Insper

Programa Institucional de Bolsas de Iniciação Científica (PIBIC)

Ciências Econômicas

Análise Econométrica do Comportamento do Minério de Ferro frente a Variáveis Macroeconômicas e de Mercado

Gabriel Nobrega Belisiario

Orientadora: Adriana Bruscato Bortoluzzo

São Paulo

2025

Programa Institucional de Bolsas de Iniciação Cientifica (PIBIC)

Gabriel Nobrega Belisiario
Ciências Econômicas

Análise Econométrica do Comportamento do Minério de Ferro frente a Variáveis Macroeconômicas e de Mercado

Relatório final apresentado ao Programa Institucional de Bolsas de Iniciação Cientifica (PIBIC)

Orientadora: Adriana Bruscato Bortoluzzo

São Paulo

2025

Resumo

Este estudo investiga o comportamento dinâmico dos preços do minério de ferro em resposta a variáveis macroeconômicas e de mercado relevantes. Combinando técnicas econométricas de previsão e modelagem de volatilidade, a pesquisa emprega análises de correlação cruzada, juntamente com modelos ARIMAX e ARIMAX-GARCH, para identificar relações temporais e efeitos de transmissão de volatilidade. A investigação abrange determinantes exógenos, como preços de commodities industriais, taxas de câmbio, taxas de juros e indicadores de atividade, avaliando sua significância e impactos defasados na formação do preço do minério de ferro. Os resultados indicam que os preços de metais industriais apresentam poder explicativo imediato e consistente, enquanto as commodities energéticas e variáveis macroeconômicas contribuem com efeitos mais defasados e menos robustos. Essas conclusões oferecem uma compreensão mais aprofundada sobre a formação de preços e a transmissão de volatilidade no mercado global de minério de ferro, trazendo implicações práticas para participantes do mercado e formuladores de políticas.

Palavras-chave: minério de ferro; econometria; ARIMAX; GARCH; volatilidade

Abstract

This study examines the dynamic behavior of iron ore prices in response to relevant macroeconomic and market variables. By integrating both econometric forecasting techniques and volatility modeling, the research employs cross-correlation analysis alongside ARIMAX and ARIMAX-GARCH models to unravel temporal relationships and volatility spillovers. The investigation spans exogenous drivers such as industrial commodity prices, exchange rates, interest rates, and activity indicators, evaluating their significance and lagged impacts on the denomination of iron ore prices. Findings highlight that industrial metal prices exhibit immediate and consistent explanatory power, whereas energy commodities and macroeconomic variables contribute with delayed and less robust effects. These insights contribute to a deeper understanding of price formation and volatility transmission in the global iron ore market, offering practical implications for market participants and policymakers.

Keywords: iron ore; econometrics; ARIMAX; GARCH; volatility.

Lista de Siglas

AUS – Austrália

AUD – Dólar Australiano

BR – Brasil

BRL – Real Brasileiro

EUR – Euro

USA – Estados Unidos da América

USD – Dólar Americano

CHY – Yuan Chinês (CNY)

BCI – Baltic Capesize Index

BDI – Baltic Dry Index

CPI – Consumer Price Index (Índice de Preços ao Consumidor)

IR – Interest Rate (Taxa de Juros)

MA – Moving Average (Média Móvel)

Sumário

1. Introdução 1
1.1 Objetivos do Estudo 1
2. Revisão de Literatura 3
2.1 Formação de Preços do Minério de Ferro 3
2.2 Aplicação de Modelos Econométricos 3
3. Base de Dados 5
3.1 Resultados Esperados dos Dados 6
4. Análise Descritiva 7
5. Análise de Relação Temporal 12
5.1 Análise de Correlação Cruzada 17
6. Metodologia Econométrica 23
6.1 Modelo ARIMAX 23
6.2 Modelo ARIMAX - GARCH 26
7. Considerações Finais 30
Referências 31
Apêndice A – Código R Markdown 32

1. Introdução

O minério de ferro é um dos pilares do comércio internacional e da produção industrial moderna. Presente em quase todas as cadeias produtivas, sua principal aplicação é na fabricação de aço - destino de cerca de 98% da produção global Esse aço, por sua vez, é insumo fundamental para setores como construção civil, transporte, infraestrutura e manufatura pesada, tornando o minério de ferro um elo central da economia mundial. Em 2023, o mercado global de minério de ferro foi estimado em US$ 279 bilhões, com projeção de atingir US$ 397 bilhões até 2032, o que representa um crescimento médio anual próximo de 4%.

A importância desse minério vai além do seu peso econômico. Ele é considerado um termômetro sensível da atividade global, refletindo com rapidez as mudanças na demanda, na oferta e nas expectativas do mercado. Movimentos no preço do minério muitas vezes antecipam tendências macroeconômicas, já que a sua demanda está fortemente ligada ao ritmo de crescimento de economias industrializadas e emergentes. Por exemplo, em 2025, mesmo diante de uma queda de 5,5% nas importações chinesas (o maior consumidor mundial) o preço se manteve acima de US$ 90 por tonelada, sustentado por uma combinação de demanda resiliente e choques de oferta.

Diante desse cenário, torna-se fundamental compreender o comportamento do preço do minério de ferro e suas interações com variáveis econômicas e de mercado. O objetivo deste trabalho é justamente analisar essas relações, investigando como indicadores macroeconômicos e oscilações de mercado se conectam às variações do minério. Essa compreensão pode oferecer informações valiosas tanto para agentes econômicos diretamente ligados ao setor mineral quanto para investidores e formuladores de políticas, que dependem dessas dinâmicas para antecipar tendências e tomar decisões estratégicas.

1.1 Objetivos do Estudo

O objetivo central deste trabalho é analisar o comportamento do preço do minério de ferro e compreender como ele reage a diferentes variáveis econômicas e de mercado. A proposta é identificar padrões e relações que possam explicar as

oscilações desse ativo, considerando seu contexto no comércio internacional e sua ligação com indicadores macroeconômicos e setoriais. Com isso, busca-se oferecer uma visão abrangente sobre os fatores que influenciam o minério de ferro, contribuindo para um melhor entendimento de sua dinâmica no cenário global.

Para alcançar esse objetivo, o estudo combina diferentes abordagens quantitativas, incluindo análises de correlação cruzada e modelos econométricos como ARIMAX e ARIMAX-GARCH. Essas ferramentas permitem identificar padrões de influência e intensidade nas relações entre o minério de ferro e variáveis-chave, tanto de caráter macroeconômico quanto de mercado. Dessa forma, é possível não apenas mapear as conexões existentes, mas também compreender como choques em diferentes frentes (como preços de commodities correlatas, taxas de juros, câmbio e índices de atividade) reverberam no comportamento do minério ao longo do tempo.

2. Revisão de Literatura

2.1 Formação de Preços do Minério de Ferro

Historicamente, a formação do preço do minério de ferro seguia o Sistema de Preço de Referência (Benchmark Pricing System), no qual mineradoras e siderúrgicas definiam anualmente um valor fixo para contratos de longo prazo. Esse formato assegurava previsibilidade às transações, mas limitava a capacidade de resposta às mudanças no mercado. A partir de 2010, esse sistema foi gradualmente substituído por negociações baseadas em preços spot e contratos futuros, impulsionado principalmente pelo crescimento acelerado da demanda chinesa e pela maior volatilidade nos mercados globais. Essa transição alterou profundamente a dinâmica de precificação, tornando os preços mais sensíveis a choques de oferta e demanda no curto prazo.

Nesse contexto, a hipótese de mercados eficientes (Fama, 1970) sustenta que os preços das commodities incorporam rapidamente todas as informações disponíveis, não apresentando memória. Entretanto, estudos como o de Wårell (2018) demonstram que o mercado de minério de ferro nem sempre segue esse padrão, uma vez que choques exógenos — como intervenções governamentais na China, crises financeiras globais ou disrupções logísticas — podem gerar distorções persistentes nos preços. Isso reforça a necessidade de utilizar abordagens econométricas capazes de captar essas variações dinâmicas e identificar como diferentes variáveis macroeconômicas e de mercado influenciam o comportamento do minério de ferro ao longo do tempo.

2.2 Aplicação de Modelos Econométricos

Diversos estudos evidenciam a aplicação de modelos de séries temporais, como ARIMA, na análise de preços de commodities. Huarnq, Yu e Hsu (2012) comparam a eficácia de modelos ARIMA, GARCH e híbridos na previsão de mercados de commodities, mostrando que abordagens combinadas podem melhorar a acurácia das projeções. Já Engle (1982) e Bollerslev (1986) introduziram as bases do GARCH e de sua generalização, permitindo capturar a heterocedasticidade condicional e a volatilidade persistente que são tão comum entre as series de preços de ativos e commodities. Ademais, Ghoshray (2011) reforça essa importância ao analisar a

volatilidade entre mercados futuros de commodities, destacando interdependências que tornam imprescindível a modelagem da variância.

No campo de previsões com variáveis exógenas, Alquist e Kilian (2010) exploram a informação contida nos preços futuros de petróleo bruto, utilizando abordagens econométricas para inferir expectativas de mercado e impactos macroeconômicos. Esse tipo de aplicação se aproxima do uso do ARIMAX, no qual variáveis externas são incorporadas para melhorar a capacidade preditiva. Pindyck e Rotemberg (1990), por sua vez, discutem o excesso de co-movimento entre preços de commodities, ressaltando a importância de considerar fatores comuns e choques globais nos modelos, algo que também pode ser incorporado via abordagens multivariadas.

Assim, a literatura indica que modelos como ARIMA, ARIMAX e GARCH são fundamentais para capturar tanto a dinâmica temporal dos preços de commodities quanto os efeitos de volatilidade e interdependência entre mercados. Essas ferramentas oferecem suporte robusto para análises que buscam entender não apenas movimentos passados, mas também antecipar reações do mercado frente a choques internos e externos.

3. Base de Dados

A seleção das variáveis utilizadas neste estudo foi guiada pela necessidade de compreender os múltiplos fatores que influenciam o comportamento do preço do minério de ferro no mercado global. As variáveis escolhidas abrangem preços de commodities (como carvão, gás natural, petróleo Brent, alumínio, cobre e soja), considerando que o minério de ferro integra cadeias produtivas nas quais esses insumos têm papel relevante, seja como substitutos, complementares ou indicadores indiretos de demanda. Essa abordagem permite avaliar como choques de oferta ou demanda em outras commodities podem impactar o minério de ferro.

Conforme apontam Pindyck e Rotemberg (1990), os preços das commodities frequentemente apresentam co-movimento excessivo devido a fatores macroeconômicos comuns, expectativas de mercado e interdependência setorial. Nesse sentido, foram incluídos preços de commodities como carvão, gás natural, petróleo Brent, alumínio, cobre e soja, que se relacionam direta ou indiretamente com a cadeia produtiva do aço e, portanto, com a demanda por minério de ferro. Esses dados foram combinados a indicadores macroeconômicos e setoriais para capturar uma visão mais abrangente do mercado, com atenção especial à seleção de variáveis para mitigar riscos de multicolinearidade no modelo.

Além disso, foram incorporados indicadores macroeconômicos e financeiros, como taxas de juros, câmbio e dados de atividade, de economias-chaves como Brasil, EUA, Austrália, União Europeia e China. A escolha desses países se deve à sua relevância na produção, exportação e consumo de minério de ferro, além da influência que exercem sobre o comércio e as finanças globais. Indicadores de atividade econômica, como produção industrial, vendas no varejo e mercado imobiliário, também foram incluídos para captar sinais de aquecimento ou desaquecimento das economias, que impactam diretamente a demanda por aço e, consequentemente, por minério de ferro.

Por fim, variáveis relacionadas ao setor de transporte marítimo, como o Baltic Dry Index (BDI) e o Baltic Capesize Index (BCI), foram consideradas devido à importância logística no comércio internacional do minério de ferro. O transporte marítimo representa a principal via de escoamento desse produto, sendo que, as oscilações nos custos de frete podem afetar tanto a competitividade quanto a formação de preços

no mercado global. Essa combinação de variáveis garante uma visão abrangente dos fatores estruturais e conjunturais que moldam a dinâmica do minério de ferro.

3.1 Resultados Esperados dos Dados

Ao analisar as variáveis da base de dados, espera-se que o carvão tenha relação positiva com o preço do minério de ferro, pois ambos são complementares na produção de aço. O gás natural pode ter efeito ambíguo: encarece custos industriais, mas também pode sinalizar atividade econômica elevada. O petróleo Brent provavelmente tem uma relação positiva por refletir tanto o ritmo da economia global quanto custos de transporte. Metais como alumínio e cobre devem ter relação positiva, pois compartilham a mesma base de demanda. A soja apresentaeia um efeito indireto e ambíguo, influenciando câmbio e termos de troca, mas espera-se que tenha uma baixa significância por ter um impacto indireto.

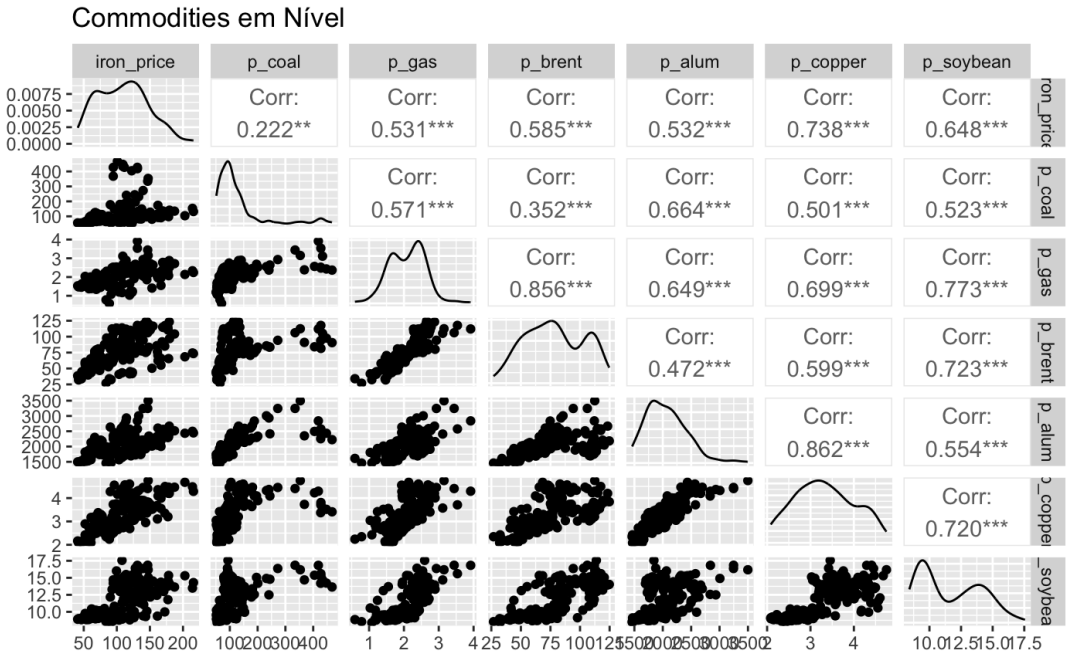
Entre as variáveis macroeconômicas, as taxas de juros tendem a ter relação negativa, já que juros mais altos reduzem atividade e demanda industrial. O câmbio pode ter um efeito ambíguo: moeda mais fraca favorece exportações, mas encarece insumos. A produção industrial e as vendas no varejo devem ter relação positiva, sinalizando expansão econômica. O mercado imobiliário também tende a ter relação positiva, pois a construção civil é intensiva em aço.

Por fim, o Baltic Dry Index (BDI) e o Baltic Capesize Index (BCI) devem ter relação positiva, já que refletem a demanda por transporte marítimo de commodities, sendo o BCI diretamente ligado ao transporte de minério de ferro. Altas nesses índices indicam maior atividade comercial internacional, o que tende a sustentar preços mais elevados do minério.

4. Análise Descritiva

Inicialmente, antes da aplicação dos modelos econométricos, foi realizada uma análise descritiva com o objetivo de identificar padrões, relações e possíveis colinearidades entre o preço do minério de ferro e variáveis macroeconômicas e de mercado selecionadas. Foram analisados cinco grandes grupos de fatores: preços de outras commodities, taxas de juros internacionais, taxas de câmbio, indicadores de atividade econômica e dados do mercado imobiliário.

Gráfico 1 – Análise de Correlação entre commodities e o minério de ferro, em nível



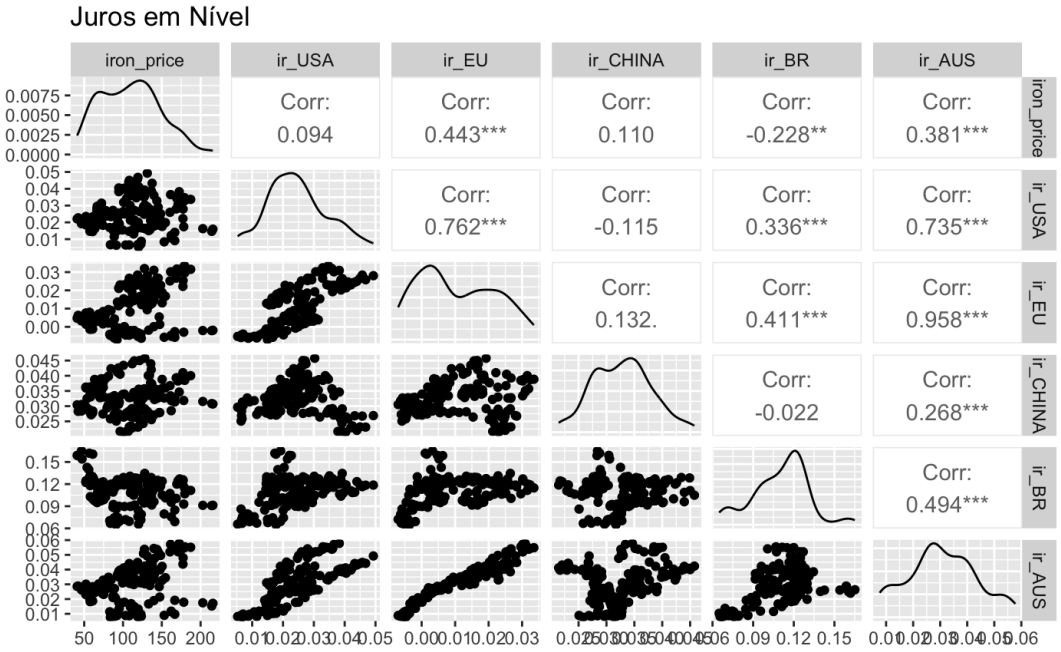
Fonte: Elaborado pelo autor

Analisando a matriz de correlação entre commodities e o preço spot do minério de ferro (Gráfico 1), pode-se observar que o preço do minério de ferro apresenta correlação positiva moderada com carvão (0,222), petróleo Brent (0,585) e gás natural (0,531), sugerindo que movimentos nos mercados energéticos tendem a estar associados às oscilações do minério. No entanto, as correlações mais fortes ocorrem com o cobre (0,738) e a soja (0,648), possivelmente refletindo fatores macroeconômicos globais comuns, como ciclos de demanda industrial e crescimento econômico, que afetam simultaneamente esses mercados.

Ademais, entre as demais commodities observa-se alta correlação entre petróleo Brent e gás natural (0,856), assim como entre cobre e alumínio (0,862),

indicando comportamento conjunto entre algumas commodities. Em suma, os padrões de dispersão sugerem relações predominantemente lineares, embora alguns pares indiquem heterocedasticidade visual, ou seja, volatilidade mais acentuada em determinados intervalos de preços. Esses resultados indicam que o minério de ferro está fortemente inserido em um contexto de preços globais interligados, e que choques em commodities correlacionadas, especialmente metais industriais, tendem a impactar diretamente seu comportamento.

Gráfico 2 – Análise de Correlação entre juros dos principais países e o minério de ferro, em nível



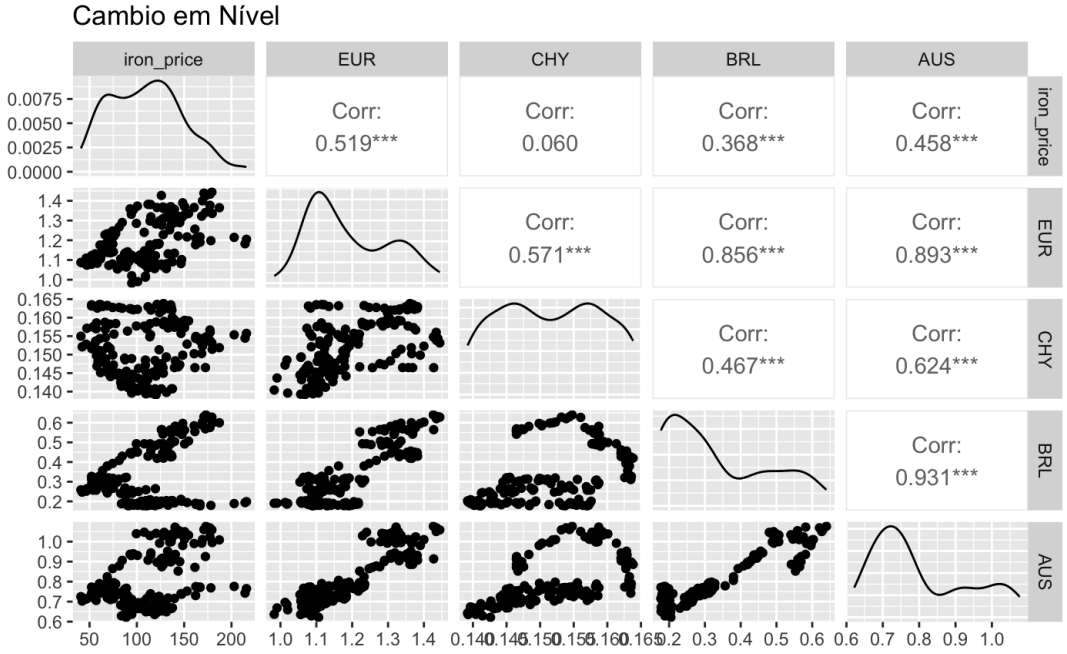
Fonte: Elaborado pelo autor

A matriz de correlação entre o preço do minério de ferro e diferentes taxas de juros (Gráfico 2) mostra relações mais fracas do que aquelas observadas com commodities, mas ainda sim podem ser relevantes em alguns casos. A maior correlação positiva é com a taxa de juros da União Europeia (0,443), seguida pela Austrália (0,381), sugerindo que políticas monetárias nessas regiões podem estar associadas a variações no preço do minério, possivelmente via impactos na demanda global e no comércio internacional. A relação com os juros dos EUA é praticamente nula (0,094), e com a China (0,110) também se mantém fraca, indicando que o preço do minério não reage de forma linear simples às taxas dessas economias. Já com o Brasil, observa-se correlação negativa (-0,228), possivelmente refletindo movimentos

opostos entre condições monetárias domésticas e o preço internacional da commodity.

Entre as próprias taxas de juros, as correlações são bastante altas em pares como EUA–UE (0,762), EUA–Austrália (0,735) e UE–Austrália (0,958), evidenciando alinhamento entre políticas monetárias de economias desenvolvidas. Também se nota que a taxa da China tem baixas correlações com os demais países, reforçando um padrão de independência relativa na condução de sua política monetária. Esses resultados indicam que, embora existam algumas conexões relevantes, especialmente com Europa e Austrália, o impacto direto de juros sobre o minério de ferro pode depender de outros fatores intermediários, como demanda global, crescimento industrial e dinâmica cambial.

Gráfico 3 – Análise de Correlação entre câmbio dos principais países e o minério de ferro, em nível



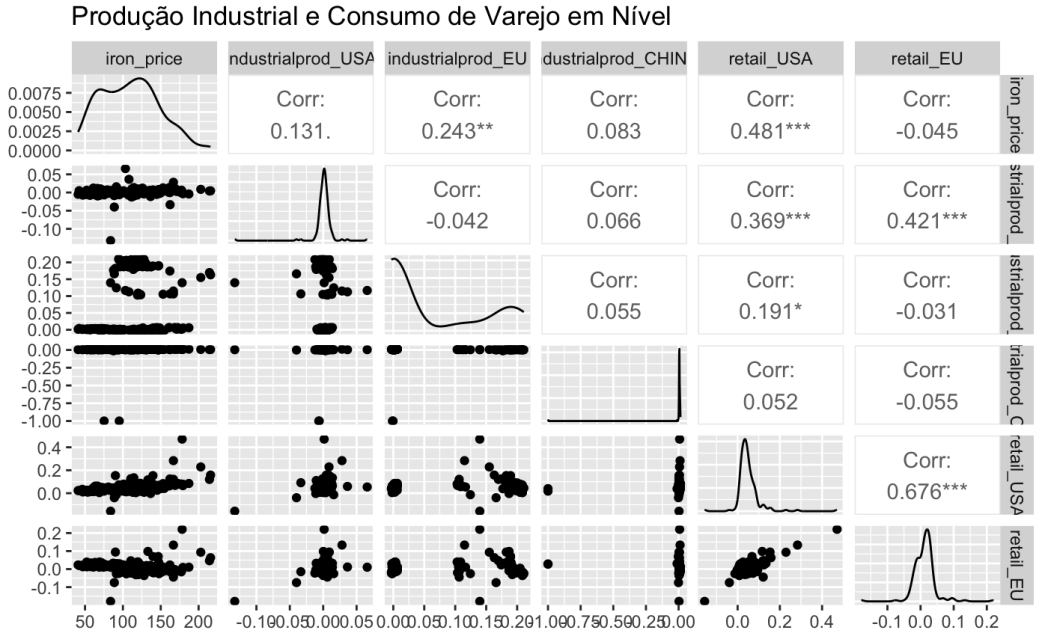
Fonte: Elaborado pelo autor

Ao observar-se a correlação cambial com o minério de ferro (Gráfico 3), pode-se analisar que o preço do minério de ferro mantém relação positiva moderada com o euro (0,519), o dólar australiano (0,458) e o real brasileiro (0,368), sugerindo que as moedas de países exportadores (Brasil e Australia) acompanham variações do minério, possivelmente por estarem associadas ao desempenho do setor mineral e a fluxos de comércio internacional. Já o yuan chinês apresenta correlação praticamente

nula (0,060), indicando que, no período analisado, o câmbio da China não se moveu de forma relevante com o minério, possivelmente devido ao controle cambial do país.

Ademais, entre as moedas observam-se correlações muito elevadas, como EUR–AUS (0,893) e BRL–AUS (0,931), que refletem movimentos comuns ligados ao ciclo de commodities e ao apetite global por risco. A forte correlação EUR–BRL (0,856) e a moderada CHY–BRL (0,467) indicam que moedas de países emergentes exportadores tendem a se mover em sintonia com moedas fortes em determinados contextos, especialmente quando influenciadas por preços internacionais de commodities. Esses resultados reforçam a importância de considerar variáveis cambiais no modelo de previsão de preços do minério.

Gráfico 4 – Análise de Correlação entre dados de atividade e o minério de ferro, em nível



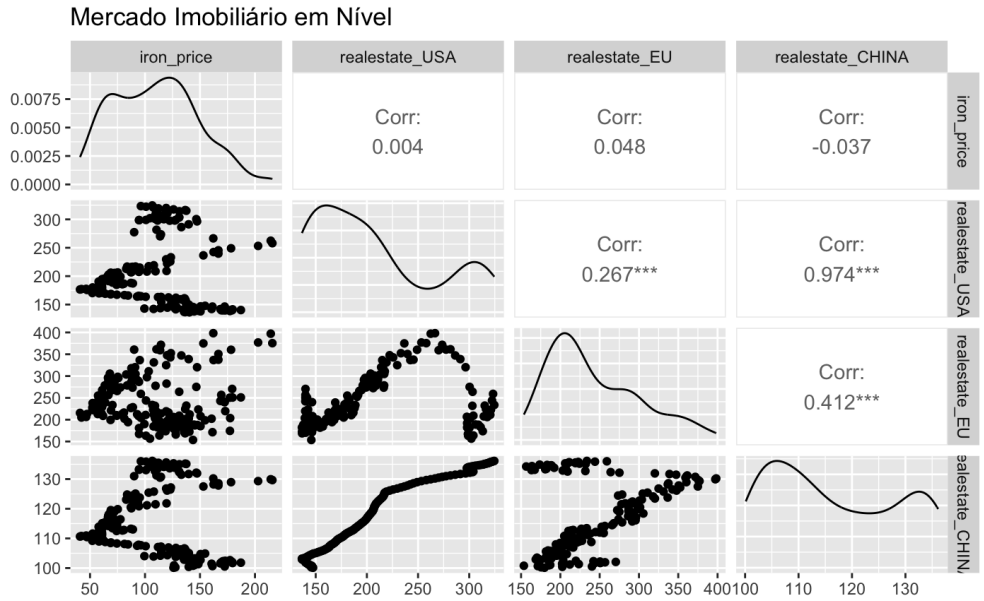
Fonte: Elaborado pelo autor

No gráfico acima (Gráfico 4), pode-se análisar que o preço do minério de ferro apresenta correlação positiva moderada com o consumo de varejo nos EUA (0,481), sugerindo que períodos de maior atividade de consumo no país tendem a estar associados a preços mais altos da commodity, possivelmente refletindo maior demanda industrial indireta. Também há correlação positiva, embora mais fraca, com a produção industrial da União Europeia (0,243), enquanto a produção industrial dos EUA (0,131) e da China (0,083) não demonstram relação estatisticamente relevante,

indicando que, no período analisado, o preço do minério não respondeu de forma direta à produção industrial desses países.

Entre as variáveis macroeconômicas, destacam-se correlações significativas entre varejo e produção industrial dentro das mesmas regiões. Por exemplo, varejo dos EUA com produção industrial dos EUA (0,369) e com varejo da UE (0,421), e varejo da UE com produção industrial da UE (0,676). Esses resultados sugerem que o comportamento de consumo e produção está fortemente alinhado nas economias desenvolvidas, mas a ligação direta com o preço do minério é mais evidente no caso do consumo norte-americano.

Gráfico 5 – Análise de Correlação entre dados de mercado imobiliário e o minério de ferro, em nível



Fonte: Elaborado pelo autor

Dando continuidade à análise descritiva, os dados de correlação do mercados imobiliários (Gráfico 5) indicam que o preço do minério de ferro não apresenta correlação estatisticamente significativa com os índices do mercado imobiliário dos EUA (0,004), da União Europeia (0,048) ou da China (-0,037). Isso sugere que, no período analisado, as oscilações no setor imobiliário dessas regiões não tiveram relação direta com a variação do preço da commodity, possivelmente porque a demanda por minério de ferro esteve mais ligada a outros setores industriais e ao comércio internacional.

Outra justificativa plausível para a ausência de relação significativa é que o mercado imobiliário, apesar de demandar aço e cimento (minério de ferro indiretamente), representa apenas uma fração da demanda global pela commodity, sendo superado por setores como infraestrutura pesada, indústria automotiva e produção de bens de capital. Além disso, fatores como estoques estratégicos, contratos de longo prazo e políticas de importação/exportação (especialmente no caso da China) podem suavizar o impacto de flutuações no setor imobiliário sobre o preço internacional do minério.

Ademais, entre os próprios mercados imobiliários, observa-se uma correlação extremamente forte entre EUA e China (0,974), e moderada-alta entre UE e China (0,412) e entre UE e EUA (0,267). Esses resultados indicam um alto grau de co-movimento entre as atividades imobiliárias dessas regiões, possivelmente refletindo tendências globais de crédito, taxas de juros e condições macroeconômicas comuns, mas sem que isso se traduza, de forma direta, em impacto sobre o preço do minério no intervalo estudado.

5. Análise de Relação Temporal

Tabela 2 – Teste ADF para o preço do Minério de Ferro em Nível e em Log-Dif

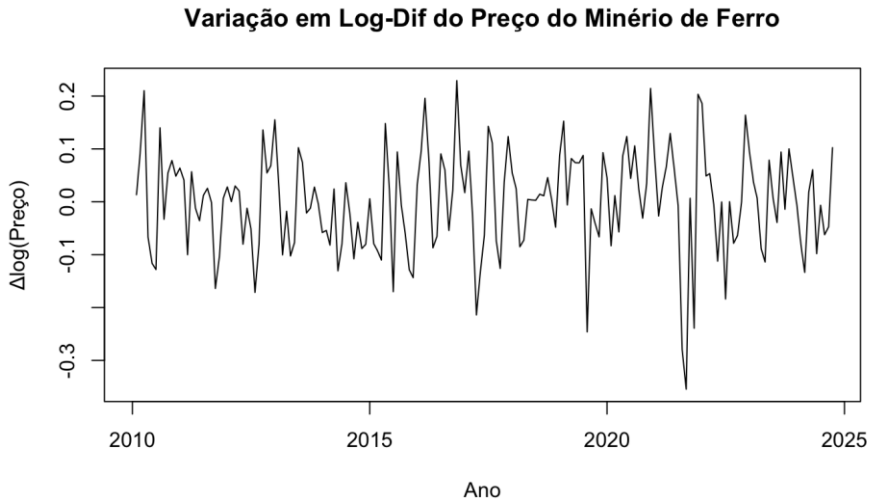
Variável	**Estatística de Teste**	**Valor crítico 1%**	**Valor crítico 5%**	**Valor crítico 10%**	**P-valor**	**Conclusão**
Preço em nível	-2,61	-3,46	-2,88	-2,57	0,00972	Não rejeita H0 (não estacionário)
Log-diferença do preço	-9,40	-3,46	-2,88	-2,57	< 2e-16	Rejeita H0 (estacionário)

Fonte: Elaborado pelo autor

Primeiramente, foi feito o teste Augmented Dickey Fuller (ADF) da série temporal do preço do minério de ferro a partir do preço em nível (Tabela 2), visando avaliar a estacionariedade e a presença de padrões sazonais. Inicialmente, observa-se que teste de raiz unitária Augmented Dickey-Fuller (ADF). Os resultados indicam que a série em nível não é estacionária, pois a estatística do teste (-2.7) não é menor que os valores críticos de 5% e 1%, o que impede a rejeição da hipótese nula de raiz

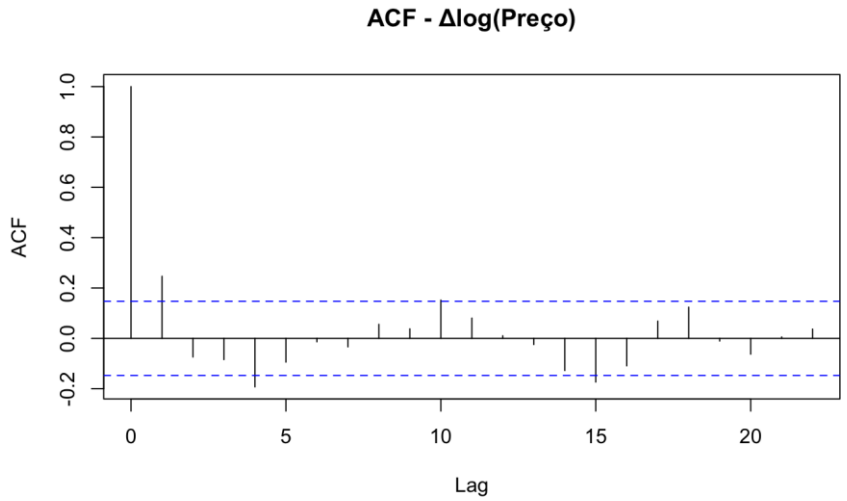
unitária. No entanto, ao aplicar o teste na série transformada pela log-diferença, a estatística do teste foi de -9.40, inferior a todos os valores críticos. Isso permite rejeitar a hipótese nula e concluir que a série log-diferenciada é estacionária, ou seja, integrada de ordem 1 (I(1))

Gráfico 6 – Variação em Log-Diferença do preço do minério de ferro



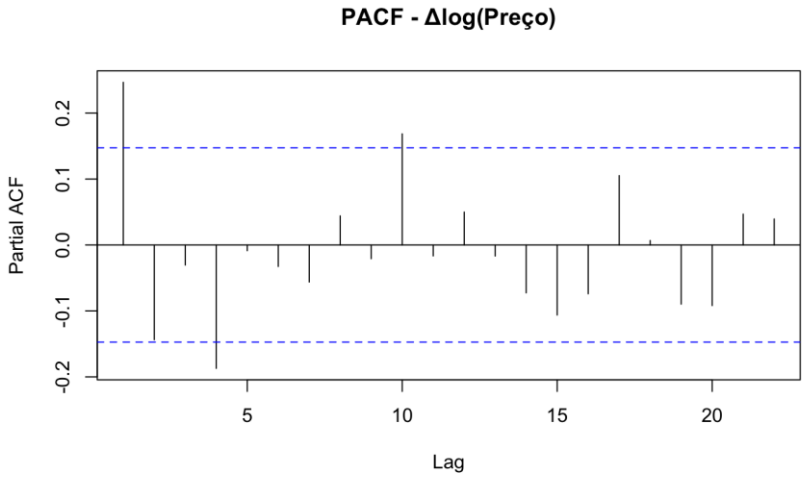
Fonte: Elaborado pelo autor

Gráfico 7 – Função de Autocorrelação (ACF) da Variação Logarítmica do Preço do Minério de Ferro



Fonte: Elaborado pelo autor

Gráfico 8 – Função de Autocorrelação Parcial (PACF) da Variação Logarítmica do Preço do Minério de Ferro



Fonte: Elaborado pelo autor

A série da variação logarítmica do preço do minério de ferro, observada entre 2010 e 2025, não apresenta tendência ou sazonalidade visível, e oscila ao redor de uma média próxima de zero com volatilidade relativamente estável (Gráfico 6). Ademais o gráfico da ACF (Gráfico 7) mostra autocorrelação significativa apenas no lag 1, com dissipação imediata nos lags seguintes, enquanto a PACF (Gráfico 8) reforça esse padrão com baixa significância nos demais lags. Esses resultados indicam que a série possui baixa memória temporal e é bem representada por um modelo de média móvel de ordem 1 (MA(1)).

Na prática, isso significa que os preços do minério de ferro reagem rapidamente a choques pontuais e tais efeitos não se prolongam no tempo. Isso é característico de mercados globais de commodities metálicas (Pindyck, 1999), que argumenta que esses ativos tendem a seguir processos estocásticos simples, com pouca previsibilidade e ajustes rápidos a novas informações.

Tabela 3 – Modelo ARIMA (0,1,1)

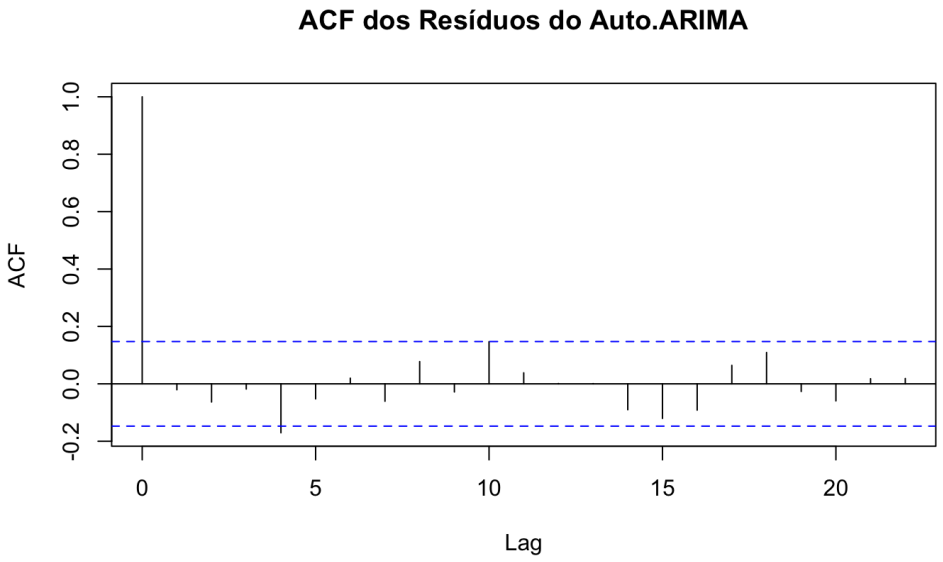
Parâmetro	Valor
ma1	0,31
sigma^2	0,008
log likelihood	171,02
AIC	-338,04
AICc	-337,97
BIC	-331,69

Fonte: Elaborado pelo autor

Com o objetivo de identificar a melhor especificação de modelo ARIMA para a série log-diferenciada do preço, foi implementado um processo de seleção automática de modelo, com base em critérios de informação, como o Akaike Information Criterion (AIC), sua versão corrigida (AICc) e o Bayesian Information Criterion (BIC). Em suma, o modelo selecionado foi um ARIMA(0,0,1), ou seja, um processo de média móvel de ordem 1 (MA(1)), aplicado sobre a série já transformada por log-diferença. O coeficiente estimado para o parâmetro ma1 foi 0.31 (Tabela 3), com erro padrão de 0.0784, sendo estatisticamente significativo. O valor do log-likelihood (171.02), juntamente com AIC (-338.04), AICc (-337.97) e BIC (-331.69) bastante baixos, confirmam a qualidade do ajuste e sustentam a escolha do modelo como ótima entre as alternativas consideradas. A seleção automática priorizou esse modelo justamente por apresentar o menor valor de AIC, critério que busca maximizar a qualidade do ajuste penalizando o excesso de parâmetros.

Sendo assim, com base na teoria estatística e nos resultados obtidos, conclui-se que o modelo ARIMA(0,0,1) representa uma especificação adequada, simples e robusta para descrever a dinâmica de curto prazo da variação logarítmica do preço do minério de ferro.

Gráfico 9 – ACF dos Resíduos do Auto Arima MA(1) para a difereça Logaritmica do Minério de Ferro



Fonte: Elaborado pelo autor

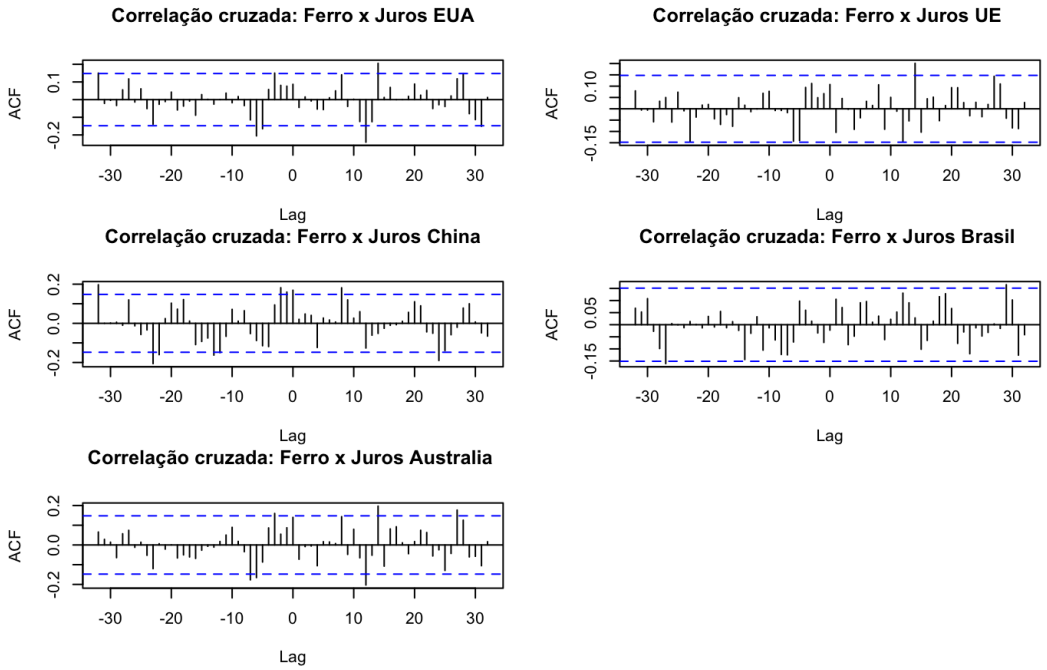
O gráfico da função de autocorrelação dos resíduos do modelo Auto.ARIMA (Gráfico 9) mostra que praticamente todos os coeficientes de autocorrelação estão dentro dos limites de significância com exceção do lag 0, que reflete a variância dos próprios resíduos e não indica autocorrelação. A ausência de autocorrelação significativa nos resíduos confirma que o modelo ARIMA ajustado capturou adequadamente a estrutura temporal da série, e que os resíduos se comportam como ruído branco. Esse resultado valida a adequação do modelo, conforme esperado em uma modelagem ARIMA bem especificada, em linha com as recomendações metodológicas da literatura de séries temporais.

O comportamento capturado por um MA(1) revela que as variações no preço são essencialmente aleatórias, com apenas uma leve dependência do erro observado no período anterior. Isso está alinhado com o funcionamento típico de mercados de commodities metálicas, como o minério de ferro, que operam de forma contínua e são influenciados por informações frequentemente imprevisíveis, como por decisões de grandes importadores (por exemplo, a China), greves em mineradoras, ou eventos climáticos que afetam a logística. Esses choques afetam o preço no curto prazo, mas rapidamente se dissipam, o que impede a formação de padrões duradouros.

Ademais, é importante ressaltar que essa especificação econométrica é esperada, já que os preços de commodities naturais tendem a seguir caminhadas aleatórias ou processos estacionários de baixa ordem, pois são determinados em mercados muito líquidos e eficientes, onde choques de oferta e demanda são rapidamente incorporados aos preços (Pindyck, 1999). Especificamente, ele observa que muitos preços de commodities exibem pouca previsibilidade fora dos movimentos aleatórios de curto prazo, o que é compatível com um modelo MA(1) como no caso do minério de ferro.

5.1 Análise de Correlação Cruzada

Gráfico 10 – Correlação Cruzada entre Minério de Ferro e juros de 5 países selecionados
3pm



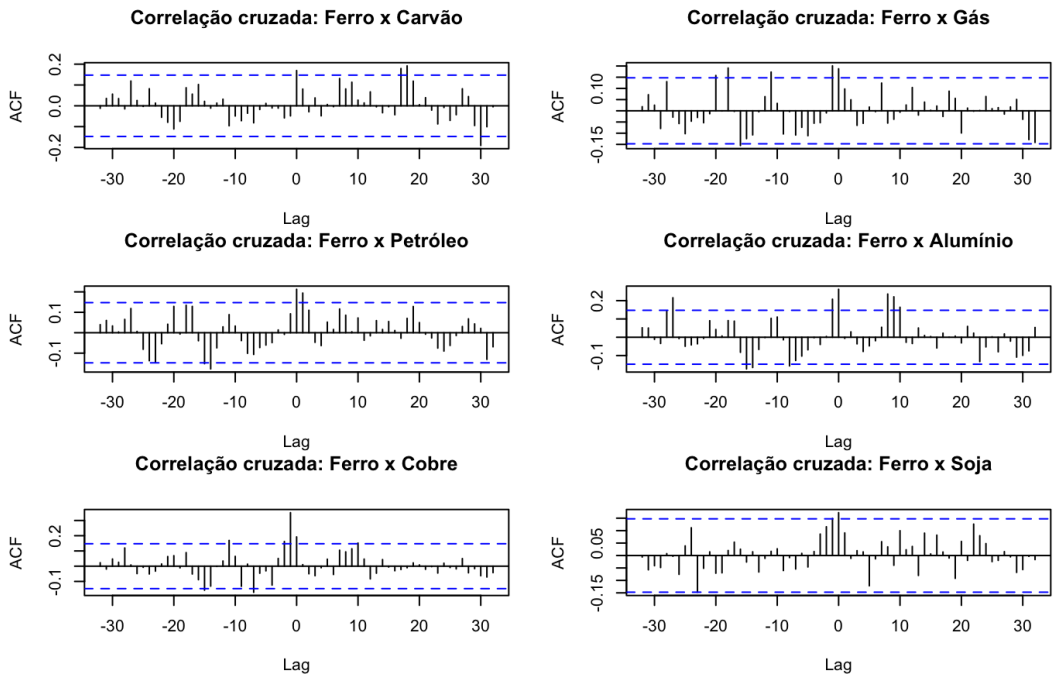
Fonte: Elaborado pelo autor

As correlações cruzadas (Gráfico 10) indicam que, para EUA e Austrália, há picos positivos significativos em lags positivos, o que sugere que aumentos nas taxas de juros tendem a ser seguidos por elevações no preço do minério de ferro. Também há momentos de correlação negativa, apontando que, em alguns períodos, altas nos juros podem estar associadas a quedas no preço do minério. Já ao analisar a China, observa-se padrão semelhante, com correlações positivas relevantes em lags positivos, coerente com o peso do minério de ferro para sua indústria e possíveis

efeitos de política monetária sobre o mercado de commodities. Já na União Europeia e no Brasil, as correlações são mais fracas e menos frequentes, sugerindo que o impacto direto das taxas de juros sobre o preço do minério é menos consistente nessas economias.

Também há momentos em que o minério de ferro se move antes dos juros, especialmente em alguns países mais ligados ao mercado de commodities. Isso sugere que, em certos contextos, o preço do minério pode funcionar como um sinal antecipado das condições monetárias. Mesmo assim, o padrão geral é que as mudanças nos juros costumam ter um impacto mais consistente sobre o minério do que o contrário.

Gráfico 11 – Correlação Cruzada entre Minério de Ferro e 6 Commodities selecionadas



Fonte: Elaborado pelo autor

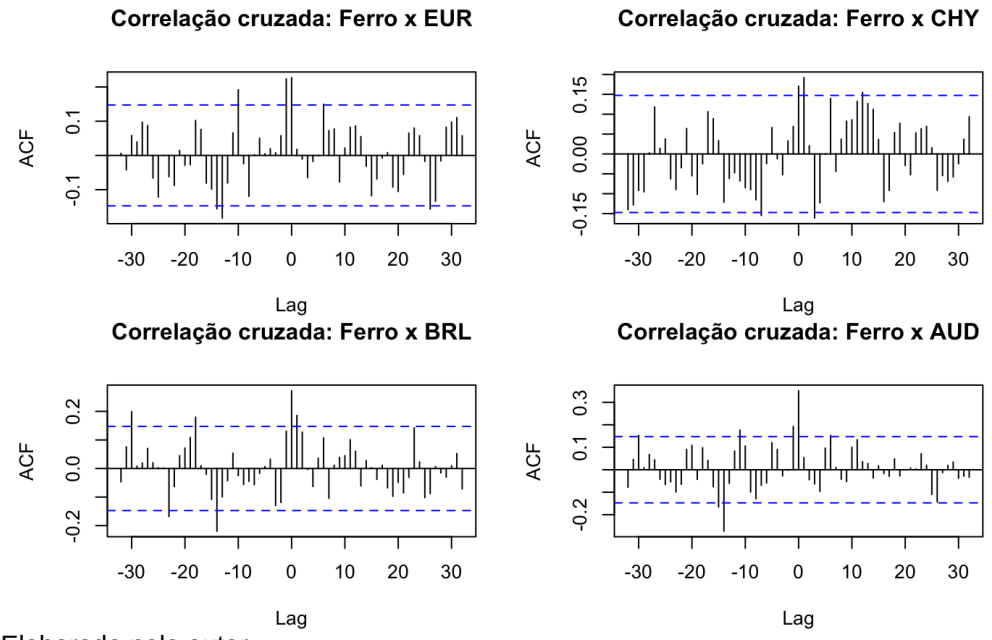
Com base nos gráficos de correlação cruzada, a análise entre o minério de ferro e outras commodities revela que algumas apresentam relação mais imediata e intensa com o preço do minério, enquanto outras demonstram efeitos mais defasados ou fracos. O cobre, por exemplo, exibe correlações positivas relevantes em lags curtos, sugerindo que movimentos no preço do cobre tendem a ser acompanhados rapidamente por movimentos semelhantes no minério de ferro, possivelmente por

ambos estarem ligados à demanda global por metais industriais. O alumínio e o carvão também mostram padrões consistentes de correlação, ainda que com intensidade um pouco menor.

No caso do petróleo Brent e do gás natural, as correlações aparecem de forma menos direta e, em alguns casos, mais defasada. Isso indica que sua influência sobre o minério de ferro pode ocorrer de maneira indireta, por meio de custos de produção, transporte ou condições macroeconômicas mais amplas que afetam o setor mineral. Em alguns lags positivos, observa-se que aumentos nesses preços podem anteceder altas no minério de ferro, sugerindo que mudanças no custo energético têm potencial de se refletir nos preços do minério com certo atraso.

Por fim, a soja, representando commodities agrícolas, apresenta uma relação mais fraca e menos consistente com o minério de ferro. Isso reforça a ideia de que o minério de ferro é mais sensível a commodities ligadas à indústria e à construção civil do que a produtos agrícolas. De forma geral, os resultados indicam que seu comportamento está mais sincronizado com mercados de metais básicos e, em menor grau, com commodities energéticas, sendo essas conexões relevantes para compreender sua dinâmica de preços.

Gráfico 11 – Correlação Cruzada entre Minério de Ferro e Taxa de Câmabio de 4 selecionadas (Modela Local / Dolar)



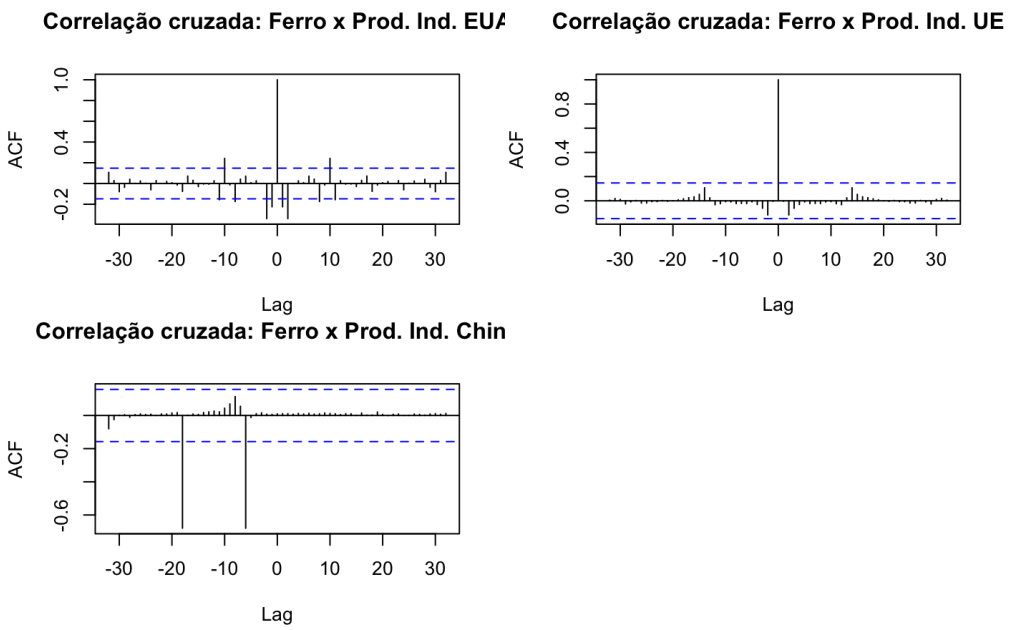
Fonte: Elaborado pelo autor

Ademais, ao análisar os gráficos de correlação cruzada (Gráfico 11) entre o preço do minério de ferro e as taxas de câmbio (EUR, CNY, BRL e AUD), em todos os casos lags positivos indicam que movimentos no câmbio precedem variações no preço do minério, enquanto lags negativos sugerem que o minério tende a influenciar o câmbio.

No caso do euro (EUR), há correlações positivas significativas em lags curtos e médios positivos, sugerindo que uma valorização do euro em relação ao dólar pode anteceder altas no preço do minério. Para o yuan (CNY), o padrão é mais disperso, com correlações positivas em alguns lags positivos, o que sugere que a moeda chinesa pode ter algum papel antecedente nas oscilações do minério, mas de forma menos consistente.

Já o real brasileiro (BRL) apresenta correlações positivas mais evidentes em lags curtos positivos, reforçando a hipótese de que a valorização do real pode estar associada a aumentos posteriores no preço do minério. Já o dólar australiano (AUD) exibe um pico marcante de correlação positiva em lag positivo próximo de zero, coerente com o papel da Austrália como um dos maiores exportadores de minério de ferro, onde oscilações cambiais estão fortemente ligadas à dinâmica do mercado do minério.

Gráfico 12 – Correlação Cruzada entre Minério de Ferro e Produção Industrial



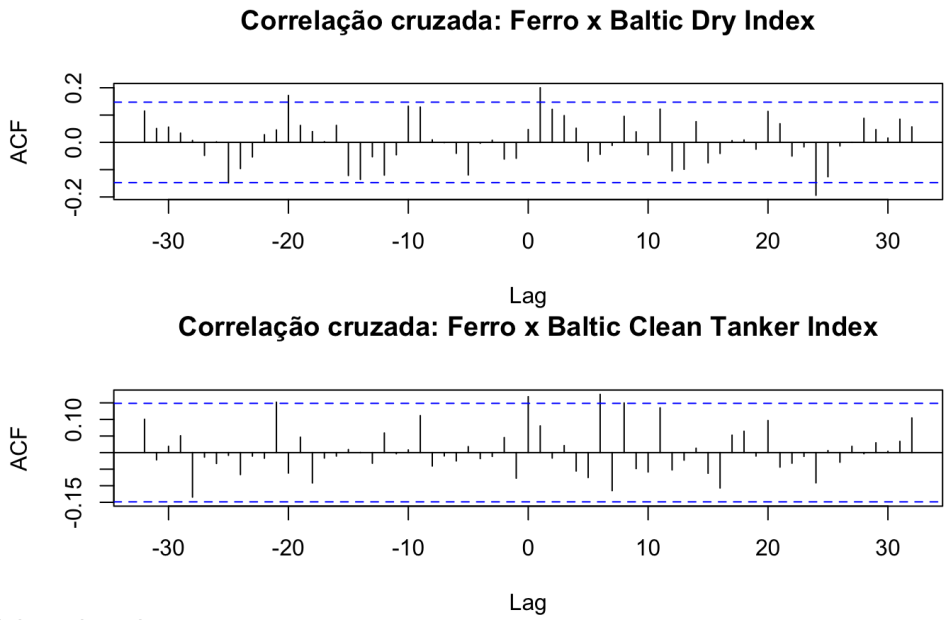
Fonte: Elaborado pelo autor

Ao análisar a correlação cruzada entre o preço do minério de ferro e a produção industrial mostra que, para os EUA, pode-se observar que há alguns picos de correlação positiva em lags positivos curtos, sugerindo que aumentos na produção industrial podem anteceder elevações no preço do minério. No entanto, o padrão não é forte ou consistente, indicando que essa relação pode ser pontual.

Na União Europeia, os gráficos não mostram correlações significativas em lags positivos, sugerindo que a produção industrial do bloco não apresenta relação clara de antecedência sobre o preço do minério de ferro. A maior parte das barras está dentro do intervalo de confiança, reforçando a ausência de efeito estatisticamente relevante.

Já no caso da China, é importante ressaltar que há correlações negativas relevantes em lags negativos mais longos, indicando que movimentos no preço do minério de ferro podem preceder quedas na produção industrial chinesa, ao invés do contrário. Não se observam sinais relevantes em lags positivos, o que aponta para uma relação de influência mais no sentido do minério afetando a produção, e não o oposto.

Gráfico 13 – Correlação Cruzada entre Minério de Ferro e Índices de Transporte



Fonte: Elaborado pelo autor

A análise da correlação cruzada entre o minério de ferro e o Baltic Dry Index (BDI) indica que os sinais mais relevantes ocorrem em lags negativos entre aproximadamente -10 e -25. Isso sugere que variações no preço do minério de ferro tendem a anteceder movimentos no BDI, o que é coerente, já que o índice reflete o custo do frete marítimo de commodities a granel, sendo o minério um dos principais itens transportados. Há também sinais positivos modestos em lags próximos de zero, indicando efeitos pontuais de resposta do minério a mudanças no BDI no curto prazo. Já no caso do Baltic Clean Tanker Index (BCTI), as correlações são mais fracas e menos consistentes. Observa-se alguns picos positivos em lags curtos, sugerindo que variações no transporte de derivados de petróleo podem ter algum efeito limitado sobre o minério de ferro. No entanto, a alternância entre correlações positivas e negativas reforça que a relação entre essas séries é menos estável e provavelmente influenciada por fatores externos não capturados diretamente pelo índice.

6. Metodologia Econométrica

Nesta seção, são apresentados os procedimentos econométricos utilizados para analisar a dinâmica dos retornos mensais do preço do minério de ferro. A estratégia parte de um modelo ARIMAX, que incorpora variáveis macroeconômicas e financeiras externas como determinantes dos preços, e evolui para uma modelagem mais robusta com a inclusão da variância condicional via modelos GARCH. O objetivo é capturar não apenas os fatores que influenciam a trajetória esperada do minério de ferro, mas também a natureza volátil de seu comportamento ao longo do tempo, permitindo uma compreensão mais precisa da interação entre choques macroeconômicos globais e o mercado de commodities metálicas.

6.1 Modelo ARIMAX

O ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) é uma extensão do modelo ARIMA tradicional que, além de considerar componentes autorregressivos (AR), de média móvel (MA) e diferenciação (I) para capturar a dinâmica temporal e tornar a série estacionária, incorpora variáveis exógenas, que representam fatores externos que podem influenciar diretamente a variável dependente. Sendo assim, o ARIMAX permite modelar o comportamento de uma série temporal considerando não apenas seu próprio histórico, mas também o impacto de variáveis explicativas externas.

Inicialmente, parti de um modelo amplo para não correr o risco de excluir, a priori, determinantes relevantes do preço do minério. Agrupei as explicativas em cinco blocos: câmbio, atividade, juros, commodities e índices de transporte. Para capturar defasagens de transmissão e reduzir simultaneidade, gerei lags de 0 a 3 meses para cada série, e utilizei como variável dependente o $\Delta\log$ do preço do minério (retorno mensal), o que torna a série aproximadamente estacionária e elimina a necessidade de termo constante.

O modelo foi estimado como ARIMAX(0,0,1), sendo que a escolha do MA(1) segue a literatura para retornos de commodities (ruído com dependência curta) e análise prévias de ACF/PACF. Ademais, o modelo identificou 49 variáveis significativas, abrangendo choques contemporâneos e defasados de preços de commodities (carvão, cobre, alumínio, petróleo Brent e gás), taxas de juros (Austrália,

Brasil, China e EUA), indicadores de atividade (produção industrial e vendas no varejo nos EUA e UE), além de taxas de câmbio de diversas moedas (BRL, EUR, CHY, AUS) e variáveis do setor imobiliário. Sendo assim, a presença de múltiplas defasagens mostra que o preço do minério de ferro responde tanto a movimentos imediatos quanto a efeitos acumulados ao longo do tempo. Por outro lado, algumas categorias inicialmente incluídas no modelo, como os índice de transporte, variáveis de mercado imobiliário nos EUA e algumas moedas, não apresentaram significância estatística.

Dando continuidade ao modelo, foram feitos repetidos modelos ARIMAX (0,0,1), removendo todas que não tinham significância relevante. Sendo que a única exceção foi o preço do carvão que foi mantido devido a sua importância teórica como matéria prima para o aço, e influenciando diretamente o preço do minério de ferro.

Tabela 3 - Resultados do Modelo ARIMAX para o Preço do Minério de Ferro (MA(1))

parametro	coef	se	t	p_value
log_diff_p_copper_lag1	0,03	0,01	4,21	0,00
log_diff_AUS_lag0	0,02	0,01	2,64	0,01
log_diff_p_copper_lag2	0,01	0,01	2,34	0,02
diff_ir_AUS_lag2	0,01	0,01	2,29	0,02
ma1	0,18	0,09	2,15	0,03
diff_ir_USA_lag3	0,01	0,01	1,92	0,06
log_diff_p_copper_lag0	0,01	0,01	1,90	0,06
diff_ir_CHINA_lag3	-0,01	0,01	-1,84	0,07
log_diff_p_coal_lag0	0,01	0,01	1,58	0,12

Fonte: Elaborado pelo autor

Tabela 4 - Testes Diagnósticos do Modelo ARIMAX

Teste	p_value
Ljung-Box (12)	0,5801
Ljung-Box (24)	0,6321
Shapiro-Wilk	0,03902
Jarque-Bera	0,001278
Breusch-Pagan	0,8681
ARCH-LM (12)	0,003078

Fonte: Elaborado pelo autor

Tabela 5 - Índice de Inflação Variância (VIF) das Variáveis Explicativas

Variável	VIF
log_diff_p_copper_lag1	1,185602
log_diff_p_copper_lag0	1,279241
log_diff_p_copper_lag2	1,044023
log_diff_p_coal_lag0	1,03274
log_diff_AUS_lag0	1,351832
diff_ir_AUS_lag2	1,027244
diff_ir_USA_lag3	1,085593
diff_ir_CHINA_lag3	1,078694

Fonte: Elaborado pelo autor

Os testes diagnósticos do modelo ARIMAX (Tabela 4) indicam que os resíduos não apresentam autocorrelação, conforme os resultados do Ljung-Box com 12 e 24 defasagens (p-valores de 0,5801 e 0,6321). Isso sugere que o modelo é capaz de capturar a dinâmica temporal dos dados. No entanto, os testes de normalidade, Shapiro-Wilk e Jarque-Bera, revelam p-valores de 0,03902 e 0,001278, respectivamente, indicando que os resíduos não seguem uma distribuição normal, o que pode afetar a validade dos intervalos de confiança e testes de hipóteses tradicionais.

O teste de Breusch-Pagan não identificou heterocedasticidade geral (p-valor de 0,8681), mas o teste ARCH-LM com 12 defasagens mostrou forte evidência de heterocedasticidade condicional (p-valor de 0,003078). Isso significa que a variância dos resíduos muda ao longo do tempo, comportamento típico de séries econômicas e financeiras, e que não pode ser tratado adequadamente apenas com o modelo ARIMAX.

Além disso, os valores de VIF de todas as variáveis explicativas (Tabela 5) estão abaixo de 1,4, indicando ausência de multicolinearidade relevante. Com isso, os coeficientes estimados são estáveis e não há necessidade de ajustes nas variáveis por esse motivo.

Diante disso, torna-se necessária a utilização de um modelo ARIMAX-GARCH, que além de modelar a média condicional, também modela a variância condicional dos retornos. Isso permite maior precisão nas previsões e maior robustez nos testes estatísticos, além de capturar adequadamente a natureza volátil dos preços de commodities. Estudos como o de Alquist e Kilian (2010) destacam a importância de

modelar a volatilidade em mercados de commodities, e outros como Huarng et al. (2012) aplicam explicitamente modelos ARIMAX-GARCH em commodities para melhorar previsões e inferências.

6.2 Modelo ARIMAX - GARCH

O modelo ARIMAX-GARCH combina duas abordagens: a modelagem da média com um componente autorregressivo de médias móveis com variáveis exógenas (ARIMAX), e a modelagem da variância com um processo GARCH. No ARIMAX, a série temporal do ativo é explicada por seus próprios valores passados (autoregressivo), por choques passados (média móvel) e por variáveis externas relevantes, como preços de commodities ou taxas de juros. Já o GARCH (Generalized Autoregressive Conditional Heteroskedasticity) é aplicado sobre os resíduos desse modelo de média e permite capturar a volatilidade condicional, ou seja, a variância dos erros que muda ao longo do tempo em função de choques anteriores. Esse tipo de modelo é muito utilizado em finanças e mercados de commodities porque reflete o comportamento mais realista de volatilidade, períodos de alta ou baixa volatilidade que se agrupam, e permite entender não apenas a direção esperada da série, mas também o risco associado a ela.

Inicialmente foi necessário identificar a distribuição do modelo GARCH a ser utilizado, com o **eGARCH com distribuição t-student padrão** (std_eGARCH) apresentando o melhor desempenho em termos de equilíbrio entre qualidade de ajuste (AIC) e diagnóstico dos resíduos (teste de Ljung-Box), superando outros modelos. Isso sugere que esse modelo consegue capturar de forma mais eficiente a heterocedasticidade condicional presente na série de preços do minério, lidando bem com assimetrias e choques nos dados.

Além disso, os modelos que apresentam menor persistência na volatilidade, ou seja, menor influência de choques passados na variância futura, foram favorecidos na pontuação final (score), o que é consistente com um mercado onde os choques não perpetuam longamente a volatilidade. Modelos mais rígidos, como o iGARCH, ficaram em último lugar, indicando que a suposição de persistência total da volatilidade não é adequada para esse mercado de commodities específico. Isso reforça a necessidade de flexibilidade na modelagem da variância condicional.

Tabela 6 - Teste de Viés de Sinal (Sign Bias) para Assimetria na Volatilidade

	t-value	prob
Sign Bias	1,88	0,06
Negative Sign Bias	0,47	0,64
Positive Sign Bias	1,57	0,11
Joint Effect	3,10	0,26

Fonte: Elaborado pelo autor

Ao analisar a os resultados do teste de viés de sinal (Tabela 6) aplicado aos resíduos padronizados de um modelo GARCH, o único resultado marginalmente significativo é o “Sign Bias” (t = 1.88, p = 0.061), indicando uma leve assimetria na resposta do modelo a choques positivos em relação à variância condicional. Já os testes de “Negative Sign Bias” e “Positive Sign Bias” não são significativos (p > 0.10), e o “Joint Effect” tampouco é (p = 0.26), sugerindo que, de forma geral, o modelo não apresenta assimetria estatisticamente relevante nos resíduos. Esse resultado é consistente com a literatura sobre precificação de commodities, como ferro, que frequentemente demonstra que a volatilidade dos preços é mais associada a choques de oferta e demanda globais do que a efeitos assimétricos, dado que choques positivos e negativos tendem a afetar a variância de forma relativamente simétrica (Pindyck, 2004). Assim, o baixo nível de viés de sinal reforça a adequação do modelo GARCH simétrico para capturar a dinâmica da volatilidade dessas commodities.

Tabela 7 - Teste de Viés de Sinal (Sign Bias) para Assimetria na Volatilidade

parametro	coef	se_rob	t_rob	p_rob
ma1	0,20	0,03	7,33	0,00
log_diff_AUS_lag0	0,02	0,00	6,26	0,00
log_diff_p_copper_lag2	0,02	0,00	6,15	0,00
log_diff_p_copper_lag1	0,02	0,00	6,07	0,00
diff_ir_CHINA_lag3	-0,01	0,00	-4,47	0,00
beta1	0,67	0,16	4,32	0,00
diff_ir_AUS_lag2	0,01	0,00	3,07	0,00
omega	-1,69	0,79	-2,14	0,03
diff_ir_USA_lag3	0,01	0,00	1,98	0,05
gamma1	0,33	0,18	1,80	0,07

Fonte: Elaborado pelo autor

Analisando os resultados apresentados pelo ARIMAX-GARCH, pode-se observar uma dinâmica estatisticamente significativa entre o preço do minério de ferro e variáveis macroeconômicas relacionadas ao mercado global de commodities e à

política monetária de países-chave. Em primeiro lugar, os coeficientes associados aos preços do cobre, variação logarítmica defasada em um e dois períodos do preço do cobre (log_diff_p_copper_lag1 e log_diff_p_copper_lag2), são positivos e altamente significativos. Isso indica que aumentos passados no preço do cobre tendem a antecipar movimentos positivos no preço do minério de ferro. Essa relação é economicamente plausível, considerando que ambas as commodities são amplamente utilizadas na indústria de construção e infraestrutura, especialmente em economias emergentes e em desenvolvimento como a China. Além disso, os mercados dessas commodities muitas vezes respondem de forma semelhante a choques de demanda global, o que corrobora a correlação observada.

Outro destaque importante é o coeficiente positivo de variação logarítmica da taxa de câmbio da Austrália no mesmo período (log_diff_AUS_lag0), que representa o impacto imediato da valorização do dólar australiano sobre o preço do minério. Como a Austrália é o maior exportador mundial de minério de ferro, a valorização da moeda local pode refletir maior demanda por exportações minerais, elevando o preço da commodity (Ghoshray, 2011).

Do lado da política monetária, o coeficiente negativo e significativo do diferencial da taxa de juros da China defasado em três períodos (diff_ir_CHINA_lag3) indica que aumentos nas taxas de juros chinesas há três meses tendem a reduzir o preço do minério. Isso é coerente com o efeito contracionista da política monetária, que reduz a demanda por commodities metálicas utilizadas na indústria. Os efeitos dos diferenciais de juros da Austrália (diff_ir_AUS_lag2, dois meses defasado) e dos EUA (diff_ir_USA_lag3, três meses defasado) também aparecem significativos, ainda que com menor intensidade, sugerindo que a política monetária dessas economias influencia o mercado de minério de forma indireta, por meio do câmbio e da aversão ao risco global.

A parte autoregressiva do modelo, evidenciada pelo coeficiente de média móvel de ordem 1 (ma1), é positiva e altamente significativa, confirmando a existência de autocorrelação nas variações do preço do minério. Além disso, o parâmetro de persistência na volatilidade (beta1) indica uma alta persistência, uma característica típica dos mercados de commodities, que estão sujeitos a choques frequentes de oferta e demanda. Ademais, mercados como o de petróleo e metais são fortemente

influenciados por choques estruturais, especialmente relacionados à demanda global, reforçando o caráter persistente e volátil dos preços (Kilian e Murphy, 2014)

O modelo GARCH ainda revela que choques passados têm efeito moderado sobre a volatilidade atual (omega, coeficiente da constante de variância), enquanto o parâmetro que captura assimetria de choques (gamma1) sugere algum nível de resposta diferenciada a choques positivos e negativos, embora com menor significância estatística. No geral, os resultados corroboram a importância de variáveis macroeconômicas e financeiras globais na precificação do minério de ferro, ressaltando que movimentos no preço não são puramente aleatórios, mas refletem fundamentos econômicos e decisões de política monetária em mercados-chave.

7. Considerações Finais

Com base nos resultados obtidos, é possível concluir que a dinâmica dos preços analisados está fortemente ligada tanto a fatores de mercado global quanto a condições específicas de política monetária em grandes economias. Observou-se que o preço do cobre tem influência significativa e imediata sobre o comportamento da variável dependente, indicando que oscilações nesse mercado são rapidamente incorporadas às expectativas e aos preços futuros. Além disso, variações nos juros da Austrália, Estados Unidos e, principalmente, China também apresentaram impacto relevante, revelando que a política monetária desses países exerce efeito significativo, ainda que com defasagens distintas.

O diferencial de juros da China (há três períodos) se destacou com efeito negativo, sugerindo que o aumento de juros no país asiático tem potencial de pressionar para baixo a variável dependente. Esse resultado está em linha com a literatura que aponta a China como um dos maiores consumidores de commodities do mundo, cujas políticas econômicas influenciam fortemente a demanda e os preços internacionais desses produtos. Além disso, a sensibilidade do mercado às mudanças recentes nos juros da Austrália e Estados Unidos reforça a tese de que expectativas de crescimento global e custo de capital são elementos centrais na formação de preços.

Em suma, os resultados indicam que há uma forte interligação entre os preços analisados e fatores macroeconômicos globais, sendo possível afirmar que esses preços funcionam como síntese de múltiplas forças – oferta e demanda, liquidez global e percepção de risco. Essa conclusão é especialmente relevante para investidores e formuladores de política, pois demonstra que movimentos nos preços não são aleatórios, mas refletem, de forma estruturada, o contexto econômico global em que estão inseridos.

Referências

ALQUIST, Ron; KILIAN, Lutz. What do we learn from the price of crude oil futures? **Journal of Applied Econometrics**, v. 25, n. 4, p. 539–573, 2010. Disponível em: https://www.jstor.org/stable/40657192. Acesso em: 13 ago. 2025.

BOLLERSLEV, Tim. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, v. 31, n. 3, p. 307–327, 1986. Disponível em: https://doi.org/10.1016/0304-4076(86)90063-1. Acesso em: 13 ago. 2025.

ENGLE, Robert F. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, v. 50, n. 4, p. 987–1007, 1982. Disponível em: https://www.jstor.org/stable/1912773. Acesso em: 13 ago. 2025.

GHOSHRAY, Atanu. A re-examination of volatility spillover in commodity futures markets. **Journal of Empirical Finance**, v. 18, n. 4, p. 786–803, 2011. Disponível em: https://doi.org/10.1016/j.jempfin.2011.02.004. Acesso em: 13 ago. 2025.

HUARNG, Kai; YU, Hsiao-Kang; HSU, Yu-Wei. Forecasting commodity markets: ARIMA, GARCH and hybrid models. **Economic Modelling**, v. 29, n. 3, p. 1035–1043, 2012. Disponível em: https://doi.org/10.1016/j.econmod.2012.03.014. Acesso em: 13 ago. 2025.

PINDYCK, Robert S.; ROTEMBERG, Julio J. The excess co-movement of commodity prices. **Economic Journal**, v. 100, n. 403, p. 1173–1189, 1990. Disponível em: https://www.jstor.org/stable/2233961. Acesso em: 13 ago. 2025.

Apêndice A – Código R Markdown

Importando as bibliotecas e os dados de Excel

```{r}
library(GGally)
library(tidyverse)
library(vars)
library(urca)
library(readxl)
library(tseries)
library(dplyr)
library(forecast)
library(dLagM)
library(FinTS)
library(car)
library(caret)
library(rugarch)

Model <- read_excel("/Users/gabriel/Desktop/Iniciação Científica/Modelo_Ferro.xlsx")
Model$Date <- as.Date(Model$Date)
head(Model)

```

Teste de Estacionariedade

Teste ADF na Série do Minério de Ferro

```{r}
adf_iron <- ur.df(Model$iron_price, type = "drift", selectlags = "AIC")
summary(adf_iron)

Model$log_diff_iron_price <- c(NA, diff(log(Model$iron_price)))

# Testes ADF nas séries transformadas
adf_log_diff_iron <- ur.df(na.omit(Model$log_diff_iron_price), type = "drift", selectlags = "AIC")
summary(adf_log_diff_iron)


plot(Model$Date, Model$log_diff_iron_price, type = "l",
     main = "Variação em Log-Dif do Preço do Minério de Ferro",
     xlab = "Ano", ylab = "Δlog(Preço)")

acf(na.omit(Model$log_diff_iron_price), main = "ACF - Δlog(Preço)")
pacf(na.omit(Model$log_diff_iron_price), main = "PACF - Δlog(Preço)")

# Procurando o melhor modelo
auto_model <- auto.arima(Model$log_diff_iron_price)
summary(auto_model)
acf(na.omit(residuals(auto_model)), main = "ACF dos Resíduos do Auto.ARIMA")


```

Transformação em Log-Diff

```{r}
```

```
Model$diff_ir_USA    <- c(NA, diff(Model$ir_USA))
Model$diff_ir_BR     <- c(NA, diff(Model$ir_BR))
Model$diff_ir_EU     <- c(NA, diff(Model$ir_EU))
Model$diff_ir_CHINA  <- c(NA, diff(Model$ir_CHINA))
Model$diff_ir_AUS    <- c(NA, diff(Model$ir_AUS))


Model$log_diff_EUR  <- c(NA, diff(log(Model$EUR)))
Model$log_diff_CHY  <- c(NA, diff(log(Model$CHY)))
Model$log_diff_BRL  <- c(NA, diff(log(Model$BRL)))
Model$log_diff_AUS  <- c(NA, diff(log(Model$AUS)))


Model$log_diff_p_coal    <- c(NA, diff(log(Model$p_coal)))
Model$log_diff_p_gas     <- c(NA, diff(log(Model$p_gas)))
Model$log_diff_p_alum    <- c(NA, diff(log(Model$p_alum)))
Model$log_diff_p_copper  <- c(NA, diff(log(Model$p_copper)))
Model$log_diff_p_soybean <- c(NA, diff(log(Model$p_soybean)))
Model$log_diff_p_brent <- c(NA, diff(log(Model$p_brent)))

Model$log_diff_BC_index  <- c(NA, diff(log(Model$BC_Index)))
Model$log_diff_BD_index <- c(NA, diff(log(Model$BD_Index)))

Model$diff_industry_USA    <- c(NA, diff(Model$industrialprod_USA))
Model$diff_industry_EU    <- c(NA, diff(Model$industrialprod_EU))
Model$diff_industry_CHINA    <- c(NA, diff(Model$industrialprod_CHINA))

Model$diff_retail_USA    <- c(NA, diff(Model$retail_USA))
Model$diff_retail_EU    <- c(NA, diff(Model$retail_EU))
```

```
Model$log_diff_RealEstate_USA    <- c(NA, diff(log(Model$realestate_USA)))
Model$log_diff_RealEstate_EU    <- c(NA, diff(log(Model$realestate_EU)))
Model$log_diff_RealEstate_CHINA    <- c(NA, diff(log(Model$realestate_CHINA)))
```

```
# Gráficos de correlação

variáveis_commodities <- dplyr::select(Model, iron_price, p_coal, p_gas, p_brent, p_alum, p_copper, p_soybean)
variáveis_juros <- dplyr::select(Model, iron_price, ir_USA, ir_EU, ir_CHINA, ir_BR, ir_AUS)
variáveis_cambio <- dplyr::select(Model, iron_price, EUR, CHY, BRL, AUS)
variáveis_industria_varejo <- dplyr::select(Model, iron_price, industrialprod_USA, industrialprod_EU, industrialprod_CHINA,retail_USA, retail_EU)
variáveis_imobiliario <- dplyr::select(Model, iron_price,realestate_USA, realestate_EU, realestate_CHINA)

ggpairs(variáveis_commodities, title = "Commodities em Nível")
ggpairs(variáveis_juros, title = "Juros em Nível")
ggpairs(variáveis_cambio, title = "Cambio em Nível")
ggpairs(variáveis_industria_varejo, title = "Produção Industrial e Consumo de Varejo em Nível")
ggpairs(variáveis_imobiliario, title = "Mercado Imobiliário em Nível")
```

```
# Correlação Cruzada
```

```{r}

par(mfrow = c(3, 2), mar = c(4, 4, 3, 2))

ccf(na.omit(Model$diff_ir_USA), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Juros EUA")

ccf(na.omit(Model$diff_ir_EU), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Juros UE ")

ccf(na.omit(Model$diff_ir_CHINA), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Juros China")

ccf(na.omit(Model$diff_ir_BR), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Juros Brasil ")

ccf(na.omit(Model$diff_ir_AUS), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Juros Australia ")


par(mfrow = c(3, 2), mar = c(4, 4, 3, 2))

ccf(na.omit(Model$log_diff_p_coal), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Carvão")

ccf(na.omit(Model$log_diff_p_gas), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Gás")

ccf(na.omit(Model$log_diff_p_brent), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Petróleo")

ccf(na.omit(Model$log_diff_p_alum), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Alumínio")

ccf(na.omit(Model$log_diff_p_copper), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Cobre")

ccf(na.omit(Model$log_diff_p_soybean), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Soja")


par(mfrow = c(2, 2), mar = c(4, 4, 3, 2))

ccf(na.omit(Model$log_diff_EUR), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x EUR")

ccf(na.omit(Model$log_diff_CHY), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x CHY")

ccf(na.omit(Model$log_diff_BRL), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x BRL")
```

```
ccf(na.omit(Model$log_diff_AUS), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x AUD")


par(mfrow = c(2, 2), mar = c(4, 4, 3, 2))

ccf(na.omit(Model$diff_industry_USA), na.omit(Model$diff_industry_USA), lag.max = 32, main = "Ferro x Prod. Ind. EUA")

ccf(na.omit(Model$diff_industry_EU), na.omit(Model$diff_industry_EU), lag.max = 32, main = "Ferro x Prod. Ind. UE")

ccf(na.omit(Model$industrialprod_CHINA), na.omit(Model$diff_industry_EU), lag.max = 32, main = "Ferro x Prod. Ind. China")


par(mfrow = c(2, 2), mar = c(4, 4, 3, 2))

ccf(na.omit(Model$log_diff_RealEstate_USA), na.omit(Model$log_diff_RealEstate_USA), lag.max = 32, main = "Correlação cruzada: Ferro x Mercado Imobiliário nos EUA")

ccf(na.omit(Model$log_diff_RealEstate_EU), na.omit(Model$log_diff_RealEstate_EU), lag.max = 32, main = "Correlação cruzada: Ferro x Mercado Imobiliário na UE")

ccf(na.omit(Model$log_diff_RealEstate_CHINA), na.omit(Model$log_diff_RealEstate_CHINA), lag.max = 32, main = "Correlação cruzada: Ferro x Mercado Imobiliário nos China")


par(mfrow = c(2, 1), mar = c(4, 4, 3, 2))

ccf(na.omit(Model$log_diff_BD_index), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Baltic Dry Index")

ccf(na.omit(Model$log_diff_BC_index), na.omit(Model$log_diff_iron_price), lag.max = 32, main = "Correlação cruzada: Ferro x Baltic Clean Tanker Index")




```

ARIMAX

Modelo com todas as Variáveis

ARIMAX(0,0,1)

```{r}
# 1) Listas
cambios <- c("log_diff_BRL","log_diff_EUR","log_diff_CHY","log_diff_AUS")
atividade_vars <- c("diff_industry_USA","diff_industry_EU","diff_industry_CHINA",
                    "diff_retail_USA","diff_retail_EU",

"log_diff_RealEstate_USA","log_diff_RealEstate_EU","log_diff_RealEstate_CHINA")
juros_vars <- c("diff_ir_USA","diff_ir_EU","diff_ir_CHINA","diff_ir_BR","diff_ir_AUS")
commodities <- c("log_diff_p_coal","log_diff_p_gas","log_diff_p_brent",
                 "log_diff_p_alum","log_diff_p_copper","log_diff_p_soybean")
transportes <- c("log_diff_BC_Index","log_diff_BD_Index")
max_lag <- 3

# 2) Lags
add_lags <- function(df, base_names, max_lag){
  for (var in base_names) {
    if (!var %in% names(df)) next
    for (l in 0:max_lag) {
      lag_name <- paste0(var, "_lag", l)
      df <- df |> dplyr::mutate(!!lag_name := dplyr::lag(.data[[var]], l))
    }
  }
  df
}
Model <- Model |>
  (\(df){df |> add_lags(cambios,max_lag) |>
            add_lags(atividade_vars,max_lag) |>
            add_lags(juros_vars,max_lag) |>
            add_lags(commodities,max_lag) |>
            add_lags(transportes,max_lag) })()
```

```
# 3) Seleção
pat_cambio    <- paste0("^(", paste(cambios, collapse="|"), ")_lag[0-3]$")
pat_atividade <- paste0("^(", paste(atividade_vars, collapse="|"), ")_lag[0-3]$")
pat_juros     <- paste0("^(", paste(juros_vars, collapse="|"), ")_lag[0-3]$")
pat_commod    <- paste0("^(", paste(commodities, collapse="|"), ")_lag[0-3]$")
pat_transp    <- paste0("^(", paste(transportes, collapse="|"), ")_lag[0-3]$")
xreg_vars_all <-
grep(paste(pat_cambio,pat_atividade,pat_juros,pat_commod,pat_transp,sep="|"),
                 names(Model), value = TRUE)

# 4) Base final
data_final_all <- Model |>
  dplyr::select(log_diff_iron_price, dplyr::all_of(xreg_vars_all)) |>
  stats::na.omit()

y_all <- data_final_all$log_diff_iron_price
xreg_all <- data_final_all |>
  dplyr::select(-log_diff_iron_price) |>
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) |>
  as.matrix()

ok_row <- is.finite(y_all) & apply(xreg_all, 1, \(r) all(is.finite(r)))
y_all   <- y_all[ok_row]
xreg_ok <- xreg_all[ok_row, , drop = FALSE]

keep_col <- apply(xreg_ok, 2, \(c) all(is.finite(c)))
xreg_ok  <- xreg_ok[, keep_col, drop = FALSE]

sd_col <- apply(xreg_ok, 2, sd)
xreg_ok <- xreg_ok[, sd_col > 0 & is.finite(sd_col), drop = FALSE]
```

```
lin <- caret::findLinearCombos(xreg_ok)
if (!is.null(lin$remove)) xreg_ok <- xreg_ok[, -lin$remove, drop = FALSE]

xreg_ok <- scale(xreg_ok)
stopifnot(nrow(xreg_ok) == length(y_all))

# 5) ARIMAX MA(1)
modelo_arimax_all_ma1 <- forecast::Arima(
  y_all,
  order = c(0,0,1),
  xreg  = xreg_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_all_ma1)

# 6) Tabela de significância (t, p, estrelas), ordenada por |t|
coefs <- coef(modelo_arimax_all_ma1)
ses   <- sqrt(diag(vcov(modelo_arimax_all_ma1)))
tstat <- coefs / ses
pval  <- 2*pt(-abs(tstat), df = length(y_all) - length(coefs))

sig_stars <- cut(pval,
                 breaks = c(-Inf, .001, .01, .05, .10, Inf),
                 labels = c("***","**","*",".",""))
```

```
tab_sig <- data.frame(
  parametro = names(coefs),
  coef      = as.numeric(coefs),
  se        = as.numeric(ses),
  t         = as.numeric(tstat),
  p_value   = as.numeric(pval),
  signif    = sig_stars,
  row.names = NULL,
  check.names = FALSE
)

tab_sig <- tab_sig[order(abs(tab_sig$t), decreasing = TRUE), ]

tab_sig_num <- tab_sig
num_cols <- sapply(tab_sig_num, is.numeric)
tab_sig_num[num_cols] <- round(tab_sig_num[num_cols], 4)

print(tab_sig_num, row.names = FALSE)


# Variaveis significantes
sig10_df <- tab_sig %>%
  dplyr::filter(p_value <= 0.10) %>%
  dplyr::arrange(dplyr::desc(abs(t)))

num_cols_sig <- sapply(sig10_df, is.numeric)
sig10_df[num_cols_sig] <- lapply(sig10_df[num_cols_sig], function(x) round(x, 6))

cat("\n== Variáveis significativas (<= 10%) ==\n")
```

```
print(sig10_df, row.names = FALSE)

vars_sig10 <- sig10_df$parametro[ sig10_df$parametro != "ma1" ]
cat("\nNº de variáveis (<=10%):", length(vars_sig10), "\n")
print(vars_sig10)

```

ARIMAX(1,0,0)

```{r}
max_lag <- 3

add_lags <- function(df, base_names, max_lag){
  for (var in base_names) {
    if (!var %in% names(df)) next
    for (l in 0:max_lag) {
      lag_name <- paste0(var, "_lag", l)
      df <- df %>% dplyr::mutate(!!lag_name := dplyr::lag(.data[[var]], l))
    }
  }
  df
}

Model <- add_lags(Model, cambios, max_lag)
Model <- add_lags(Model, atividade_vars, max_lag)
Model <- add_lags(Model, juros_vars, max_lag)
Model <- add_lags(Model, commodities, max_lag)
Model <- add_lags(Model, transportes, max_lag)

pat_cambio    <- paste0("^(", paste(cambios, collapse="|"), ")_lag[0-3]$")
```

```
pat_atividade <- paste0("^(", paste(atividade_vars, collapse="|"), ")_lag[0-3]$")
pat_juros     <- paste0("^(", paste(juros_vars, collapse="|"), ")_lag[0-3]$")
pat_commod    <- paste0("^(", paste(commodities, collapse="|"), ")_lag[0-3]$")
pat_transp    <- paste0("^(", paste(transportes, collapse="|"), ")_lag[0-3]$")

xreg_vars_all  <-  grep(paste(pat_cambio,  pat_atividade,  pat_juros,  pat_commod,
pat_transp, sep="|"),
                    names(Model), value = TRUE)

data_final_all <- Model %>%
  dplyr::select(log_diff_iron_price, dplyr::all_of(xreg_vars_all)) %>%
  stats::na.omit()

y_all <- data_final_all$log_diff_iron_price
xreg_all <- data_final_all %>%
  dplyr::select(-log_diff_iron_price) %>%
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) %>%
  as.matrix()

if (!requireNamespace("caret", quietly = TRUE)) utils::install.packages("caret")
library(caret)

ok_row  <- is.finite(y_all) & apply(xreg_all, 1, function(r) all(is.finite(r)))
y_all   <- y_all[ok_row]
xreg_ok <- xreg_all[ok_row, , drop = FALSE]
keep_col <- apply(xreg_ok, 2, function(c) all(is.finite(c)))
xreg_ok  <- xreg_ok[, keep_col, drop = FALSE]
sd_col   <- apply(xreg_ok, 2, stats::sd)
xreg_ok  <- xreg_ok[, sd_col > 0 & is.finite(sd_col), drop = FALSE]
lin <- findLinearCombos(xreg_ok)
if (!is.null(lin$remove)) xreg_ok <- xreg_ok[, -lin$remove, drop = FALSE]
```

```
xreg_ok <- scale(xreg_ok)
stopifnot(nrow(xreg_ok) == length(y_all))

if (!requireNamespace("forecast", quietly = TRUE)) utils::install.packages("forecast")
library(forecast)
modelo_arimax_all_ar1 <- forecast::Arima(
  y_all,
  order = c(1, 0, 0),
  xreg  = xreg_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_all_ar1)
```

ARIMAX(1,0,1)

```{r}
max_lag <- 3

add_lags <- function(df, base_names, max_lag){
  for (var in base_names) {
    if (!var %in% names(df)) next
    for (l in 0:max_lag) {
      lag_name <- paste0(var, "_lag", l)
```

```
      df <- df %>% dplyr::mutate(!!lag_name := dplyr::lag(.data[[var]], l))
    }
  }
  df
}

Model <- add_lags(Model, cambios, max_lag)
Model <- add_lags(Model, atividade_vars, max_lag)
Model <- add_lags(Model, juros_vars, max_lag)
Model <- add_lags(Model, commodities, max_lag)
Model <- add_lags(Model, transportes, max_lag)

pat_cambio    <- paste0("^(", paste(cambios, collapse="|"), ")_lag[0-3]$")
pat_atividade <- paste0("^(", paste(atividade_vars, collapse="|"), ")_lag[0-3]$")
pat_juros     <- paste0("^(", paste(juros_vars, collapse="|"), ")_lag[0-3]$")
pat_commod    <- paste0("^(", paste(commodities, collapse="|"), ")_lag[0-3]$")
pat_transp    <- paste0("^(", paste(transportes, collapse="|"), ")_lag[0-3]$")

xreg_vars_all  <-  grep(paste(pat_cambio,  pat_atividade,  pat_juros,  pat_commod,
pat_transp, sep="|"),
                    names(Model), value = TRUE)

data_final_all <- Model %>%
  dplyr::select(log_diff_iron_price, dplyr::all_of(xreg_vars_all)) %>%
  stats::na.omit()

y_all <- data_final_all$log_diff_iron_price
xreg_all <- data_final_all %>%
  dplyr::select(-log_diff_iron_price) %>%
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) %>%
  as.matrix()
```

```
if (!requireNamespace("caret", quietly = TRUE)) utils::install.packages("caret")
library(caret)

ok_row  <- is.finite(y_all) & apply(xreg_all, 1, function(r) all(is.finite(r)))
y_all   <- y_all[ok_row]
xreg_ok <- xreg_all[ok_row, , drop = FALSE]
keep_col <- apply(xreg_ok, 2, function(c) all(is.finite(c)))
xreg_ok  <- xreg_ok[, keep_col, drop = FALSE]
sd_col   <- apply(xreg_ok, 2, stats::sd)
xreg_ok  <- xreg_ok[, sd_col > 0 & is.finite(sd_col), drop = FALSE]
lin <- findLinearCombos(xreg_ok)
if (!is.null(lin$remove)) xreg_ok <- xreg_ok[, -lin$remove, drop = FALSE]
xreg_ok <- scale(xreg_ok)

if (!requireNamespace("forecast", quietly = TRUE)) utils::install.packages("forecast")
library(forecast)
modelo_arimax_all_ar1ma1 <- forecast::Arima(
  y_all,
  order = c(1, 0, 1),
  xreg  = xreg_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_all_ar1ma1)
```

```

## ARIMAX (0,0,0)
```{r}
cambios <- c("log_diff_BRL","log_diff_EUR","log_diff_CHY","log_diff_AUS")
atividade_vars <- c(
  "diff_industry_USA","diff_industry_EU","diff_industry_CHINA",
  "diff_retail_USA","diff_retail_EU",
  "log_diff_RealEstate_USA","log_diff_RealEstate_EU","log_diff_RealEstate_CHINA"
)
juros_vars <- c("diff_ir_USA","diff_ir_EU","diff_ir_CHINA","diff_ir_BR","diff_ir_AUS")
commodities <- c("log_diff_p_coal","log_diff_p_gas","log_diff_p_brent",
                 "log_diff_p_alum","log_diff_p_copper","log_diff_p_soybean")
transportes <- c("log_diff_BDI","log_diff_HARPEX","log_diff_FBX")
max_lag <- 3

add_lags <- function(df, base_names, max_lag){
  for (var in base_names) {
    if (!var %in% names(df)) next
    for (l in 0:max_lag) {
      lag_name <- paste0(var, "_lag", l)
      df <- df %>% dplyr::mutate(!!lag_name := dplyr::lag(.data[[var]], l))
    }
  }
  df
}
```
```

```
Model <- add_lags(Model, cambios, max_lag)
Model <- add_lags(Model, atividade_vars, max_lag)
Model <- add_lags(Model, juros_vars, max_lag)
Model <- add_lags(Model, commodities, max_lag)
Model <- add_lags(Model, transportes, max_lag)

pat_cambio    <- paste0("^(", paste(cambios, collapse="|"), ")_lag[0-3]$")
pat_atividade <- paste0("^(", paste(atividade_vars, collapse="|"), ")_lag[0-3]$")
pat_juros     <- paste0("^(", paste(juros_vars, collapse="|"), ")_lag[0-3]$")
pat_commod    <- paste0("^(", paste(commodities, collapse="|"), ")_lag[0-3]$")
pat_transp    <- paste0("^(", paste(transportes, collapse="|"), ")_lag[0-3]$")
xreg_vars_all  <-  grep(paste(pat_cambio,  pat_atividade,  pat_juros,  pat_commod,
pat_transp, sep="|"),
                    names(Model), value = TRUE)

data_final_all <- Model %>%
  dplyr::select(log_diff_iron_price, dplyr::all_of(xreg_vars_all)) %>%
  stats::na.omit()

y_all <- data_final_all$log_diff_iron_price
xreg_all <- data_final_all %>%
  dplyr::select(-log_diff_iron_price) %>%
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) %>%
  as.matrix()


ok_row  <- is.finite(y_all) & apply(xreg_all, 1, function(r) all(is.finite(r)))
y_all   <- y_all[ok_row]
xreg_ok <- xreg_all[ok_row, , drop = FALSE]
keep_col <- apply(xreg_ok, 2, function(c) all(is.finite(c)))
xreg_ok  <- xreg_ok[, keep_col, drop = FALSE]
```

```
sd_col   <- apply(xreg_ok, 2, stats::sd)
xreg_ok  <- xreg_ok[, sd_col > 0 & is.finite(sd_col), drop = FALSE]
lin <- findLinearCombos(xreg_ok)
if (!is.null(lin$remove)) xreg_ok <- xreg_ok[, -lin$remove, drop = FALSE]
xreg_ok <- scale(xreg_ok)

modelo_arimax_all <- forecast::Arima(
  y_all,
  order = c(0, 0, 0),   # AR(2)
  xreg  = xreg_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_all)
```

AJUSTE DO MODELO

Utilizando o MA(1) para ajuste

```{r}
```

```
# 7) Novo ARIMAX apenas com variáveis significativas (<=10%)
data_sig <- Model |>
  dplyr::select(log_diff_iron_price, dplyr::all_of(vars_sig10)) |>
  stats::na.omit()

y_sig <- data_sig$log_diff_iron_price
xreg_sig <- data_sig |>
  dplyr::select(-log_diff_iron_price) |>
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) |>
  as.matrix()

ok_row_sig <- is.finite(y_sig) & apply(xreg_sig, 1, \(r) all(is.finite(r)))
y_sig <- y_sig[ok_row_sig]
xreg_sig_ok <- xreg_sig[ok_row_sig, , drop = FALSE]

keep_col_sig <- apply(xreg_sig_ok, 2, \(c) all(is.finite(c)))
xreg_sig_ok <- xreg_sig_ok[, keep_col_sig, drop = FALSE]

sd_col_sig <- apply(xreg_sig_ok, 2, sd)
xreg_sig_ok <- xreg_sig_ok[, sd_col_sig > 0 & is.finite(sd_col_sig), drop = FALSE]

lin_sig <- caret::findLinearCombos(xreg_sig_ok)
if (!is.null(lin_sig$remove)) xreg_sig_ok <- xreg_sig_ok[, -lin_sig$remove, drop =
FALSE]

xreg_sig_ok <- scale(xreg_sig_ok)
stopifnot(nrow(xreg_sig_ok) == length(y_sig))

# ARIMAX com apenas variáveis significativas
modelo_arimax_sig_ma1 <- forecast::Arima(
```

```
  y_sig,
  order = c(0, 0, 1),
  xreg  = xreg_sig_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_sig_ma1)

# Nova tabela de significância
coefs_sig <- coef(modelo_arimax_sig_ma1)
ses_sig   <- sqrt(diag(vcov(modelo_arimax_sig_ma1)))
tstat_sig <- coefs_sig / ses_sig
pval_sig  <- 2*pt(-abs(tstat_sig), df = length(y_sig) - length(coefs_sig))

sig_stars_sig <- cut(pval_sig,
                     breaks = c(-Inf, .001, .01, .05, .10, Inf),
                     labels = c("***","**","*",".",""))

tab_sig2 <- data.frame(
  parametro = names(coefs_sig),
  coef      = as.numeric(coefs_sig),
  se        = as.numeric(ses_sig),
  t         = as.numeric(tstat_sig),
  p_value   = as.numeric(pval_sig),
  signif    = sig_stars_sig,
  row.names = NULL,
```

```
  check.names = FALSE
)

tab_sig2 <- tab_sig2[order(abs(tab_sig2$t), decreasing = TRUE), ]

tab_sig2_num <- tab_sig2
num_cols2 <- sapply(tab_sig2_num, is.numeric)
tab_sig2_num[num_cols2] <- round(tab_sig2_num[num_cols2], 4)

cat("\n== Novo modelo ARIMAX (apenas variáveis significativas) ==\n")
print(tab_sig2_num, row.names = FALSE)

# Novas variáveis significativas
sig10_df2 <- tab_sig2 %>%
  dplyr::filter(p_value <= 0.10) %>%
  dplyr::arrange(dplyr::desc(abs(t)))

num_cols_sig2 <- sapply(sig10_df2, is.numeric)
sig10_df2[num_cols_sig2]  <-  lapply(sig10_df2[num_cols_sig2],  function(x)  round(x,
6))

cat("\n== Variáveis significativas no novo modelo (<= 10%) ==\n")
print(sig10_df2, row.names = FALSE)

# Novo vetor com variáveis significativas (mantendo ma1 sempre)
vars_sig10_2 <- union(sig10_df2$parametro, "ma1")
cat("\nNº de variáveis (<=10%) + ma1 no novo modelo:", length(vars_sig10_2), "\n")
print(vars_sig10_2)
```

Filtrando novamente

```{r}
# 8) Terceiro ARIMAX com variáveis significativas do segundo modelo
vars_sig10_2_select <- setdiff(vars_sig10_2, "ma1")

data_sig2 <- Model |>
  dplyr::select(log_diff_iron_price, dplyr::all_of(vars_sig10_2_select)) |>
  stats::na.omit()

y_sig2 <- data_sig2$log_diff_iron_price
xreg_sig2 <- data_sig2 |>
  dplyr::select(-log_diff_iron_price) |>
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) |>
  as.matrix()

ok_row_sig2 <- is.finite(y_sig2) & apply(xreg_sig2, 1, \(r) all(is.finite(r)))
y_sig2 <- y_sig2[ok_row_sig2]
xreg_sig2_ok <- xreg_sig2[ok_row_sig2, , drop = FALSE]

keep_col_sig2 <- apply(xreg_sig2_ok, 2, \(c) all(is.finite(c)))
xreg_sig2_ok <- xreg_sig2_ok[, keep_col_sig2, drop = FALSE]

sd_col_sig2 <- apply(xreg_sig2_ok, 2, sd)
xreg_sig2_ok <- xreg_sig2_ok[, sd_col_sig2 > 0 & is.finite(sd_col_sig2), drop = FALSE]

lin_sig2 <- caret::findLinearCombos(xreg_sig2_ok)
if (!is.null(lin_sig2$remove)) xreg_sig2_ok <- xreg_sig2_ok[, -lin_sig2$remove, drop = FALSE]

xreg_sig2_ok <- scale(xreg_sig2_ok)
```

```
stopifnot(nrow(xreg_sig2_ok) == length(y_sig2))

# Terceiro ARIMAX com MA(1)
modelo_arimax_sig2_ma1 <- forecast::Arima(
  y_sig2,
  order = c(0, 0, 1),
  xreg  = xreg_sig2_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_sig2_ma1)

# Nova tabela de significância (3ª iteração)
coefs_sig2 <- coef(modelo_arimax_sig2_ma1)
ses_sig2   <- sqrt(diag(vcov(modelo_arimax_sig2_ma1)))
tstat_sig2 <- coefs_sig2 / ses_sig2
pval_sig2  <- 2*pt(-abs(tstat_sig2), df = length(y_sig2) - length(coefs_sig2))

sig_stars_sig2 <- cut(pval_sig2,
                      breaks = c(-Inf, .001, .01, .05, .10, Inf),
                      labels = c("***","**","*",".",""))

tab_sig3 <- data.frame(
  parametro = names(coefs_sig2),
  coef      = as.numeric(coefs_sig2),
  se        = as.numeric(ses_sig2),
```

```
  t         = as.numeric(tstat_sig2),
  p_value   = as.numeric(pval_sig2),
  signif    = sig_stars_sig2,
  row.names = NULL,
  check.names = FALSE
)

tab_sig3 <- tab_sig3[order(abs(tab_sig3$t), decreasing = TRUE), ]

tab_sig3_num <- tab_sig3
num_cols3 <- sapply(tab_sig3_num, is.numeric)
tab_sig3_num[num_cols3] <- round(tab_sig3_num[num_cols3], 4)

cat("\n== Terceiro modelo ARIMAX (refinado) ==\n")
print(tab_sig3_num, row.names = FALSE)

# Variáveis significativas (p <= 10%) no novo modelo
sig10_df3 <- tab_sig3 %>%
  dplyr::filter(p_value <= 0.10) %>%
  dplyr::arrange(dplyr::desc(abs(t)))

num_cols_sig3 <- sapply(sig10_df3, is.numeric)
sig10_df3[num_cols_sig3]  <-  lapply(sig10_df3[num_cols_sig3],  function(x)  round(x,
6))

cat("\n== Variáveis significativas no modelo final (<= 10%) ==\n")
print(sig10_df3, row.names = FALSE)

# Novo vetor com variáveis significativas (mantendo ma1)
vars_sig10_3 <- union(sig10_df3$parametro, "ma1")
cat("\nNº de variáveis (<=10%) + ma1 no modelo final:", length(vars_sig10_3), "\n")
```

```
print(vars_sig10_3)

```
```
```

Filtrando (3a vez)

```{r}
# 9) Quarto ARIMAX com variáveis significativas do terceiro modelo
vars_sig10_3_select <- setdiff(vars_sig10_3, "ma1")

data_sig3 <- Model |>
  dplyr::select(log_diff_iron_price, dplyr::all_of(vars_sig10_3_select)) |>
  stats::na.omit()

y_sig3 <- data_sig3$log_diff_iron_price
xreg_sig3 <- data_sig3 |>
  dplyr::select(-log_diff_iron_price) |>
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) |>
  as.matrix()

ok_row_sig3 <- is.finite(y_sig3) & apply(xreg_sig3, 1, \(r) all(is.finite(r)))
y_sig3 <- y_sig3[ok_row_sig3]
xreg_sig3_ok <- xreg_sig3[ok_row_sig3, , drop = FALSE]

keep_col_sig3 <- apply(xreg_sig3_ok, 2, \(c) all(is.finite(c)))
xreg_sig3_ok <- xreg_sig3_ok[, keep_col_sig3, drop = FALSE]

sd_col_sig3 <- apply(xreg_sig3_ok, 2, sd)
xreg_sig3_ok <- xreg_sig3_ok[, sd_col_sig3 > 0 & is.finite(sd_col_sig3), drop = FALSE]
```

```
lin_sig3 <- caret::findLinearCombos(xreg_sig3_ok)
if (!is.null(lin_sig3$remove)) xreg_sig3_ok <- xreg_sig3_ok[, -lin_sig3$remove, drop =
FALSE]

xreg_sig3_ok <- scale(xreg_sig3_ok)
stopifnot(nrow(xreg_sig3_ok) == length(y_sig3))

# Quarto ARIMAX com MA(1)
modelo_arimax_sig3_ma1 <- forecast::Arima(
  y_sig3,
  order = c(0, 0, 1),
  xreg  = xreg_sig3_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_sig3_ma1)

# Nova tabela de significância (4ª iteração)
coefs_sig3 <- coef(modelo_arimax_sig3_ma1)
ses_sig3   <- sqrt(diag(vcov(modelo_arimax_sig3_ma1)))
tstat_sig3 <- coefs_sig3 / ses_sig3
pval_sig3  <- 2*pt(-abs(tstat_sig3), df = length(y_sig3) - length(coefs_sig3))

sig_stars_sig3 <- cut(pval_sig3,
                      breaks = c(-Inf, .001, .01, .05, .10, Inf),
                      labels = c("***","**","*",".",""))
```

```
tab_sig4 <- data.frame(
  parametro = names(coefs_sig3),
  coef      = as.numeric(coefs_sig3),
  se        = as.numeric(ses_sig3),
  t         = as.numeric(tstat_sig3),
  p_value   = as.numeric(pval_sig3),
  signif    = sig_stars_sig3,
  row.names = NULL,
  check.names = FALSE
)

tab_sig4 <- tab_sig4[order(abs(tab_sig4$t), decreasing = TRUE), ]

tab_sig4_num <- tab_sig4
num_cols4 <- sapply(tab_sig4_num, is.numeric)
tab_sig4_num[num_cols4] <- round(tab_sig4_num[num_cols4], 4)

cat("\n== Quarto modelo ARIMAX (refinado) ==\n")
print(tab_sig4_num, row.names = FALSE)

# Variáveis significativas (p <= 10%) no modelo refinado
sig10_df4 <- tab_sig4 %>%
  dplyr::filter(p_value <= 0.10) %>%
  dplyr::arrange(dplyr::desc(abs(t)))

num_cols_sig4 <- sapply(sig10_df4, is.numeric)
sig10_df4[num_cols_sig4]  <-  lapply(sig10_df4[num_cols_sig4],  function(x)  round(x,
6))

cat("\n== Variáveis significativas no modelo refinado (<= 10%) ==\n")
print(sig10_df4, row.names = FALSE)
```

```
# Novo vetor com variáveis significativas (mantendo ma1)
vars_sig10_4 <- union(sig10_df4$parametro, "ma1")
cat("\nNº de variáveis (<=10%) + ma1 no modelo refinado:", length(vars_sig10_4), "\n")
print(vars_sig10_4)
```

Filtrando (4a vez)

```{r}
# 10) Quinto ARIMAX com variáveis significativas do quarto modelo
vars_sig10_4_select <- setdiff(vars_sig10_4, "ma1")

data_sig4 <- Model |>
  dplyr::select(log_diff_iron_price, dplyr::all_of(vars_sig10_4_select)) |>
  stats::na.omit()

y_sig4 <- data_sig4$log_diff_iron_price
xreg_sig4 <- data_sig4 |>
  dplyr::select(-log_diff_iron_price) |>
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) |>
  as.matrix()

ok_row_sig4 <- is.finite(y_sig4) & apply(xreg_sig4, 1, \(r) all(is.finite(r)))
y_sig4 <- y_sig4[ok_row_sig4]
xreg_sig4_ok <- xreg_sig4[ok_row_sig4, , drop = FALSE]

keep_col_sig4 <- apply(xreg_sig4_ok, 2, \(c) all(is.finite(c)))
xreg_sig4_ok <- xreg_sig4_ok[, keep_col_sig4, drop = FALSE]
```

```
sd_col_sig4 <- apply(xreg_sig4_ok, 2, sd)
xreg_sig4_ok <- xreg_sig4_ok[, sd_col_sig4 > 0 & is.finite(sd_col_sig4), drop = FALSE]

lin_sig4 <- caret::findLinearCombos(xreg_sig4_ok)
if (!is.null(lin_sig4$remove)) xreg_sig4_ok <- xreg_sig4_ok[, -lin_sig4$remove, drop = FALSE]

xreg_sig4_ok <- scale(xreg_sig4_ok)
stopifnot(nrow(xreg_sig4_ok) == length(y_sig4))

# Quinto ARIMAX com MA(1)
modelo_arimax_sig4_ma1 <- forecast::Arima(
  y_sig4,
  order = c(0, 0, 1),
  xreg  = xreg_sig4_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_sig4_ma1)

# Nova tabela de significância (5ª iteração)
coefs_sig4 <- coef(modelo_arimax_sig4_ma1)
ses_sig4   <- sqrt(diag(vcov(modelo_arimax_sig4_ma1)))
tstat_sig4 <- coefs_sig4 / ses_sig4
pval_sig4  <- 2*pt(-abs(tstat_sig4), df = length(y_sig4) - length(coefs_sig4))
```

```
sig_stars_sig4 <- cut(pval_sig4,
                      breaks = c(-Inf, .001, .01, .05, .10, Inf),
                      labels = c("***","**","*",".",""))

tab_sig5 <- data.frame(
  parametro = names(coefs_sig4),
  coef      = as.numeric(coefs_sig4),
  se        = as.numeric(ses_sig4),
  t         = as.numeric(tstat_sig4),
  p_value   = as.numeric(pval_sig4),
  signif    = sig_stars_sig4,
  row.names = NULL,
  check.names = FALSE
)

tab_sig5 <- tab_sig5[order(abs(tab_sig5$t), decreasing = TRUE), ]

tab_sig5_num <- tab_sig5
num_cols5 <- sapply(tab_sig5_num, is.numeric)
tab_sig5_num[num_cols5] <- round(tab_sig5_num[num_cols5], 4)

cat("\n== Quinto modelo ARIMAX (refinado) ==\n")
print(tab_sig5_num, row.names = FALSE)

# Variáveis significativas (p <= 10%) no novo modelo
sig10_df5 <- tab_sig5 %>%
  dplyr::filter(p_value <= 0.10) %>%
  dplyr::arrange(dplyr::desc(abs(t)))

num_cols_sig5 <- sapply(sig10_df5, is.numeric)
```

```
sig10_df5[num_cols_sig5] <- lapply(sig10_df5[num_cols_sig5], function(x) round(x, 6))

cat("\n== Variáveis significativas no modelo final (<= 10%) ==\n")
print(sig10_df5, row.names = FALSE)

# Novo vetor com variáveis significativas (mantendo ma1)
vars_sig10_5 <- union(sig10_df5$parametro, "ma1")
cat("\nNº de variáveis (<=10%) + ma1 no modelo final:", length(vars_sig10_5), "\n")
print(vars_sig10_5)
```

Filtrando (5a vez)

```{r}
vars_final <- c(
  "log_diff_p_copper_lag1",
  "log_diff_p_copper_lag0",
  "log_diff_p_copper_lag2",
  "log_diff_p_coal_lag0",
  "log_diff_AUS_lag0",
  "diff_ir_AUS_lag2",
  "diff_ir_USA_lag3",
  "diff_ir_CHINA_lag3"
)


# Seleção das variáveis finais
data_final_model <- Model |>
  dplyr::select(log_diff_iron_price, dplyr::all_of(vars_final)) |>
  stats::na.omit()
```

```
y_final <- data_final_model$log_diff_iron_price
xreg_final <- data_final_model |>
  dplyr::select(-log_diff_iron_price) |>
  dplyr::mutate(dplyr::across(dplyr::everything(), as.numeric)) |>
  as.matrix()

ok_row_final <- is.finite(y_final) & apply(xreg_final, 1, \(r) all(is.finite(r)))
y_final <- y_final[ok_row_final]
xreg_final_ok <- xreg_final[ok_row_final, , drop = FALSE]

keep_col_final <- apply(xreg_final_ok, 2, \(c) all(is.finite(c)))
xreg_final_ok <- xreg_final_ok[, keep_col_final, drop = FALSE]

sd_col_final <- apply(xreg_final_ok, 2, sd)
xreg_final_ok <- xreg_final_ok[, sd_col_final > 0 & is.finite(sd_col_final), drop =
FALSE]

lin_final <- caret::findLinearCombos(xreg_final_ok)
if (!is.null(lin_final$remove)) xreg_final_ok <- xreg_final_ok[, -lin_final$remove, drop =
FALSE]

xreg_final_ok <- scale(xreg_final_ok)
stopifnot(nrow(xreg_final_ok) == length(y_final))

modelo_arimax_final <- forecast::Arima(
  y_final,
  order = c(0, 0, 1),
  xreg  = xreg_final_ok,
  include.mean   = FALSE,
  method         = "ML",
  optim.method   = "BFGS",
```

```
  transform.pars = FALSE,
  optim.control  = list(maxit = 5000)
)

summary(modelo_arimax_final)


coefs_final <- coef(modelo_arimax_final)
ses_final   <- sqrt(diag(vcov(modelo_arimax_final)))
tstat_final <- coefs_final / ses_final
pval_final  <- 2 * pt(-abs(tstat_final), df = length(y_final) - length(coefs_final))

sig_stars_final <- cut(pval_final,
                       breaks = c(-Inf, .001, .01, .05, .10, Inf),
                       labels = c("***", "**", "*", ".", ""))

tab_sig_final <- data.frame(
  parametro = names(coefs_final),
  coef      = as.numeric(coefs_final),
  se        = as.numeric(ses_final),
  t         = as.numeric(tstat_final),
  p_value   = as.numeric(pval_final),
  signif    = sig_stars_final,
  row.names = NULL,
  check.names = FALSE
)

tab_sig_final <- tab_sig_final[order(abs(tab_sig_final$t), decreasing = TRUE), ]

tab_sig_final_num <- tab_sig_final
```

```
num_cols_final <- sapply(tab_sig_final_num, is.numeric)
tab_sig_final_num[num_cols_final] <- round(tab_sig_final_num[num_cols_final], 4)

cat("\n== Tabela de significância - Modelo Final ==\n")
print(tab_sig_final_num, row.names = FALSE)
```

TESTES

```{r}
# Resíduos do modelo final
resid_final <- residuals(modelo_arimax_final)
n_final <- length(resid_final)

# Ljung–Box (lags 12 e 24)
cat("\n== Ljung-Box ==\n")
for (k in c(12, 24)) print(stats::Box.test(resid_final, lag = k, type = "Ljung-Box"))

# Testes de normalidade dos resíduos
cat("\n== Normalidade ==\n")
if (n_final <= 5000) print(stats::shapiro.test(resid_final))
print(tseries::jarque.bera.test(resid_final))

# Breusch–Pagan para heterocedasticidade geral
ols_aux_final <- lm(y_final ~ xreg_final_ok)
cat("\n== Breusch-Pagan ==\n")
```

```
print(lmtest::bptest(ols_aux_final))

# ARCH–LM (heterocedasticidade condicional)
cat("\n== ARCH-LM (12) ==\n")
print(FinTS::ArchTest(resid_final, lags = 12))

# VIF (multicolinearidade)
cat("\n== VIF ==\n")
Xdf_final <- as.data.frame(xreg_final_ok)
if (ncol(Xdf_final) >= 2) {
  ols_aux_final <- lm(y_final ~ ., data = Xdf_final)
  print(car::vif(ols_aux_final))
} else {
  cat("Pulando VIF: há", ncol(Xdf_final), "regressor(es). VIF requer ≥ 2.\n")
}

```
```{r}
spec_garch <- rugarch::ugarchspec(
  mean.model = list(armaOrder = c(0, 1), include.mean = FALSE,
                    external.regressors = xreg_final_ok),
  variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
  distribution.model = "std"
)

fit_garch <- rugarch::ugarchfit(spec = spec_garch, data = y_final, solver = "hybrid")
show(fit_garch)
```

```
rob <- fit_garch@fit$robust.matcoef
rotula_mx <- function(s) {
  if (grepl("^mxreg[0-9]+$", s)) {
    k <- as.integer(sub("^mxreg", "", s))
    return(colnames(xreg_final_ok)[k])
  }
  s
}
rotulos <- vapply(rownames(rob), rotula_mx, character(1))

tab_garch <- data.frame(
  parametro = rotulos,
  coef      = rob[, 1],
  se_rob    = rob[, 2],
  t_rob     = rob[, 3],
  p_rob     = rob[, 4],
  row.names = NULL,
  check.names = FALSE
)

num_cols <- sapply(tab_garch, is.numeric)
tab_garch[num_cols] <- lapply(tab_garch[num_cols], function(x) round(x, 5))
tab_garch <- tab_garch[order(abs(tab_garch$t_rob), decreasing = TRUE), ]
print(tab_garch, row.names = FALSE)

alpha <- as.numeric(rugarch::coef(fit_garch)["alpha1"])
beta  <- as.numeric(rugarch::coef(fit_garch)["beta1"])
persistencia <- alpha + beta
cat("\nPersistência (alpha+beta):", round(persistencia, 4), "\n")
cat("Desvio-padrão condicional médio:", round(mean(rugarch::sigma(fit_garch)), 6),
"\n")
```

```
res_std <- rugarch::residuals(fit_garch, standardize = TRUE)

cat("\n== Ljung-Box resíduos ==\n")
print(stats::Box.test(res_std,  lag = 12, type = "Ljung-Box"))
print(stats::Box.test(res_std,  lag = 24, type = "Ljung-Box"))

cat("\n== Ljung-Box resíduos^2 ==\n")
print(stats::Box.test(res_std^2, lag = 12, type = "Ljung-Box"))
print(stats::Box.test(res_std^2, lag = 24, type = "Ljung-Box"))

cat("\n== Jarque-Bera (resíduos padronizados) ==\n")
print(tseries::jarque.bera.test(res_std))

```


ARIMAX GARCH Ajustado

```{r}
vars_sig10 <- c(
  "log_diff_p_copper_lag1",
  "log_diff_AUS_lag0",
  "diff_ir_AUS_lag2",
  "log_diff_p_copper_lag2",
  "diff_ir_CHINA_lag3",
  "diff_ir_USA_lag3"
)

vars_sig10 <- intersect(vars_sig10, colnames(xreg_final_ok))
xreg_sig   <- xreg_final_ok[, vars_sig10, drop = FALSE]
```

```
stopifnot(nrow(xreg_sig) == length(y_final))

spec_garch_sig <- rugarch::ugarchspec(
  mean.model = list(
    armaOrder = c(0, 1),
    include.mean = FALSE,
    external.regressors = xreg_sig
  ),
  variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
  distribution.model = "std"
)

fit_garch_sig <- rugarch::ugarchfit(spec = spec_garch_sig, data = y_final, solver =
"hybrid")
show(fit_garch_sig)

rob <- fit_garch_sig@fit$robust.matcoef

rotula_mx <- function(s) {
  if (grepl("^mxreg[0-9]+$", s)) {
    k <- as.integer(sub("^mxreg", "", s))
    return(colnames(xreg_sig)[k])
  }
  s
}
rotulos <- vapply(rownames(rob), rotula_mx, character(1))

tab_garch <- data.frame(
  parametro = rotulos,
  coef      = rob[, 1],
  se_rob    = rob[, 2],
```

```
  t_rob     = rob[, 3],
  p_rob     = rob[, 4],
  stringsAsFactors = FALSE
)

tab_garch$sig10 <- ifelse(tab_garch$p_rob <= 0.10, "*", "")
tab_garch <- tab_garch[order(abs(tab_garch$t_rob), decreasing = TRUE), ]
num_cols <- sapply(tab_garch, is.numeric)
tab_garch[num_cols] <- lapply(tab_garch[num_cols], function(x) round(x, 5))
print(tab_garch, row.names = FALSE)

alpha <- as.numeric(rugarch::coef(fit_garch_sig)["alpha1"])
beta  <- as.numeric(rugarch::coef(fit_garch_sig)["beta1"])
cat("\nPersistência (alpha+beta):", round(alpha + beta, 4), "\n")
cat("Desvio-padrão condicional médio:", round(mean(rugarch::sigma(fit_garch_sig)),
6), "\n")

res_std <- rugarch::residuals(fit_garch_sig, standardize = TRUE)

cat("\n== Ljung-Box resíduos ==\n")
print(stats::Box.test(res_std,  lag = 12, type = "Ljung-Box"))
print(stats::Box.test(res_std,  lag = 24, type = "Ljung-Box"))

cat("\n== Ljung-Box resíduos^2 ==\n")
print(stats::Box.test(res_std^2, lag = 12, type = "Ljung-Box"))
print(stats::Box.test(res_std^2, lag = 24, type = "Ljung-Box"))

cat("\n== Jarque-Bera (resíduos padronizados) ==\n")
print(tseries::jarque.bera.test(res_std))
```

```{r}

x_use <- xreg_final_ok[, intersect(colnames(xreg_final_ok), c(
  "log_diff_p_copper_lag1","log_diff_AUS_lag0","log_diff_p_copper_lag2",
  "diff_ir_AUS_lag2","diff_ir_CHINA_lag3","diff_ir_USA_lag3"
)), drop = FALSE]

y_use <- y_final

cand <- list(
  list(model="sGARCH",  order=c(1,1)),
  list(model="sGARCH",  order=c(1,2)),
  list(model="sGARCH",  order=c(2,1)),
  list(model="gjrGARCH",order=c(1,1)),
  list(model="eGARCH",  order=c(1,1)),
  list(model="iGARCH",  order=c(1,1))
)
distr_vec <- c("std","sstd")

fit_list <- list()
log_msgs <- c()

for (distr in distr_vec) {
  for (c in cand) {
    spec <- rugarch::ugarchspec(
      mean.model = list(armaOrder=c(0,1), include.mean=FALSE,
                        external.regressors = x_use),
      variance.model = list(model = c$model, garchOrder = c$order),
      distribution.model = distr
    )
```

```
    tag <- paste0(distr,"_",c$model,"_",paste(c$order,collapse=","))

    fit <- tryCatch(
      rugarch::ugarchfit(spec, data = y_use, solver = "hybrid"),
      error = function(e) { log_msgs <<- c(log_msgs, paste("FAIL:", tag, "->",
e$message)); NULL }
    )

    if (!is.null(fit)) {
      res   <- residuals(fit, standardize = TRUE)
      lb12 <- tryCatch(Box.test(res^2, lag=12, type="Ljung-Box")$p.value,
error=function(e) NA_real_)
      aic   <- tryCatch(rugarch::infocriteria(fit)[1], error=function(e) NA_real_)
      a1    <- suppressWarnings(as.numeric(rugarch::coef(fit)["alpha1"]))
      b1    <- suppressWarnings(as.numeric(rugarch::coef(fit)["beta1"]))
      pers  <- suppressWarnings(a1 + b1)
      fit_list[[tag]] <- list(fit=fit, aic=aic, lb12=lb12, pers=pers)
    }
  }
}

if (length(fit_list) == 0) {
  print(unique(log_msgs))
  stop("Interrompido.")
}

summ <- do.call(rbind, lapply(names(fit_list), function(n){
  data.frame(
    model = n,
    AIC   = as.numeric(fit_list[[n]]$aic),
    LB12  = as.numeric(fit_list[[n]]$lb12),
```

```
    PERSIST = as.numeric(fit_list[[n]]$pers),
    stringsAsFactors = FALSE
  )
}))
summ <- summ[is.finite(summ$AIC), , drop = FALSE]
if (nrow(summ) == 0) { print(unique(log_msgs)); stop("Sem AIC válido.") }

summ$score <- -summ$AIC + 5 * pmin(pmax(summ$LB12, 0), 1)
summ <- summ[order(-summ$score, summ$AIC), ]
row.names(summ) <- NULL
cat("\n=== Resumo dos candidatos (top 10) ===\n"); print(head(summ, 10))

best_name <- summ$model[1]
best_fit  <- fit_list[[best_name]]$fit
cat("\n=== Melhor modelo selecionado ===\n"); cat(best_name, "\n"); show(best_fit)

rob <- best_fit@fit$robust.matcoef
mx2lbl <- function(nm) {
  if (grepl("^mxreg[0-9]+$", nm)) {
    k <- as.integer(sub("^mxreg","", nm))
    return(colnames(x_use)[k])
  }
  nm
}
rotulos <- vapply(rownames(rob), mx2lbl, FUN.VALUE = character(1))

tab_coef <- data.frame(
  parametro = rotulos,
  coef      = rob[,1],
  se_rob    = rob[,2],
```

```
  t_rob     = rob[,3],
  p_rob     = rob[,4],
  stringsAsFactors = FALSE,
  check.names = FALSE
)
tab_coef$sig10 <- ifelse(tab_coef$p_rob <= 0.10, "*", "")
tab_coef <- tab_coef[order(abs(tab_coef$t_rob), decreasing = TRUE), ]
num_cols <- sapply(tab_coef, is.numeric)
tab_coef[num_cols] <- lapply(tab_coef[num_cols], function(x) round(x, 6))

cat("\n=== Coeficientes  ===\n")
print(tab_coef, row.names = FALSE)

tab_x <- subset(tab_coef, parametro %in% colnames(x_use))
if (nrow(tab_x) > 0) {
  cat("\n=== Apenas regressoras exógenas ===\n")
  print(tab_x, row.names = FALSE)
}

res_std <- residuals(best_fit, standardize = TRUE)
cat("\n== Diagnóstico do melhor ==\n")
cat("Ljung-Box (resíduos):\n"); print(Box.test(res_std,  lag=12, type="Ljung-Box"))
cat("Ljung-Box (resíduos^2):\n"); print(Box.test(res_std^2, lag=12, type="Ljung-Box"))
cat("Jarque-Bera (resíduos padronizados):\n"); print(tseries::jarque.bera.test(res_std))

a1 <- suppressWarnings(as.numeric(coef(best_fit)["alpha1"]))
b1 <- suppressWarnings(as.numeric(coef(best_fit)["beta1"]))
cat("\nPersistência (alpha+beta):", round(a1 + b1, 4), "\n")
cat("Desvio-padrão condicional médio:", round(mean(sigma(best_fit)), 6), "\n")
```