PREVISÃO DOS PREÇOS DE MILHO E SOJA COM VARIÁVEIS EXÓGENAS: MODELOS ARIMAX, SARIMAX, PROPHET E LSTM COMBINADOS POR ENSEMBLE PONDERADO

Alexandre Augusto Angelo de Souza – Universidade Federal da Grande Dourados - FACET
alexandreasouza@ufgd.edu.br

Felipe José Carbone – Universidade Federal da Grande Dourados - FACET
felipecarbone@ufgd.edu.br

Priscila Marques Kai – Universidade Federal da Grande Dourados - FACET
priscilakai@ufgd.edu.br

Nicholas Antunes Cinto – Visum Digital
nicholaas.nick@gmail.com

RESUMO

Este trabalho avalia se a incorporação de variáveis exógenas melhora a previsão mensal dos preços de milho e soja no mercado brasileiro, medidos pelos indicadores CEPEA/ESALQ entre março de 2016 e março de 2026 (121 observações por série). Como exógenas foram adotados a taxa de câmbio R$/US$ do Banco Central do Brasil e os preços internacionais de milho e soja do Fundo Monetário Internacional, que juntos compõem a paridade de exportação dos grãos. Foram comparados cinco modelos univariados (ARIMA, SARIMA, ETS, Prophet e LSTM) e quatro versões com exógenas (ARIMAX, SARIMAX, Prophet com regressores e LSTM multivariada), além de dois ensembles ponderados pelo inverso do RMSE quadrático, um formado apenas pelos univariados e outro apenas pelos modelos com exógenas. Os pesos foram estimados em seis janelas de validação walk-forward encerradas antes do holdout de doze meses, e as exógenas do horizonte foram previstas somente com informação disponível na origem, reproduzindo uma previsão ex ante. Na média do walk-forward, as exógenas reduziram o RMSE em sete dos oito pares de modelos, com ganhos de até 40% no SARIMA da soja. No holdout de abril de 2025 a março de 2026, porém, o ganho foi modesto no milho, onde o ensemble com exógenas obteve RMSE de R$ 16,98 contra R$ 17,31 do univariado, e negativo na soja (R$ 5,79 contra R$ 4,61), porque a apreciação do real em 2025 não foi antecipada pela previsão do câmbio. Quando as exógenas observadas substituem as previstas, o RMSE do ensemble do milho cai para R$ 13,32. As elasticidades estimadas indicam transmissão quase integral do câmbio e do preço internacional para a soja e parcial para o milho. Conclui-se que as exógenas carregam informação relevante, mas que o ganho preditivo efetivo depende da qualidade da previsão das próprias exógenas.

Palavras-chave: Previsão de Preços Agrícolas; Variáveis Exógenas; SARIMAX; LSTM; Ensemble; Paridade de Exportação.

ABSTRACT

This paper assesses whether exogenous variables improve monthly forecasts of Brazilian corn and soybean prices, measured by the CEPEA/ESALQ indicators from March 2016 to March 2026 (121 observations per series). The BRL/USD exchange rate from the Central Bank of Brazil and the international corn and soybean prices from the International Monetary Fund, which together form the export parity of both grains, were used as exogenous inputs. Five univariate models (ARIMA, SARIMA, ETS, Prophet and LSTM) were compared with four exogenous counterparts (ARIMAX, SARIMAX, Prophet with regressors and multivariate LSTM), plus two ensembles weighted by the inverse squared RMSE, one built only from univariate models and the other only from exogenous models. Weights were estimated on six walk-forward windows ending before a twelve-month holdout, and the

exogenous paths over the horizon were forecast using only information available at the origin, reproducing an ex ante setting. On average over the walk-forward windows, exogenous inputs lowered the RMSE in seven of the eight model pairs, with gains of up to 40% for the soybean SARIMA. In the April 2025 to March 2026 holdout, however, the gain was modest for corn, where the exogenous ensemble reached an RMSE of R$ 16.98 against R$ 17.31 for the univariate one, and negative for soybeans (R$ 5.79 against R$ 4.61), because the appreciation of the real in 2025 was not anticipated by the exchange rate forecast. When observed exogenous values replace the forecast ones, the corn ensemble RMSE falls to R$ 13.32. The estimated elasticities indicate an almost complete pass-through of the exchange rate and the international price for soybeans and a partial one for corn. Exogenous variables carry relevant information, but the effective predictive gain depends on how well the exogenous variables themselves can be forecast.

Keywords: Agricultural Price Forecasting; Exogenous Variables; SARIMAX; LSTM; Ensemble; Export Parity.

Introdução

O agronegócio ocupa posição central na economia brasileira, e o milho e a soja respondem por parcela expressiva da produção de grãos, das exportações e da renda gerada no campo, sobretudo em estados do Centro-Oeste como Mato Grosso do Sul. A capacidade de antecipar o comportamento dos preços dessas commodities orienta decisões de comercialização, de estocagem, de contratação de crédito e de proteção de risco de produtores, cooperativas, tradings e agroindústrias, além de subsidiar a formulação de políticas públicas. Allen (1994), em revisão sobre a previsão econômica na agricultura, já observava que o valor prático dessas previsões depende tanto da escolha do modelo quanto da informação que o modelo consegue incorporar sobre os determinantes do mercado.

Boa parte dos estudos aplicados de previsão de preços agrícolas adota modelos univariados, nos quais o preço futuro é projetado apenas a partir do próprio histórico da série. Modelos da família Box-Jenkins (BOX et al., 2015), o alisamento exponencial em espaço de estados (HYNDMAN et al., 2002), o Prophet (TAYLOR; LETHAM, 2018) e as redes neurais recorrentes do tipo LSTM (HOCHREITER; SCHMIDHUBER, 1997) compartilham essa limitação, ainda que com hipóteses estruturais muito distintas. Estudos recentes comparando essas abordagens em diferentes mercados, como os de Paul et al. (2022) e de Manogna, Dharmaji e Sarang (2025), mostram que nenhuma delas domina de forma consistente, e que o desempenho relativo varia conforme a volatilidade e as quebras estruturais de cada série.

No caso brasileiro, porém, há razões econômicas claras para esperar que informação externa à série melhore a previsão. Milho e soja são commodities comercializadas internacionalmente, e seus preços domésticos tendem a seguir a paridade de exportação, isto é, o preço internacional convertido pela taxa de câmbio e ajustado por custos logísticos e prêmios portuários. Mundlak e Larson (1992) documentaram que a maior parte da variação dos preços internacionais é transmitida aos preços domésticos, e Margarido, Turolla e Bueno (2007) encontraram evidência dessa transmissão especificamente para a soja no Brasil. Incorporar o câmbio e o preço internacional como variáveis exógenas é, portanto, um caminho natural para aproximar os modelos de previsão dos fundamentos de formação de preço.

Essa incorporação, contudo, não é trivial do ponto de vista preditivo. Para prever o preço doméstico doze meses à frente com um modelo que depende do câmbio e do preço internacional, é necessário também prever essas variáveis ao longo do mesmo horizonte, o que caracteriza a chamada previsão ex ante (HYNDMAN; ATHANASOPOULOS, 2021). Avaliações que utilizam os valores efetivamente observados das exógenas no período de teste, conhecidas como previsões ex post, tendem a superestimar o ganho que seria obtido em uma aplicação real, pois transferem ao modelo uma informação que não estaria disponível no momento da previsão.

Diante disso, o objetivo deste artigo é avaliar, em condições ex ante, se a inclusão do câmbio e do preço internacional como variáveis exógenas melhora a previsão mensal dos preços de milho e soja publicados pelo CEPEA, comparando cinco modelos univariados com suas quatro contrapartes com exógenas e dois ensembles ponderados, um formado apenas por modelos univariados e outro apenas por modelos com exógenas. O estudo contribui em três frentes. A primeira é a comparação pareada, modelo a modelo, entre as versões com e sem exógenas, tanto em validação walk-forward quanto em holdout. A segunda é a decomposição do ganho das exógenas entre o que decorre da informação que elas carregam, medido pelo cenário ex post, e o que se perde ao ter de prevê-las, medido pelo cenário ex ante. A terceira é a estimação das elasticidades do preço doméstico em relação ao câmbio e ao preço internacional, que ajudam a interpretar economicamente as diferenças de desempenho entre as duas commodities.

Referencial teórico

A formação do preço doméstico de uma commodity exportável pode ser descrita, em primeira aproximação, pela lei do preço único, segundo a qual o preço interno converge para o preço internacional convertido pela taxa de câmbio, descontados custos de transporte, tributos e margens de comercialização. Mundlak e Larson (1992) estimaram elasticidades de transmissão dos preços mundiais para os preços domésticos em dezenas de países e concluíram que essa transmissão é elevada e predominantemente integral no longo prazo, ainda que sujeita a desvios temporários. Para a soja brasileira, Margarido, Turolla e Bueno (2007) identificaram que as variações do preço internacional e do câmbio são transmitidas ao preço doméstico, com efeitos que dependem do calendário da safra e do comércio. O milho, por sua vez, tem parcela maior da produção destinada ao consumo interno, na alimentação animal e mais recentemente na produção de etanol, o que tende a tornar seu preço mais sensível a choques domésticos de oferta, como o tamanho da segunda safra, e menos atrelado à paridade de exportação.

Do lado dos métodos, os modelos ARIMA de Box et al. (2015) representam a série diferenciada como um processo autorregressivo e de médias móveis, e o SARIMA acrescenta operadores sazonais. A extensão desses modelos para incluir variáveis explicativas é conhecida como regressão dinâmica, ou regressão com erros ARIMA, na qual a variável de interesse é explicada por uma combinação linear de regressores e o erro dessa regressão segue um processo ARIMA (PANKRATZ, 1991; HYNDMAN; ATHANASOPOULOS, 2021). Essa formulação, usualmente denominada ARIMAX ou SARIMAX, preserva a estrutura de autocorrelação dos modelos univariados e permite interpretar diretamente os coeficientes dos regressores. A

seleção automática das ordens por critério de informação, proposta por Hyndman e Khandakar (2008), tornou essa família de modelos aplicável de forma sistemática a muitas séries.

O alisamento exponencial com tendência amortecida, formalizado por Gardner e McKenzie (1985) e inserido no arcabouço de espaço de estados por Hyndman et al. (2002), decompõe a série em nível, tendência e sazonalidade, com um parâmetro de amortecimento que faz a tendência convergir para um patamar constante em horizontes longos. O Prophet, de Taylor e Letham (2018), adota uma decomposição aditiva com tendência linear por partes, pontos de mudança automáticos e sazonalidade por séries de Fourier, e aceita regressores externos adicionados linearmente ao modelo. As redes LSTM de Hochreiter e Schmidhuber (1997) introduzem células de memória controladas por portas que permitem aprender dependências de longo prazo, e podem receber como entrada vetores com várias variáveis a cada passo de tempo, o que as torna naturalmente multivariadas. Em amostras curtas, técnicas de regularização como o dropout (SRIVASTAVA et al., 2014) e a otimização adaptativa de Kingma e Ba (2015) são usualmente empregadas para reduzir o sobreajuste.

A combinação de previsões, proposta por Bates e Granger (1969), parte da constatação de que modelos distintos erram em momentos distintos, de modo que uma média ponderada tende a ser mais estável do que a escolha de um único modelo. Timmermann (2006) mostra que esquemas simples de ponderação, como os baseados no inverso do erro quadrático passado, costumam ser competitivos frente a esquemas mais sofisticados, sobretudo quando o número de observações para estimar os pesos é pequeno. Ribeiro e Coelho (2020) aplicaram ensembles a séries do agronegócio brasileiro e obtiveram ganhos consistentes sobre os modelos individuais. Por fim, a avaliação de previsões em séries temporais exige que o modelo seja testado apenas com dados posteriores à origem da previsão, e a validação com origem móvel, ou walk-forward, é recomendada por Tashman (2000) como forma de reduzir a dependência dos resultados em relação a um único período de teste. Hyndman e Koehler (2006) discutem as propriedades das métricas de erro mais usadas, como MAE, RMSE e MAPE.

Método de pesquisa

Trata-se de uma pesquisa aplicada, de abordagem quantitativa, conduzida por meio de experimentos computacionais de previsão sobre séries temporais públicas. As variáveis-alvo são o Indicador do Milho ESALQ/BM&FBovespa e o Indicador da Soja CEPEA/ESALQ Paraná, ambos em reais por saca de 60 kg, obtidos no portal do Centro de Estudos Avançados em Economia Aplicada (CEPEA, 2026) com periodicidade mensal de março de 2016 a março de 2026, totalizando 121 observações por série. As variáveis exógenas são a taxa de câmbio livre do dólar americano, na média mensal de venda, obtida no Sistema Gerenciador de Séries Temporais do Banco Central do Brasil (BCB, 2026) pela série 3698, e os preços internacionais de soja e de milho em dólares por tonelada, calculados pelo Fundo Monetário Internacional e obtidos na base FRED do Federal Reserve Bank of St. Louis pelas séries PSOYBUSDM e PMAIZMTUSDM (FRED, 2026). A Tabela 1 resume as estatísticas descritivas no período de estudo.

Tabela 1 – Estatísticas descritivas das séries, março de 2016 a março de 2026 (N = 121)

Série	Unidade	Mínimo	Máximo	Média	Desvio-padrão	Coef. de variação
Milho (CEPEA)	R$/saca	26,33	100,72	59,85	21,07	35,2%
Soja (CEPEA)	R$/saca	61,16	195,85	117,69	40,12	34,1%
Câmbio	R$/US$	3,10	6,10	4,62	0,91	19,6%
Milho internacional	US$/t	143,99	348,51	204,53	54,76	26,8%
Soja internacional	US$/t	305,31	621,20	413,56	86,67	21,0%

Fonte: elaborado pelos autores a partir de CEPEA (2026), BCB (2026) e FRED (2026).

Para explicitar o vínculo econômico entre as exógenas e os preços domésticos, foi calculada a paridade de exportação bruta de cada grão, definida pela Equação (1), em que o fator 0,06 converte o preço por tonelada para o preço por saca de 60 kg. Essa paridade não inclui fretes, tributos nem prêmios portuários, e por isso não coincide exatamente com o preço doméstico, mas captura a parcela do preço determinada pelo mercado internacional e pelo câmbio.

$$P_t^{par} = P_t^{ext} \times E_t \times 0{,}06 \quad (1)$$

Nessa expressão, P^{ext}_t é o preço internacional em dólares por tonelada e E_t é a taxa de câmbio em reais por dólar no mês t. Todas as séries foram transformadas por logaritmo natural antes da modelagem, o que estabiliza a variância e faz com que os coeficientes das exógenas nos modelos de regressão possam ser lidos como elasticidades. A reversão para a escala original incorpora a correção de viés log-normal, pela qual a previsão em nível é obtida como $\exp(\hat{y} + \sigma^2/2)$, em que σ^2 é a variância dos resíduos do modelo na escala logarítmica. Na estimação dessa variância foram descartados os primeiros $d + D \cdot m$ resíduos de cada modelo ARIMA, que correspondem à inicialização difusa do filtro de Kalman e assumiriam valores próximos ao próprio nível da série, inflando artificialmente a correção.

Os modelos univariados de referência foram cinco. O ARIMA e o SARIMA tiveram as ordens selecionadas automaticamente pelo critério de Akaike (HYNDMAN; KHANDAKAR, 2008), com uma diferenciação simples imposta, busca de p e q até 4 no ARIMA e até 3 no SARIMA, uma diferenciação sazonal e busca das ordens sazonais P e Q até 2, com período m = 12. O ETS adotou a forma com erro, tendência amortecida e sazonalidade aditivos, estimada por máxima verossimilhança. O Prophet foi configurado com sazonalidade anual aditiva e escala a priori dos pontos de mudança igual a 0,15. A LSTM univariada seguiu a arquitetura descrita adiante para a versão multivariada, sem as colunas de exógenas.

Os modelos com exógenas foram quatro. O ARIMAX e o SARIMAX correspondem à regressão com erros ARIMA definida pelas Equações (2) e (3), em que y_t é o logaritmo do preço doméstico, $x_{k,t}$ é o logaritmo da k-ésima exógena, β_k é a elasticidade correspondente e η_t é o erro da regressão, que segue um processo ARIMA ou SARIMA com as mesmas regras de seleção dos modelos univariados.

$$y_t = \beta_0 + \Sigma_{k=1}^{K} \beta_k x_{k,t} + \eta_t \quad (2)$$

$$\Phi_P(B^m)\phi(B)(1 - B)^d(1 - B^m)^D \eta_t = \Theta_Q(B^m)\theta(B)\varepsilon_t \quad (3)$$

Na Equação (3), $\phi(B)$ e $\theta(B)$ são os polinômios autorregressivo e de médias móveis não sazonais, Φ_P e Θ_Q são seus correspondentes sazonais, B é o operador de defasagem e ε_t é um ruído branco. No ARIMAX os termos sazonais são omitidos. Em ambos os casos foram usadas como exógenas o câmbio e o preço internacional do próprio grão, de modo que K = 2. O Prophet com regressores, denominado Prophet-X, acrescenta à decomposição original um termo linear nas exógenas padronizadas, conforme a Equação (4), em que g(t) é a tendência linear por partes, s(t) é a sazonalidade anual e γ_k são os coeficientes dos regressores. Nessa versão o modelo foi ajustado ao logaritmo do preço, para manter coerência com a especificação em elasticidades dos demais modelos.

$$y(t) = g(t) + s(t) + \sum_{k=1}^{K} \gamma_k x_k(t) + \varepsilon_t \quad (4)$$

A LSTM multivariada, denominada LSTM-X, recebe janelas de 18 meses em que cada passo contém o preço defasado e as exógenas deslocadas um mês à frente, de modo que o último passo da janela já traz as exógenas do próprio mês a ser previsto. As variáveis foram normalizadas no intervalo de zero a um com parâmetros estimados apenas no treino. A rede tem uma camada LSTM com 24 unidades, dropout de 0,2 e uma camada densa de saída, treinada pelo otimizador Adam com taxa de aprendizado de 0,01 e parada antecipada com paciência de 15 épocas, monitorando a perda em uma fração cronológica final de 15% do treino. Para reduzir a dependência da inicialização, foram treinadas cinco redes com sementes distintas, e a previsão final é a média de suas trajetórias. A previsão de h passos é recursiva, e o incremento de cada passo em relação ao valor anterior é multiplicado por $0{,}85^h$, à semelhança do amortecimento do ETS, o que contém a extrapolação linear de tendências recentes.

Como os modelos com exógenas precisam dos valores do câmbio e do preço internacional ao longo do horizonte de previsão, cada exógena foi projetada por um modelo ARIMA não sazonal com uma diferenciação, ajustado em logaritmo e selecionado automaticamente, usando exclusivamente as observações disponíveis até a origem de cada previsão. Esse procedimento define o cenário ex ante, adotado em toda a avaliação principal. Como referência, o holdout foi repetido em um cenário ex post, no qual os modelos com exógenas recebem os valores efetivamente observados das exógenas no período de teste. A diferença entre os dois cenários mede quanto do potencial informativo das exógenas se perde pela necessidade de prevê-las.

Foram construídos dois ensembles. O Ensemble-U combina os cinco modelos univariados e o Ensemble-X combina os quatro modelos com exógenas, em ambos os casos com pesos proporcionais ao inverso do RMSE quadrático obtido na validação walk-forward, conforme as Equações (5) e (6), em que $RMSE_i$ é o erro médio do modelo i nas janelas de validação, $\hat{y}_{i,t}$ é sua previsão para o mês t e n é o número de modelos do ensemble.

$$w_i = \frac{\frac{1}{RMSE_i^2}}{\sum_{j=1}^{n} \frac{1}{RMSE_j^2}} \quad (5)$$

$$\hat{y}_{ens,t} = \sum_{i=1}^{n} w_i \hat{y}_{i,t} \quad (6)$$

A avaliação seguiu dois níveis. No primeiro, a validação walk-forward usou origens expansíveis em março de cada ano, de 2019 a 2024, com horizonte de doze meses em cada janela, de forma que todos os períodos de teste se encerram antes do holdout. O RMSE médio dessas seis janelas define os pesos dos ensembles avaliados no holdout, o que evita que a informação do período de teste final participe da ponderação. No segundo nível, o holdout usa os 109 meses de março de 2016 a março de 2025 como treino e os doze meses de abril de 2025 a março de 2026 como teste. Para a projeção de abril de 2026 a março de 2027, os modelos foram reajustados com a série completa e os pesos foram recalculados com as seis janelas mais recentes, de 2020 a 2025. O desempenho foi medido pelo erro absoluto médio (MAE), pela raiz do erro quadrático médio (RMSE) e pelo erro percentual absoluto médio (MAPE), definidos nas Equações (7) a (9), em que y_t é o valor observado, $\hat{y}_t$ é o previsto e $H = 12$ é o horizonte.

$$MAE = \frac{1}{H}\sum_{t=1}^{H}|y_t - \hat{y}_t| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{H}\sum_{t=1}^{H}(y_t - \hat{y}_t)^2} \tag{8}$$

$$MAPE = \frac{100}{H}\sum_{t=1}^{H}\frac{|y_t - \hat{y}_t|}{y_t} \tag{9}$$

A faixa de incerteza das projeções futuras foi calculada como $\hat{y}_{ens,t} \pm \sigma_e\sqrt{h}$, em que σ_e é a média, ponderada pelos pesos do Ensemble-X, dos desvios-padrão dos resíduos de ajuste de cada modelo em reais por saca e h é o horizonte em meses. Essa faixa não incorpora a incerteza adicional decorrente da previsão das exógenas e deve ser lida como uma indicação da dispersão esperada, e não como um intervalo de confiança formal. Todo o procedimento foi implementado em Python, com as bibliotecas pmdarima 2.0.4, statsmodels 0.14.1 (SEABOLD; PERKTOLD, 2010), Prophet 1.1.6 e TensorFlow 2.16.2, e o código foi organizado de modo que cada execução reproduz integralmente os resultados aqui apresentados.

Resultados e discussão

A Figura 1 compara os indicadores do CEPEA com a paridade de exportação bruta da Equação (1). As duas curvas caminham juntas ao longo de toda a década, inclusive no ciclo de alta de 2020 a 2022, associado à desvalorização do real e à elevação dos preços internacionais, e na correção posterior. A correlação em nível entre preço e paridade é de 0,93 para o milho e de 0,97 para a soja, e a correlação entre as variações mensais em logaritmo é de 0,45 e 0,66, respectivamente. A razão média entre o preço doméstico e a paridade é de 1,05 no milho e de 1,01 na soja. Esses números confirmam que a soja se comporta de forma muito próxima à paridade, enquanto o milho apresenta desvios maiores e mais persistentes, como a queda acentuada de 2023 e o pico de março de 2025 seguido de reversão, episódios em que o preço doméstico se afastou do preço internacional convertido.

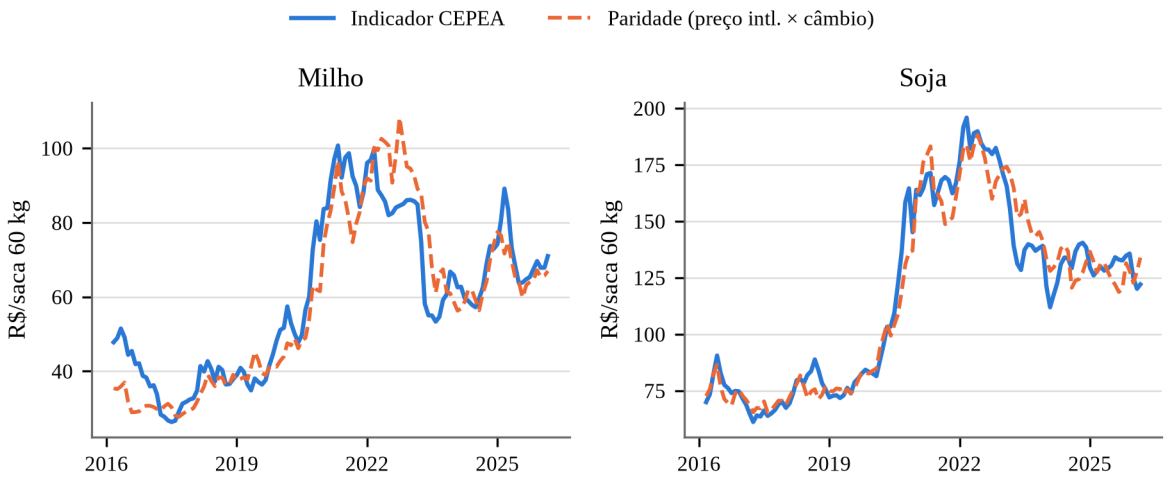


Figura 1 – Indicadores CEPEA e paridade de exportação bruta (preço internacional × câmbio × 0,06)
Fonte: elaborado pelos autores a partir de CEPEA (2026), BCB (2026) e FRED (2026).

A Tabela 2 apresenta as ordens selecionadas para os modelos Box-Jenkins no holdout e as elasticidades estimadas nos modelos com exógenas. Na soja, a elasticidade do preço doméstico em relação ao câmbio foi de 0,92 no ARIMAX e de 0,93 no SARIMAX, e em relação ao preço internacional de 0,71 e 0,80, valores próximos da unidade que indicam transmissão quase integral, em linha com Mundlak e Larson (1992) e com Margarido, Turolla e Bueno (2007). No milho, as elasticidades ficaram entre 0,53 e 0,61, indicando que apenas cerca de metade das variações do câmbio e do preço internacional chega ao preço doméstico no mesmo mês. Esse resultado é coerente com a maior importância do mercado interno na formação do preço do milho, e antecipa que as exógenas devem explicar uma parcela menor da dinâmica desse grão.

Tabela 2 – Ordens selecionadas e elasticidades estimadas no treino do holdout (mar/2016 a mar/2025)

Produto	Modelo	Ordem com exógenas	Ordem univariada	β câmbio	β preço intl.
Milho	ARIMAX	(0,1,1)	(1,1,0)	0,548	0,527
Milho	SARIMAX	(0,1,1)(0,1,2)12	(1,1,1)(2,1,0)12	0,582	0,614
Soja	ARIMAX	(1,1,1)	(0,1,1)	0,923	0,710
Soja	SARIMAX	(2,1,1)(0,1,1)12	(0,1,1)(2,1,0)12	0,931	0,800

Fonte: elaborado pelos autores.

A Tabela 3 reúne o RMSE médio de cada modelo nas seis janelas de validação walk-forward e os pesos resultantes nos dois ensembles. A comparação pareada é favorável às exógenas na maioria dos casos. Na soja, as quatro versões com exógenas superaram suas contrapartes univariadas, com reduções de 14,2% no ARIMA, 40,4% no SARIMA, 27,1% no Prophet e 2,3% na LSTM. No milho, as reduções foram de 37,4% no SARIMA e de 31,1% no Prophet, com ganho marginal de 1,0% na LSTM, enquanto o ARIMAX ficou praticamente empatado com o ARIMA, com RMSE 0,6% maior. O efeito mais expressivo ocorreu justamente nos modelos que, sem exógenas, tinham o pior desempenho, o SARIMA e o Prophet, o que sugere que as exógenas funcionam como uma âncora de nível que corrige extrapolações indevidas de tendência e de sazonalidade. Como consequência, os pesos do Ensemble-X ficaram mais equilibrados entre os quatro modelos, ao passo que no Ensemble-U o SARIMA e o Prophet receberam pesos reduzidos.

Tabela 3 – RMSE médio (R$/saca) nas seis janelas walk-forward (origens de mar/2019 a mar/2024) e peso de cada modelo no respectivo ensemble

Modelo	RMSE Milho	Peso Milho	RMSE Soja	Peso Soja
ARIMA	13,79	0,258	25,37	0,278
ARIMAX	13,88	0,248	21,77	0,318
SARIMA	22,12	0,101	42,53	0,099
SARIMAX	13,84	0,250	25,35	0,235
ETS	13,32	0,277	28,52	0,220
Prophet	19,13	0,134	37,02	0,131
Prophet-X	13,18	0,275	26,98	0,207
LSTM	14,63	0,230	25,64	0,272
LSTM-X	14,49	0,228	25,06	0,240

Fonte: elaborado pelos autores. Pesos dos univariados referem-se ao Ensemble-U e pesos dos modelos com exógenas ao Ensemble-X.

O holdout de abril de 2025 a março de 2026 coincidiu com um período atípico para o milho. Após atingir R$ 89,12 por saca em março de 2025, último mês do treino, o indicador caiu para R$ 63,63 em julho, com a entrada da segunda safra, e se recuperou apenas parcialmente até R$ 70,96 em março de 2026. A Tabela 4 e a Figura 2 mostram que o Prophet univariado foi o único modelo a acompanhar essa reversão, com RMSE de R$ 7,93, pois seus pontos de mudança de tendência já haviam registrado a inflexão recente da série. Os demais modelos, com ou sem exógenas, projetaram a continuidade do patamar elevado. Nos pares, as exógenas reduziram o RMSE do ARIMA de R$ 24,67 para R$ 23,60, do SARIMA de R$ 18,18 para R$ 17,05 e da LSTM de R$ 16,18 para R$ 15,09, mas elevaram o do Prophet de R$ 7,93 para R$ 13,85. O Ensemble-X obteve RMSE de R$ 16,98, 1,9% abaixo do Ensemble-U (R$ 17,31).

Tabela 4 – Métricas no holdout do milho (abr/2025 a mar/2026), cenário ex ante

Modelo	Exógenas	MAE (R$)	RMSE (R$)	MAPE (%)
ARIMA	Não	24,04	24,67	35,63
ARIMAX	Sim	22,99	23,60	34,07
SARIMA	Não	16,11	18,18	23,88
SARIMAX	Sim	15,92	17,05	23,71
ETS	Não	21,42	22,47	31,76
Prophet	**Não**	**6,51**	**7,93**	**9,13**
Prophet-X	Sim	12,52	13,85	18,42
LSTM	Não	15,31	16,18	22,87
LSTM-X	Sim	14,27	15,09	21,31
Ensemble-U	Não	16,36	17,31	24,38
Ensemble-X	Sim	16,03	16,98	23,88

Fonte: elaborado pelos autores. Em negrito, o menor RMSE.

A pequena vantagem das exógenas no holdout do milho tem duas explicações complementares. A primeira é que a queda de 2025 foi essencialmente um choque doméstico de oferta, e não um movimento da paridade, de modo que nem o câmbio nem o preço internacional, que permaneceu próximo de US$ 200 por tonelada, sinalizavam uma redução daquela magnitude. A segunda é a própria previsão das exógenas. Na origem de março de 2025 o câmbio estava em R$ 5,75, e o modelo ARIMA projetou sua manutenção em torno de R$ 5,8 a R$ 6,0 por dólar, quando o valor observado recuou de forma contínua até R$ 5,23 em março de 2026, com média

de R$ 5,44 no período. Com elasticidade positiva, esse erro de previsão do câmbio empurrou as previsões dos modelos com exógenas para cima, justamente quando o preço doméstico caía.

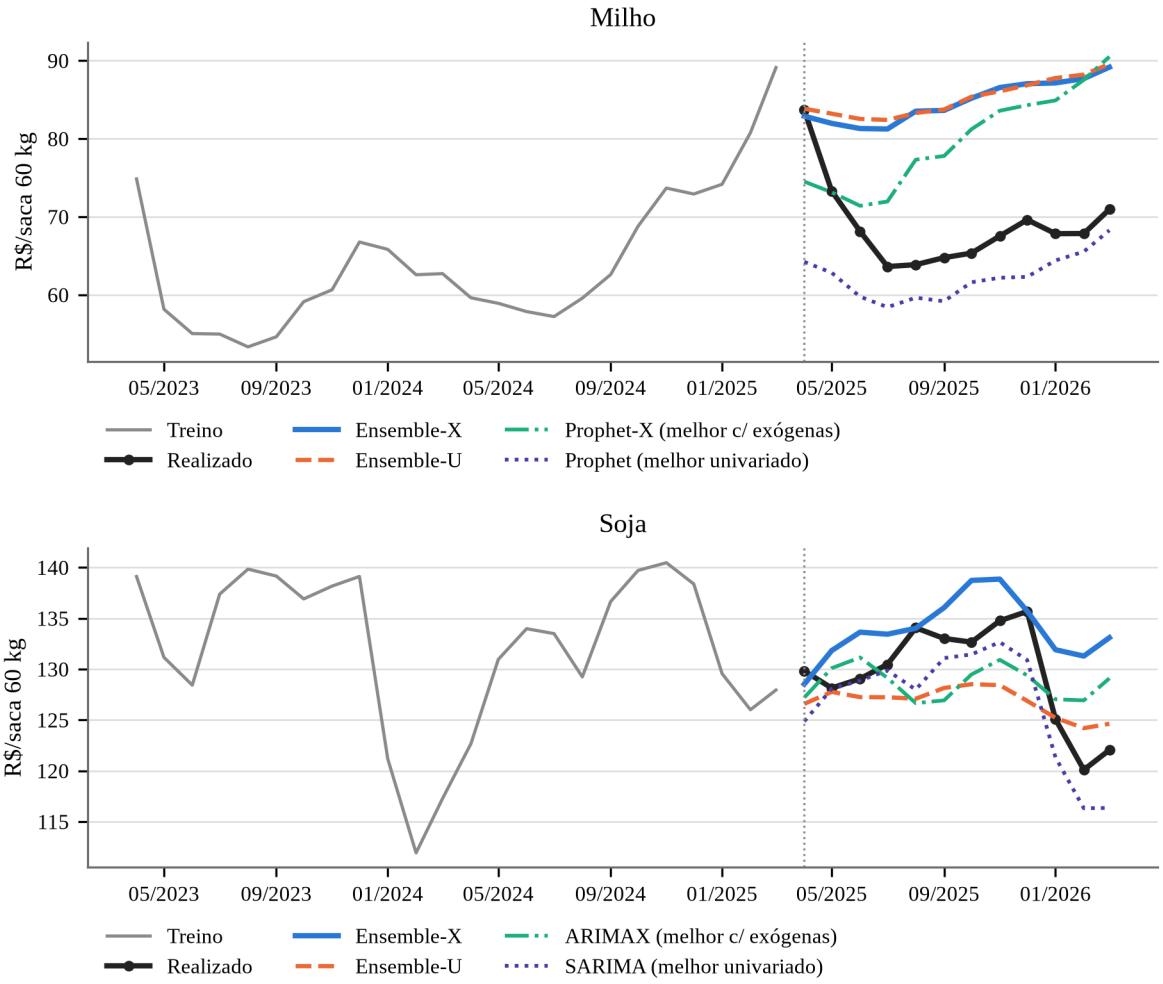


Figura 2 – Previsões no holdout (abr/2025 a mar/2026) dos ensembles e dos melhores modelos individuais com e sem exógenas

Fonte: elaborado pelos autores.

Na soja, a série permaneceu em um patamar relativamente estável no holdout, oscilando entre R$ 120 e R$ 136 por saca, e os modelos univariados de estrutura mais simples lideraram. Conforme a Tabela 5, o SARIMA obteve o menor RMSE, de R$ 3,60, seguido pelo ETS (R$ 3,86) e pelo Ensemble-U (R$ 4,61). O Ensemble-X ficou em R$ 5,79. Apenas o Prophet-X melhorou de forma expressiva frente ao Prophet, de R$ 17,24 para R$ 12,11, e o ARIMAX empatou com o ARIMA. Como a elasticidade da soja em relação ao câmbio é próxima de um, o erro de previsão do câmbio discutido para o milho foi transmitido quase integralmente às previsões, que superestimaram o preço de modo sistemático no segundo semestre de 2025 e no início de 2026, como se observa na Figura 2.

Tabela 5 – Métricas no holdout da soja (abr/2025 a mar/2026), cenário ex ante

Modelo	Exógenas	MAE (R$)	RMSE (R$)	MAPE (%)
ARIMA	Não	4,07	4,85	3,16
ARIMAX	Sim	4,22	4,77	3,27
SARIMA	**Não**	**2,93**	**3,60**	**2,27**
SARIMAX	Sim	5,33	6,10	4,13
ETS	Não	3,29	3,86	2,54
Prophet	Não	16,79	17,24	12,91
Prophet-X	Sim	10,95	12,11	8,51
LSTM	Não	3,95	4,85	3,10
LSTM-X	Sim	5,51	7,45	4,40
Ensemble-U	Não	3,88	4,61	2,96
Ensemble-X	Sim	4,57	5,79	3,62

Fonte: elaborado pelos autores. Em negrito, o menor RMSE.

Para separar o valor informativo das exógenas do erro introduzido por sua previsão, a Tabela 6 compara o RMSE dos modelos com exógenas nos cenários ex ante e ex post. No milho, quando os modelos recebem o câmbio e o preço internacional efetivamente observados, o RMSE do Ensemble-X cai de R$ 16,98 para R$ 13,32, uma redução de 21,6%, e passa a ficar 23,1% abaixo do Ensemble-U. O SARIMAX e a LSTM-X atingem R$ 13,02 e R$ 13,16, respectivamente. Na soja, o Ensemble-X melhora de R$ 5,79 para R$ 5,01 e o SARIMAX de R$ 6,10 para R$ 4,24, mas o conjunto permanece ligeiramente acima do Ensemble-U, e o ARIMAX chega a piorar. Esse comportamento indica que, em 2025, o preço doméstico da soja não acompanhou integralmente a queda do câmbio, possivelmente porque prêmios de exportação e fatores internos compensaram parte da valorização do real, um desvio temporário da paridade que modelos com elasticidade estimada próxima de um tendem a penalizar.

Tabela 6 – RMSE (R$/saca) no holdout com exógenas previstas (ex ante) e observadas (ex post)

Modelo	Milho ex ante	Milho ex post	Soja ex ante	Soja ex post
ARIMAX	23,60	19,44	4,77	7,77
SARIMAX	17,05	13,02	6,10	4,24
Prophet-X	13,85	10,56	12,11	11,07
LSTM-X	15,09	13,16	7,45	5,77
Ensemble-X	16,98	13,32	5,79	5,01
Ensemble-U (referência)	17,31	–	4,61	–

Fonte: elaborado pelos autores.

A Figura 3 sintetiza a comparação pareada no holdout ex ante. Os resultados mostram que a informação contida nas exógenas é real e economicamente interpretável, já que melhora a maior parte dos modelos na média das janelas de validação e melhora substancialmente o milho quando as exógenas são conhecidas. Ao mesmo tempo, evidenciam que o ganho efetivo em uma aplicação ex ante é limitado pela capacidade de prever o câmbio, variável notoriamente difícil de antecipar em horizontes de um ano. Em um único período de teste, um choque cambial não previsto pode anular ou inverter a vantagem observada em média, o que reforça a importância de avaliações com múltiplas origens, como a walk-forward, para conclusões sobre o valor preditivo das exógenas (TASHMAN, 2000).

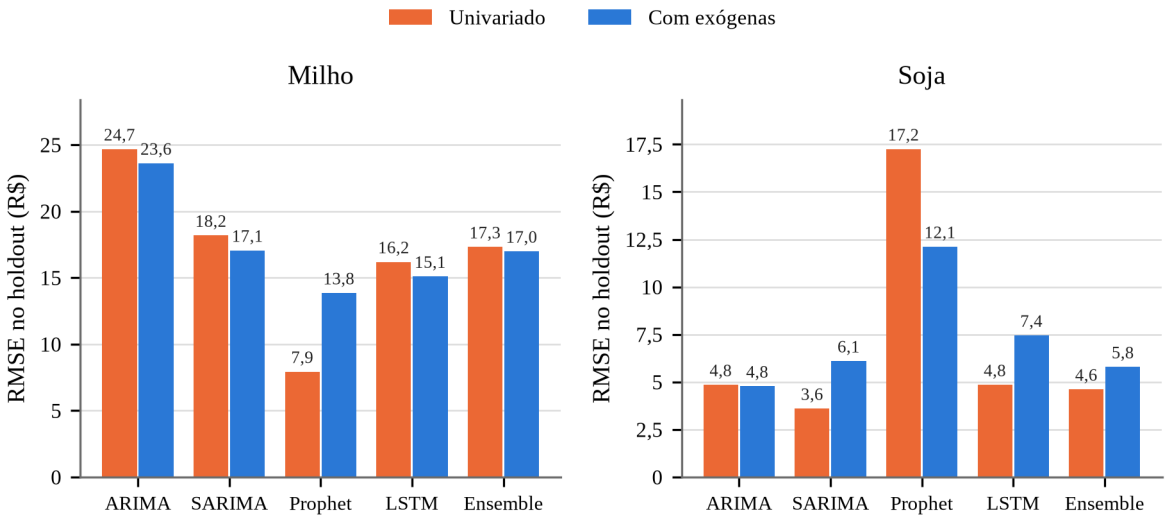


Figura 3 – RMSE no holdout (ex ante) de cada modelo em sua versão univariada e com exógenas
Fonte: elaborado pelos autores.

As projeções de abril de 2026 a março de 2027 estão na Tabela 7 e na Figura 4. Nesse exercício, os valores do câmbio e dos preços internacionais de abril a julho de 2026 já estavam publicados na data de coleta e foram usados diretamente, e os meses seguintes foram projetados pelo modelo ARIMA das exógenas. Para o milho, o Ensemble-X indica um recuo até R$ 62,93 em junho de 2026, seguido de recuperação gradual até R$ 71,85 em março de 2027, trajetória próxima à do Ensemble-U, com faixa de incerteza que se amplia de R$ 63,45 a R$ 72,40 no primeiro mês para R$ 56,34 a R$ 87,36 no último. Para a soja, as duas combinações divergem mais. O Ensemble-X projeta alta até R$ 134,15 em setembro de 2026, puxada pela recuperação do preço internacional observada até julho de 2026, e acomodação em R$ 128,16 ao fim do horizonte, enquanto o Ensemble-U permanece entre R$ 121,65 e R$ 126,05. Essa divergência ilustra a principal utilidade prática dos modelos com exógenas, que é permitir que a projeção do preço doméstico incorpore sinais já disponíveis no mercado internacional e no câmbio, algo que os modelos univariados só captariam com defasagem.

Tabela 7 – Projeções dos ensembles e faixa ±σ√h do Ensemble-X (R$/saca), abr/2026 a mar/2027

Mês	Milho Ens-U	Milho Ens-X	Milho faixa	Soja Ens-U	Soja Ens-X	Soja faixa
04/2026	69,58	67,93	63,45 a 72,40	123,54	121,00	114,04 a 127,96
05/2026	67,98	65,63	59,30 a 71,96	123,98	123,95	114,10 a 133,79
06/2026	66,83	62,93	55,17 a 70,68	123,63	123,79	111,73 a 135,85
07/2026	66,26	65,92	56,97 a 74,88	124,20	130,16	116,24 a 144,09
08/2026	66,93	68,26	58,25 a 78,26	124,80	131,93	116,36 a 147,49
09/2026	67,24	68,25	57,28 a 79,21	125,67	134,15	117,10 a 151,21
10/2026	68,52	68,28	56,44 a 80,12	125,87	133,48	115,06 a 151,90
11/2026	69,07	69,10	56,44 a 81,76	126,05	132,29	112,59 a 151,98
12/2026	69,98	70,44	57,01 a 83,86	125,28	132,14	111,26 a 153,03
01/2027	70,80	71,24	57,09 a 85,40	122,84	130,29	108,28 a 152,31
02/2027	71,14	71,25	56,41 a 86,10	121,65	128,51	105,42 a 151,60
03/2027	73,28	71,85	56,34 a 87,36	123,38	128,16	104,04 a 152,27

Fonte: elaborado pelos autores.

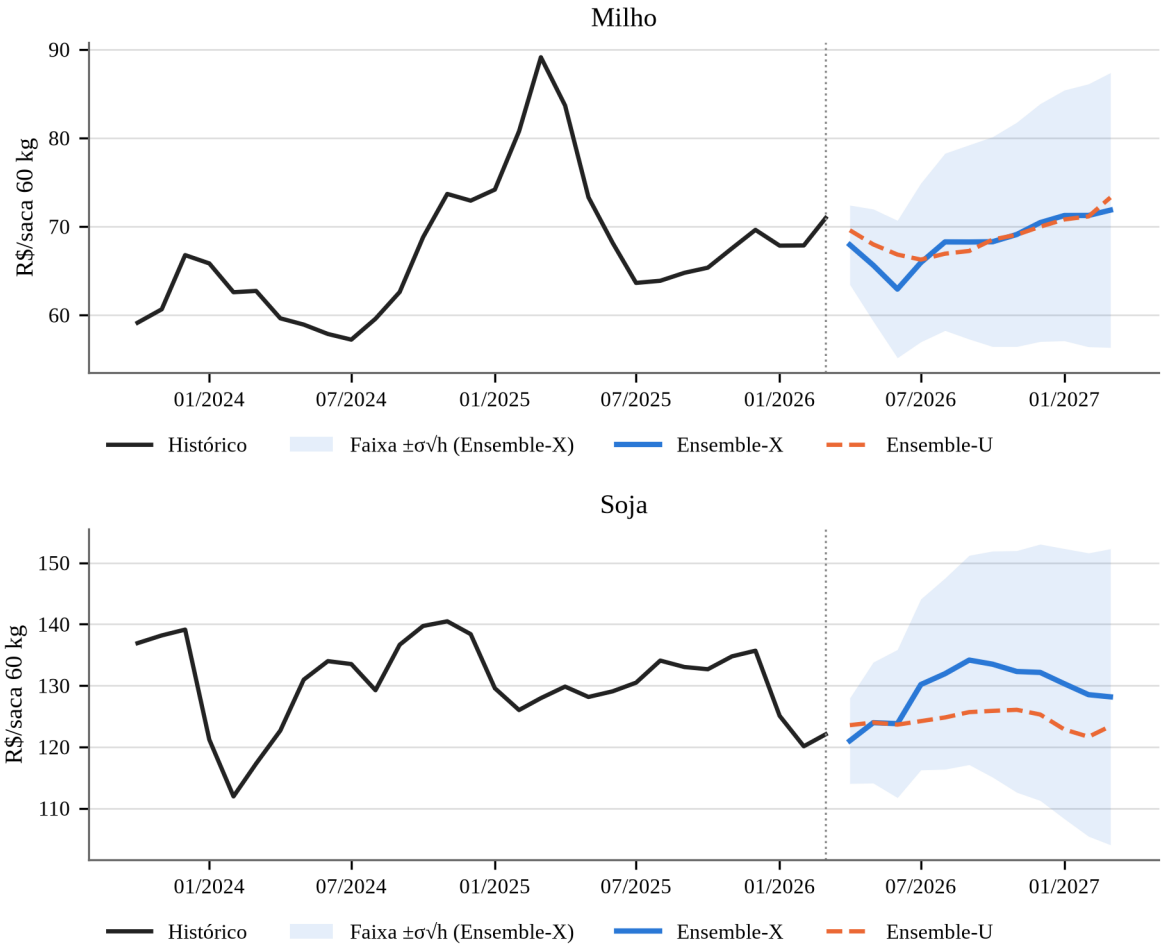


Figura 4 – Projeções de abril de 2026 a março de 2027 com faixa de incerteza do Ensemble-X
Fonte: elaborado pelos autores.

Considerações finais

Este artigo avaliou a incorporação do câmbio e do preço internacional como variáveis exógenas na previsão mensal dos preços de milho e soja do CEPEA, comparando cinco modelos univariados, quatro modelos com exógenas e dois ensembles ponderados, em um desenho de avaliação ex ante no qual as exógenas do horizonte são previstas apenas com informação disponível na origem e os pesos dos ensembles são estimados sem uso do período de teste final. As elasticidades estimadas confirmaram a transmissão quase integral da paridade de exportação para o preço da soja e uma transmissão parcial, em torno de metade, para o preço do milho, o que é coerente com a literatura de transmissão de preços e com o maior peso do mercado interno na formação do preço do milho.

Na validação walk-forward com seis origens, as exógenas reduziram o RMSE em sete dos oito pares de modelos, com ganhos maiores justamente nos modelos univariados mais propensos a extrapolar tendência e sazonalidade, o SARIMA e o Prophet. No holdout de abril de 2025 a março de 2026, porém, o ganho foi modesto no milho e negativo na soja, porque a apreciação do real ao longo de 2025 não foi antecipada pela previsão do câmbio. O cenário ex post mostrou

que, com as exógenas conhecidas, o ensemble com exógenas supera o univariado no milho por margem expressiva, o que permite concluir que a informação das exógenas é relevante e que o gargalo está na qualidade de sua previsão. Em termos práticos, recomenda-se usar os modelos com exógenas principalmente como ferramenta de cenários, alimentada por projeções de mercado para o câmbio e para os preços internacionais, e manter a comparação com os modelos univariados como salvaguarda.

O estudo tem limitações que orientam trabalhos futuros. A avaliação final se apoia em um único período de teste de doze meses, e a comparação estatística formal entre os modelos, por testes como o de Diebold e Mariano, exigiria um número maior de origens de previsão. As exógenas foram previstas por modelos ARIMA simples, e a substituição dessas projeções pela curva de contratos futuros da CBOT e da B3 e pelas expectativas de câmbio do Boletim Focus tende a aproximar o cenário ex ante do cenário ex post. Também se recomenda incluir variáveis domésticas de oferta, como estimativas de safra e estoques da CONAB, que poderiam capturar choques como a queda do milho em 2025, avaliar horizontes mais curtos, de um a três meses, nos quais a informação recente das exógenas tende a ser mais valiosa, e estender a análise a outras commodities agrícolas.

Referências

ALLEN, P. G. Economic forecasting in agriculture. *International Journal of Forecasting*, v. 10, n. 1, p. 81-135, 1994.

BANCO CENTRAL DO BRASIL (BCB). *Sistema Gerenciador de Séries Temporais*: série 3698, taxa de câmbio livre, dólar americano (venda), média de período, mensal. Brasília: BCB, 2026. Disponível em: https://www3.bcb.gov.br/sgspub. Acesso em: 26 set. 2026.

BATES, J. M.; GRANGER, C. W. J. The combination of forecasts. *Operational Research Quarterly*, v. 20, n. 4, p. 451-468, 1969.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time series analysis*: forecasting and control. 5. ed. Hoboken: John Wiley & Sons, 2015.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). *Indicadores de preços*: milho e soja. Piracicaba: ESALQ/USP, 2026. Disponível em: https://www.cepea.org.br. Acesso em: 26 set. 2026.

FEDERAL RESERVE BANK OF ST. LOUIS (FRED). *FRED Economic Data*: Global price of Soybeans (PSOYBUSDM) e Global price of Corn (PMAIZMTUSDM). Fonte original: International Monetary Fund, Primary Commodity Prices. St. Louis: FRED, 2026. Disponível em: https://fred.stlouisfed.org. Acesso em: 26 set. 2026.

GARDNER, E. S.; McKENZIE, E. Forecasting trends in time series. *Management Science*, v. 31, n. 10, p. 1237-1246, 1985.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting*: principles and practice. 3. ed. Melbourne: OTexts, 2021.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, v. 27, n. 3, p. 1-22, 2008.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, v. 22, n. 4, p. 679-688, 2006.

HYNDMAN, R. J.; KOEHLER, A. B.; SNYDER, R. D.; GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, v. 18, n. 3, p. 439-454, 2002.

KINGMA, D. P.; BA, J. Adam: a method for stochastic optimization. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 3., 2015, San Diego. *Proceedings* [...]. San Diego: ICLR, 2015.

MANOGNA, R. L.; DHARMAJI, V.; SARANG, S. Enhancing agricultural commodity price forecasting with deep learning. *Scientific Reports*, v. 15, art. 20903, 2025.

MARGARIDO, M. A.; TUROLLA, F. A.; BUENO, C. R. F. The world market for soybeans: price transmission into Brazil and effects from the timing of crop and trade. *Nova Economia*, v. 17, n. 2, p. 241-270, 2007.

MUNDLAK, Y.; LARSON, D. F. On the transmission of world agricultural prices. *The World Bank Economic Review*, v. 6, n. 3, p. 399-422, 1992.

PANKRATZ, A. *Forecasting with dynamic regression models*. New York: John Wiley & Sons, 1991.

PAUL, R. K. et al. Machine learning techniques for forecasting agricultural prices: a case of brinjal in Odisha, India. *PLoS ONE*, v. 17, n. 7, e0270553, 2022.

RIBEIRO, M. H. D. M.; COELHO, L. S. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing*, v. 86, 105837, 2020.

SEABOLD, S.; PERKTOLD, J. Statsmodels: econometric and statistical modeling with Python. In: PYTHON IN SCIENCE CONFERENCE, 9., 2010, Austin. *Proceedings* [...]. Austin: SciPy, 2010. p. 92-96.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, v. 15, n. 56, p. 1929-1958, 2014.

TASHMAN, L. J. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, v. 16, n. 4, p. 437-450, 2000.

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. *The American Statistician*, v. 72, n. 1, p. 37-45, 2018.

TIMMERMANN, A. Forecast combinations. In: ELLIOTT, G.; GRANGER, C. W. J.; TIMMERMANN, A. (ed.). *Handbook of economic forecasting*. Amsterdam: Elsevier, 2006. v. 1, p. 135-196.