componentes de tendência, ciclo e perturbação, o que os torna especialmente úteis para analisar séries que possuem ciclo, como é o caso das culturas agrícolas [1]. Diante disso, o objetivo desse estudo foi comparar o desempenho de Modelos de Decomposição em Componentes Não Observáveis com o desempenho de Modelos Autoregressivos de Médias Móveis para a predição

¹lucas.mielke@usp.br

²paulino.villas-boas@embrapa.br

de séries temporais não estacionárias, no caso o preço das commodities do Arroz, Café, Milho, Soja e Trigo.

2 Materiais e Métodos

Base de dados

Para elaboração do modelo proposto, foram coletados preços à vista em Reais e Dólar de Arroz, Café, Milho, Soja e Trigo da pesquisa de preço realizada pelo Centro de Estudos Avançados em Economia Aplicada (CEPEA) da Universidade de São Paulo (USP) para o período entre 30 de Junho de 2012 e 30 de Junho de 2022, totalizando 10 anos. O preço à vista em reais foi utilizado como variável preditiva ou dependente, enquanto que o preço em dólar foi utilizado para cálculo da taxa de dólar de cada amostra, a ser utilizada como variável explicativa ou independente.

Análises e Aprendizagem de Máquina

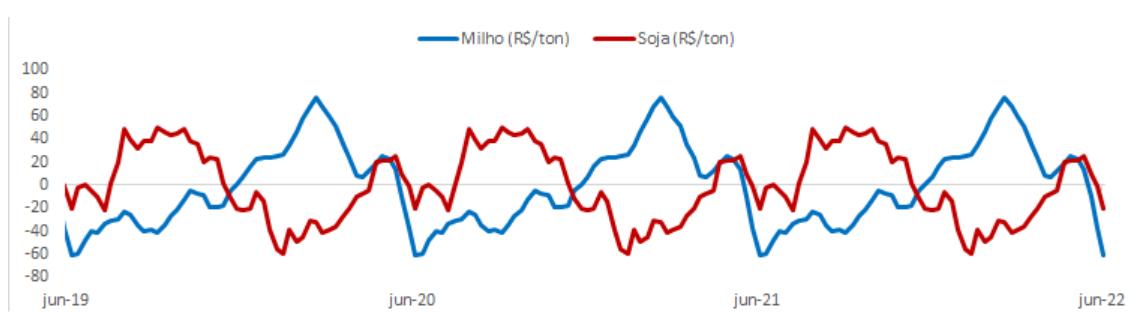
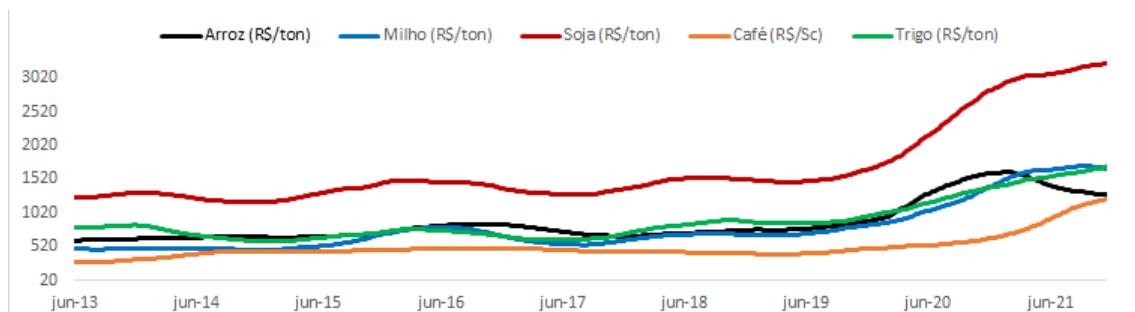
Para certificarmos que as séries temporais analisadas não são estacionárias, realizamos o teste de raiz unitária Dickey–Fuller Aumentado (ADF, Augmented Dickey–Fuller em inglês), onde a hipótese nula ($p\text{-value} > 0.05$) é de que a serie possui raízes unitárias e pertence a séries não estacionárias e, a alternativa ($p\text{-value} \leq 0.05$), de que não existe raíz unitária e a série é uma sequência estacionária. A Tabela 1 exibe o resultado do testes para cada uma das culturas analisadas. Também realizamos a decomposição da série em tendência, ciclo e perturbação a fim de verificar o comportamento desses componentes para todas as séries e identificar padrões de comportamento [2].

Para a aprendizagem de máquina, expressamos o problema na forma de equação onde temos uma variável Y de interesse que pode ser explicada por vetores X , chamados de variáveis explicativas ou independentes da forma como ocorre em um modelo linear clássico. O conjunto de dados com os vetores foram preparados para treino e testes da aprendizagem de máquina. Foi mantida a sequência original das séries, das quais 85% dos dados mais antigos foram usados para treino e 15% dos dados mais recentes, para teste.

Os dados foram treinados e testados com os modelos UCM e ARIMA sazonal com fatores exógenos SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors, em inglês). O segundo permite trabalhar com séries Sazonais, como é o caso de culturas agrícolas, e incluir variáveis exógenas na equação, no caso a taxa de dólar. Em seguida o resultado de cada um dos modelos foi comparado para cada uma das culturas pela medição do Erro Percentual Absoluto Médio (MAPE, Mean Absolute Percentage Error em Inglês). A Tabela 1 exibe o MAPE de ambos modelos para todas culturas estudadas.

3 Resultados e discussão

O Tesde ADF obteve valor-p superior a 0,05 para todas as culturas agrícolas analisadas, indicando que o preço de nenhuma delas é estacionário como era esperado. A Tabela 1 apresenta os resultados do teste ADF para todas as séries. A decomposição das séries gerou curvas de tendência, ciclo e ruído. As curvas de tendência nos permitiu verificar relativa estabilidade de



preço para todas as commodities entre o início do período analisado até o fim do ano de 2019, quando houve crescimento acentuado de preços, especialmente entre o início de 2020 e metade de 2021 para o Milho, Soja e Trigo. O Arroz também apresentou tendência de crescimento acentuado a partir de 2020, mas desde janeiro de 2021 apresenta tendência de redução enquanto o café teve tendência de crescimento mais tarde, a partir de dezembro de 2020. As curvas de ciclo, por sua vez, nos demonstraram a sazonalidade anual dos preços das culturas analisadas, com pico de preço ocorrendo no segundo semestre, com exceção do Milho, cujo período de maior sazonalidade ocorre muito próximo ao momento de menor sazonalidade da Soja. A Figura 1 apresenta as curvas de tendência de todas as séries analisadas enquanto a Figura 2 apresenta as curvas de ciclo do Milho e da Soja.

Ao se analisar os resultados dos modelos de aprendizagem de máquina, verificou-se que o MAPE foi inferior com o uso do modelo UCM em todas as culturas, indicando que este modelo pode ser mais promissor para o tipo de análise desse trabalho. Entretanto, cabe destacar que o erro ficou superior a 18% para todas as culturas, indicando que ambos os modelos não descreveram de forma razoável a variável preditiva em função da variável explicativa e que são necessários mais estudos com inclusão mais variáveis explicativas. A Tabela 1 apresenta o MAPE para todos os estudos feitos.

Cultura	Teste ADF	MAPE (%)	
		UCM	SARIMAX
Arroz	0,5424	18,7%	19,6%
Café	0,9968	31,4%	41,7%
Milho	0,9623	32,5%	42,3%
Soja	0,9933	25,7%	31,5%
Trigo	0,9978	20,8%	22,9%

Tabela 1: Testes ADF e MAPE.

4 Conclusões

Neste trabalho, verificamos que o preço das culturas de Arroz, Café, Milho, Soja e Trigo não são estacionárias pelo teste ADF e possuíram tendência de crescimento acentuado em alguns períodos conforme Figura 1. Também verificamos que o Modelo UCM apresentou menor MAPE que o Modelo SARIMAX em todas as culturas, indicando que este pode ser mais promissor para o tipo de análise feita nesse trabalho. Entretanto, o MAPE foi superior a 18% em todos os casos, indicando que ambos os modelos não descreveram a variável preditiva de forma razoável, sendo necessário explorar outras variáveis a fim de melhorar a capacidade desses modelos em descrever a variável preditiva ‘preço’ das commodities estudadas.

Referências

- [1] N. K. K. Brintha et al. Use of Unobserved Components Model for Forecasting Non-stationary Time Series: A Case of Annual National Coconut Production in Sri Lanka. Disponível em: <https://pdfs.semanticscholar.org/761e/0ac6a53020b7bee95751e489b16275397cdd.pdf>. 2014.
- [2] J. Brownlee. How to Decompose Time Series Data into Trend and Seasonality. 2017. *Machine Learning Mastery*. <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>
- [3] J. E. C. Lima et al. Aplicação do Modelo SARIMA na Previsão de Demanda no Setor Calçadista. *Revista Multidisciplinar de Psicologia*. 2019. Disponível em: <https://idonline.emnuvens.com.br/id/article/view/1875>.
- [4] L. Rabelo. Princípios básicos para criar previsões de Séries Temporais. 2019. *Ensina.AI*. Disponível em: <https://medium.com/ensina-ai/princ>