Antonio Pedro Morize

1812887

Um estudo comparativo de diferentes modelos de machine learning para previsões do preço do ouro

Projeto de Monografia

Orientador: Prof. Cláudio Flores

Rio de Janeiro

Dezembro de 2021

Sumário

Lista de Figuras

Lista de tabelas

1 Introdução

O ouro é um dos metais preciosos mais antigos conhecidos pelo homem e por anos foi avaliado como uma moeda global, um investimento, uma mercadoria e um objeto de beleza. Além do seu belo visual, nenhum outro mineral extraído da terra é considerado tão útil quanto o ouro. Sua utilidade deriva da sua diversidade de propriedades especiais, como por exemplo ser um excelente condutor de eletricidade, não oxidante, altamente maleável e facilmente ligado a outros metais. No entanto, estima-se que somente 10% do ouro novo produzido vai para usos práticos, enquanto os outros 50% vão para joalheria e 40% para fins de investimento. O motivo para o ouro não ser muito utilizado em processos industriais é simplesmente por causa do seu alto valor, oriundo das suas características monetárias que ressalto abaixo.

Por milhares de anos, ouro e metais preciosos foram usados como dinheiro. A sua escassez, e suas propriedades físicas e químicas o tornam uma forma estável e previsível de preservar a riqueza. Mesmo no mundo financeiramente incerto de hoje, ouro e prata ainda estão entre as mercadorias mais estáveis para se comprar. As propriedades inatas do ouro o tornam altamente valioso; no entanto, a sua escassez é o fator mais predominante, pois diz aos investidores que a oferta de ouro permanecerá aproximadamente estável. A partir disso, existe um extenso debate na literatura (Baur, D.G., & Lucey, B.M. (2010)) para determinar se o ouro é de fato um investimento seguro, especialmente durante o tempo de recessão com alto risco de inflação e flutuações da taxa de câmbio.

O ouro desempenha várias funções na economia mundial e sua ligação com variáveis financeiras e macroeconômicas está bem estabelecida (Lili, L., & Chengmei, D. (2013). Com isso em mente e tendo entendido a importância do ouro no mercado financeiro, não é surpresa que econometristas tentam estimar o preço do ouro há anos. Essa tendência ganhou ainda mais força, pois com o aumento do uso de grandes bases de dados, surgiu-se uma tendência de aplicação da econometria e do machine learning de forma integral. Essa abundância de dados, permitiu que econometristas aprimorassem os algoritmos dos seus modelos por meio da experiência e do uso de dados. Nesse sentido, o “machine learning” também é visto como uma ferramenta de inteligência artificial.

Nesse contexto, este artigo visa avaliar a adequação de uma variedade de técnicas de previsão existentes e novas ferramentas automatizadas como a de regressão em árvores, para fornecer previsões para o preço do ouro. Entre os métodos lineares que testarei estão os modelos econométricos padrão, mais especificamente, ARIMA, Random Walk e uma regressão linear em variáveis macroeconômicas. Também me aventuro nas regressões penalizadas, como os modelos Lasso e Ridge. Com relação aos modelos não lineares, estarei testando métodos baseados em árvores, como florestas aleatórias e modelos de suavização exponencial.

2 Motivação

Tudo isso considerado, podemos concluir que de fato o preço do ouro é de extrema relevância. Logo, um modelo que modelasse o gráfico do ouro poderia ter implicações diretas no trabalho de gestores de portfólio global, investidores e até para formadores de políticas públicas. Para gestores de carteiras e investidores, nossos resultados de previsão de retorno do ouro e volatilidade não só oferecem oportunidades para exploração, mas também redução de risco. Portanto, para as empresas nesse ramo mitigarem o seu risco devido a incerteza na flutuação do preço do ouro, nossas descobertas podem auxiliar nas suas estratégias de hedging e investimentos futuros.

Para formuladores de políticas, nossas evidências poderiam oferecer uma base de estudo mais sofisticada para fortalecer as regulações de mercado. Além disso, os nossos resultados poderiam ajudar futuros econometristas a entenderem melhor o funcionamento de ferramentas preditivas com machine learning no mercado de commodities.

3 Revisão Bibliográfica

Em geral, a maior parte da literatura existente sobre o mercado de ouro concentra-se em seus fatores de influência ou na melhoria dos modelos tradicionais para aumentar ainda mais a precisão das análises e previsões de preços.

Muitos estudos analisaram o padrão dos preços do ouro (Worthington, Pahlavani, 2007) para identificar os fatores que influenciam seu preço. Alguns dos fatores identificados são a inflação, a taxa de câmbio, os preços dos títulos, o desempenho do mercado e os preços do petróleo. Nesse mesmo campo, (Kausik Gangopadhyay, 2016), descobriu que o preço do ouro tem uma relação de cointegração com a taxa de câmbio, índice de preço do consumidor e com o índice do mercado de ações, sendo esse último uma relação negativa com o preço do ouro, apoiando a narrativa do papel do ouro como hedge para portfólios.

Em termos de métodos, apesar da grande aceitação dos econometristas pelos sofisticados modelos de machine learning, muitos estudiosos provam que modelos tradicionais econométricos podem ser extremamente eficientes. Dos muitos modelos possíveis disponíveis, os modelos ARMA e ARIMA estão entre os mais usados. Ru e Ren (2012), por exemplo, desenvolveram um modelo ARMA para capturar a dinâmica do mercado de alumínio e afirmam que o modelo, em uma janela de preços de curto prazo, prevê os preços de forma eficiente. No entanto, os autores destacaram que devido aos impactos das incertezas econômicas, políticas e tecnológicas sobre os preços, é importante que o modelo seja atualizado continuamente. Em uma nota parecida, Liu et al. (2018) comparou os modelos ARIMA e NAR (1) dos preços do ouro à vista e fez uma previsão de curto prazo dos preços do ouro. Os resultados mostram que o modelo autorregressivo não paramétrico (NAR (P)) se ajusta melhor. Isso é mais consistente com a avaliação feita em (GillianDooley, 2005), no

qual ele concluiu que os modelos ARMA têm desvantagens significativas que resultam em uma aproximação insatisfatória dos problemas do mundo real. A respeito dos modelos que utilizam machine learning, (Adegbenga Olayiwola, 2016) na sua tentativa de modelar o preço do cobre, provou que em uma comparação com o modelo ARIMA, a árvore de decisão produziu resultados com muito maior precisão.

Como resultado, outros pesquisadores utilizaram as técnicas ARCH e GARCH para medir a volatilidade na precificação de derivativos e na análise de risco. Para citar alguns neste campo, Tully e Lucey defenderam o uso do Poder Assimétrico GARCH (APGARCH) para modelar os efeitos das variáveis macroeconômicas sobre os preços do ouro, e Claudio Morana (2001) observou que os modelos GARCH são adequados não apenas para prever os preços do petróleo no curto prazo, mas também em horizontes diferentes, sem a necessidade de incluir mudanças estruturais.

A modelagem do mercado de commodities também foi testada com o método de Valor em risco (VaR). Este método ajuda a capturar o risco embutido no preço usando um único número real que mede e quantifica o nível de risco financeiro dentro de uma empresa, carteira ou posição em um período específico de tempo. Além de modelar VaR para índices de ações e taxas de câmbio, muitos pesquisadores também estudaram modelos de VaR no mercado de commodities. Cabedo e Moya (2003) compararam o desempenho de três modelos de VaR, abordagem padrão da simulação histórica, simulação histórica com previsão ARMA e modelo ARCH para o petróleo bruto sob o pressuposto de distribuição normal dos retornos do preço do petróleo. Eles relataram que o ARMA com simulação histórica é um modelo superior em comparação com os outros dois modelos.

Finalmente, outro método que também se mostrou bem sucedido nessa literatura foi o método STL-ETS testado em (JianChai, 2021). Além desse modelo que combina a decomposição de tendência e sazonalidade com um algoritmo de suavização exponencial, este artigo usou a rede neural e o modelo de série temporal estrutural Bayesiana para prever os retornos do preço do ouro e comparar seu desempenho com modelos de benchmark. Eles concluíram que o modelo STL-ETS pode modelar com precisão a tendência dos retornos do ouro.

4 Metodologia

O período amostral utilizado neste trabalho abrangerá entre 15 de janeiro de 1997 até 24 de junho de 2021, com um total de 8868 observações para 8 variáveis.

Está bem documentado na academia, (veja Himani Gupta e Manisha Gupta - 2017 e LiLi . L (2013)), que o ouro é altamente correlacionado a algumas variáveis macroeconômicas importantes como PIB, dólar americano, preços do petróleo bruto e inflação. Estas são apenas algumas das dezenas de outras variáveis que escolherei para me ajudar a prever o preço real do ouro. Apesar do objetivo de prever o preço do ouro, minha outra grande preocupação será avaliar a eficácia dos modelos de previsão que utilizarei. Para poder dizer com eficácia que determinado modelo teve as melhores previsões para um determinado intervalo de tempo, é necessário que eu adote certas medidas.

Nesse contexto, cada modelo é inicialmente testado usando um método denominado validação cruzada K-Fold. Essa abordagem envolve a divisão aleatória do conjunto de observações em k grupos, de tamanho aproximadamente igual. O primeiro grupo é tratado como um conjunto de validação e o método é ajustado nos k - 1 grupos restantes. O erro quadrático médio, MSE1, é então calculado nas observações no grupo retido. Este procedimento é repetido k vezes; onde em cada vez um grupo diferente de observações é tratado como um conjunto de validação. Este processo resulta em estimativas do erro de teste, MSE1, MSE2,. . . , MSEk. A estimativa de CV k-fold é calculada pela média desses valores.

Isto posto, quando finalizadas as estimativas dos modelos e iniciado o processo de comparação, será dada consideração especial aos modelos que fornecem previsões que superam o método do passeio aleatório. Esse será o caso, pois a prática de usar o passeio aleatório como benchmark de performance é amplamente aceito na literatura[1], dado que o modelo simplesmente prevê que o valor esperado de amanhã para o ouro será o valor de hoje.

Quando tratamos dos modelos auto regressivos, ou mais especificamente os ARIMA(P,D,Q), precisamos determinar de forma criteriosa os valores de p, d, q. A identificação consiste em 3 partes. Primeiro verificamos se existe a necessidade de transformações na série original, com o objetivo de estabilizar a variância, como por exemplo uma transformação logarítmica. Após isso, tomamos as diferenças da série original, quantas forem necessárias de modo a obter uma série estacionária. Podemos ver que a série transformada é estacionária por meio das FAC (decaem rapidamente) ou por meio de testes (Dickey-Fuller (ADF)). Por fim, identificamos o processo ARMA(p,q) resultante através das FAC e FACP estimadas e/ou por meio de critérios de informação como AIC e BIC.

Finalmente há o diagnóstico do modelo ajustado, através de uma análise de resíduos, para saber se o modelo é adequado. Isso pode ser feito através da verificação de Ruído Branco, que pode ser aproximada por testes de autocorrelação nos resíduos como a verificação da

[1] Benchmarks for forecasting - (https://robjhyndman.com/hyndsight/benchmarks/)

FAC do quadrado dos resíduos ou observando os resultados do teste Ljung-Box (entre outros testes de autocorrelação). Para a verificação de normalidade dos resíduos, podemos entre outras possibilidades, analisar o formato do histograma dos resíduos e fazer o teste Jarque-Bera (teste de normalidade).

Abaixo descrevo os modelos que serão estimados:

4.1 Regressão Linear Multivariada

A regressão linear múltipla tenta modelar a relação entre duas ou mais variáveis explicativas e uma variável observável ajustando uma equação linear aos dados observados. Cada valor da variável independente x está associado a um valor da variável dependente y. Formalmente, o modelo para regressão linear múltipla, dadas n observações, é:

$$_1 = \beta_0 + \beta_1 \, _1 + \beta_2 \, _2 + \ldots \beta \quad + \varepsilon, \qquad = 1, 2, \ldots \quad .$$

4.2 Passeio aleatório

O passeio aleatório é um dos modelos de previsão mais simples e conhecidos que existem. É um AR (1), onde o coeficiente autoregressivo é exatamente igual a 1. Isso significa que o valor esperado é regredido no seu valor em um período passado. As condições neste modelo são tais que a variável de interesse é uma série temporal com média constante, mas não estacionária porque sua variância não é constante.

$$= \; _0 + \sum_{=0}^{-1} \varepsilon_{-1} \qquad (\;) = \sigma^2 = (\;)$$

4.2.1 Passeio aleatório com Constante

Uma ligeira modificação que pode ser feita no passeio aleatório é o modelo de passeio aleatório com uma constante. A modificação é adicionar um componente alpha ao processo, tornando $= \alpha + \; _{-1} + \varepsilon$. Isso pode ser reformulado em, $= \alpha + \; _0 + \sum_{=0}^{-1} \varepsilon_{-1}$. Diferente da série temporal anterior, esta é não estacionária em média, pois sua expectativa depende do tempo $(\;) = \alpha + \; _0$.

4.3 ARIMA

Simplificando, a série temporal ARIMA é um modelo ARMA que foi integrado D vezes, ou seja, foi diferenciado D vezes para obter uma série estacionária. Como o nome sugere, esta série temporal tem um processo auto-regressivo e com média móvel. Por exemplo, uma série temporal ARMA (1,1) pode ser escrita como $= \rho_{-1} + \varepsilon + \theta\varepsilon_{-1}$.

Como vimos acima, para identificar a ordem e o tipo do modelo, precisamos observar a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) da série e analisar o modelo por meio de critérios de informação, como o AIC e BIC.

Os 3 pontos abaixo indicam como identificar uma ordem e tipo de modelo através da visualização da FAC.

- Um processo AR (p) tem FAC que decai de modo geométrico
- Um processo MA (q) tem FAC finita, no sentido que apresenta um corte após o lag q
- Um processo ARMA (p, q) tem FAC que decai de modo geométrico após o lag q

4.4 Ridge

A regressão Ridge é muito semelhante aos mínimos quadrados ordinários, exceto que os coeficientes são estimados minimizando uma quantidade ligeiramente diferente. Em particular, as estimativas de coeficiente de regressão ridge $\hat{\beta}$ são os valores que minimizam:

$$\hat{\beta} = \left[\sum_{=1} (- \sum_{=1} \beta)^2 + \lambda \sum_{=1} \beta^2 \right],$$

onde lambda é um parâmetro de ajuste, a ser determinado separadamente. Essa equação faz um “tradeoff” entre dois critérios diferentes. Tal como acontece com os mínimos quadrados, a regressão ridge procura estimativas de coeficientes que se ajustem bem aos dados, tornando o RSS pequeno. No entanto, o segundo termo, chamado de termo de penalização, é pequeno quando $\beta 1, \ldots, \beta p$ são próximos de zero e, portanto, tem o efeito de reduzir as estimativas de βj para zero.. O parâmetro de ajuste λ serve para controlar o impacto relativo desses dois termos nas estimativas do coeficiente de regressão.

4.5 Lasso

A regressão Lasso é muito semelhante à regressão Ridge, mas ela contém diferenças muito importantes. A principal delas é que em vez de elevar ao quadrado a inclinação como é feito com o Ridge, a regressão Lasso assume o valor absoluto. Isso permite que a regressão Lasso possa reduzir o coeficiente de inclinação βi até zero, enquanto a Ridge até perto de zero. Isso possibilita que o Lasso exclua variáveis inúteis da equação.

Assim como com a regressão Ridge, (λ) pode ser qualquer valor de zero a infinito positivo e é determinado por exemplo por validação cruzada[2]. Outra semelhança é que o laço introduz um pouco de viés, mas com menos variância do que os mínimos quadrados para fazer melhores previsões.

$$\widehat{\beta} = \left[\sum_{=1} (\quad - \sum_{=1} \quad \beta)^2 + \lambda \sum_{=1} \left| \beta \right| \right],$$

4.6 Random Forest

Uma floresta aleatória é uma coleção de árvores de regressão, projetada para reduzir a variação de previsão usando agregação de bootstrap (bagging) de árvores de regressão construídas aleatoriamente. As árvores de decisão procuram encontrar a melhor divisão para o subconjunto dos dados e normalmente são treinadas por meio do algoritmo de árvore de classificação e regressão (CART).

Na prática, um grande problema com as árvores de regressão é sua alta variância nas previsões. Normalmente, uma pequena mudança nos dados leva a sequências de divisões muito diferentes. A principal razão para tal instabilidade é a natureza do algoritmo, o efeito de um grande erro na divisão superior é propagado para todas as divisões abaixo dela. Para superar esse problema, uma técnica chamada de "bagging" (agregação de bootstrap) pode ser usada. Neste método, uma amostra aleatória de dados em um conjunto de treinamento é selecionada com substituição. Depois que várias amostras de dados são geradas, esses modelos são treinados de forma independente e dependendo do tipo de tarefa - ou seja, regressão ou classificação - a média ou a maioria dessas previsões produzem uma estimativa mais precisa. Essa abordagem é comumente usada para reduzir a variância em um conjunto de dados não comportados.

[2] A validação cruzada é um procedimento de reamostragem usado para avaliar modelos de machine learning e acessar como o modelo será executado para um conjunto de dados de teste independente.

5 Análise da série

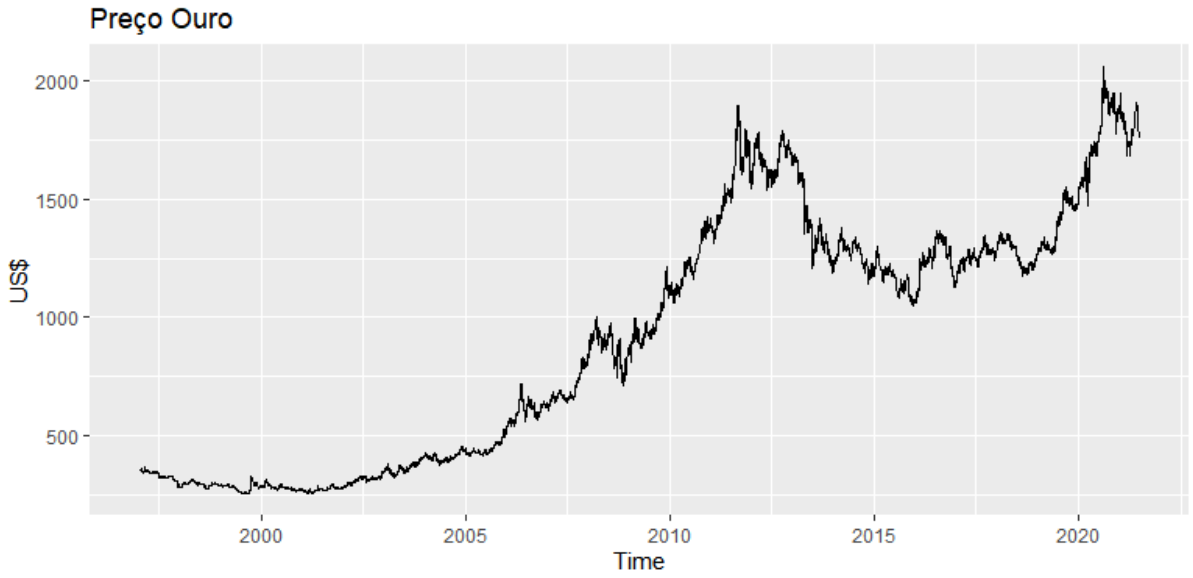

Figura 1: Preço do ouro em dólar de Janeiro de 1997 até Julho de 2021

O gráfico acima apresenta os preços reais do ouro nos últimos 23 anos precificados em dólares. Observa-se que desde o início da série em 1997, até meados de 2001 o preço permanece estável. Porém, a partir de agosto de 2001, período do qual o preço se movimenta em torno dos 270 US$, inicia-se um longo movimento de alta até o preço de 1899.60 US$ formado em Setembro de 2011. Isso é um aumento percentual em torno de 600% em uma década. Claramente, encontrar razões explicativas para esse aumento será crucial para a construção de um modelo eficiente. Após esse pico, a série apresenta uma clara tendência de baixa até encontrar um fundo de 1050 US$ em dezembro de 2015. A partir desse momento, a série teve uma reversão de tendência capaz de romper o pico de 2011 e gerar uma alta histórica de 2063.564 US$ no dia 5 de Outubro de 2020. Esse movimento foi contido, à medida que fomos nos aprofundando na pandemia do Covid 19.

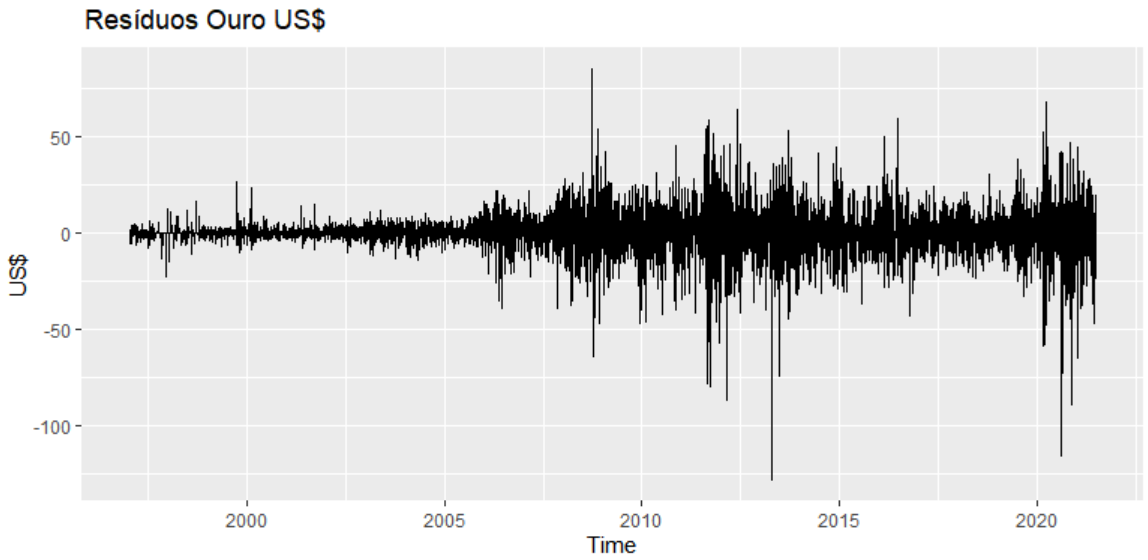

Figura 2: Resíduos do Modelo AR(1) na série do ouro

A figura acima apresenta os resíduos extraídos da série do ouro quando usamos um modelo AR(1) para modelar a série. Sob maior inspeção, é nítido que a pandemia não é o único caso de quebra estrutural no nosso modelo, vemos que após a metade da série os resíduos se tornam muito turbulentos. Mais especificamente, os resíduos demonstram maior volatilidade justamente nos períodos de forte alta da série, este é o caso com os picos formados em 2008, 2011 e 2019.

Essa maior turbulência nos dados não foi algo específico para a série do ouro, na verdade, diversas outras variáveis explicativas que escolhi sofreram um grande choque especialmente durante a pandemia do Coronavírus. Esse fato, causou uma explosão nessas variáveis, dificultando muito a previsão para o ano de 2020 e 2021. Isso pode ser visto no gráfico abaixo onde faço previsões entre os anos de 2016 e 2020, usando modelos que necessitam de variáveis explicativas.

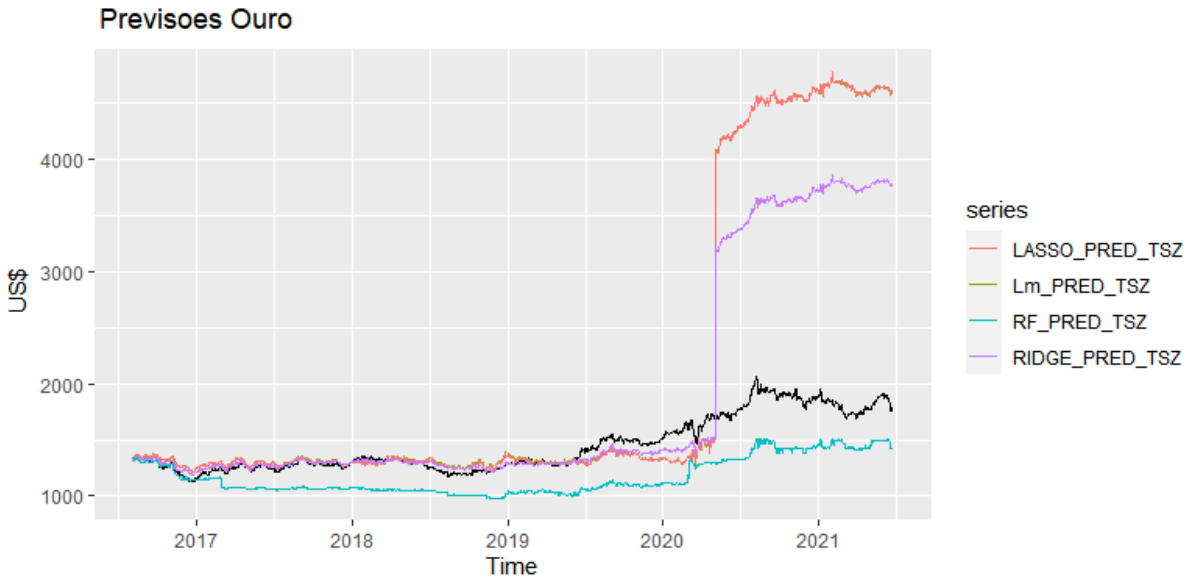

Figura 3: Previsões dos modelos Lasso, Ridge, Regressão Linear Multivariada e Random Forest para o período entre "2016-08-05" e “2021-06-24”.

Analisando o gráfico acima, vemos que somente o modelo Random Forest mantém as previsões em um patamar correto, enquanto as outras 3 apresentam uma explosão no início da pandemia que não está presente na série original . Inicialmente, o objetivo inicial era treinar o modelo entre as datas de “1997-01-15” até "2016-08-04" para gerar previsões entre o período de "2016-08-05" até “2021-06-24”. Porém, por causa da quebra estrutural observada acima, as previsões para os últimos dois anos da série foram severamente impactadas. Por esse motivo e pela percepção de que os dados do período da pandemia não representam tempos normais, mas sim uma fase de extrema incerteza, modifiquei as previsões para que elas fossem geradas entre o período de "2015-05-30" e "2019-12-31".

6 Estimando o modelo ARIMA(0,1,1)

Como definimos no capítulo acima, analisaremos a série truncada abaixo:

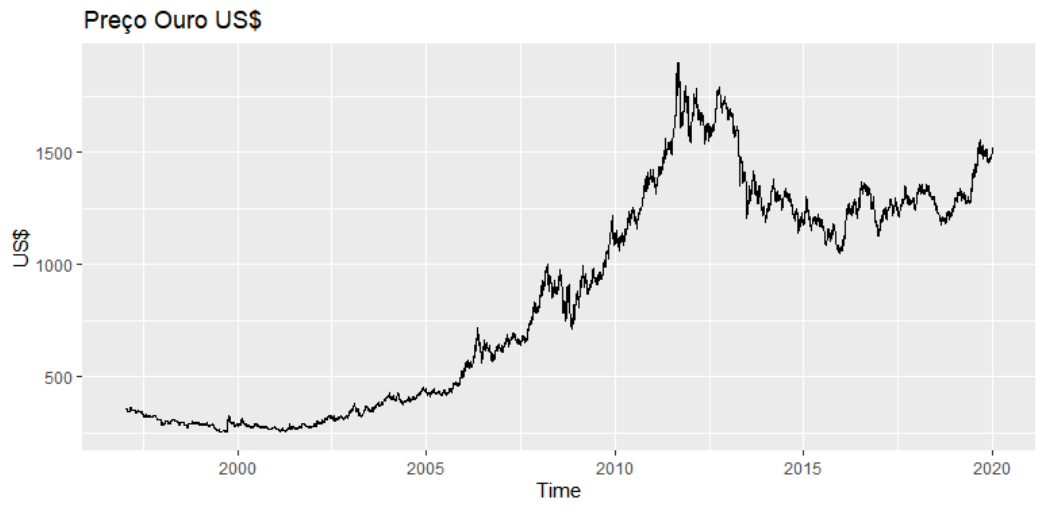

Figura 4: Preço do Ouro em dólar entre 15 de Janeiro de 1997 até 31 de Dezembro de 2019

Analisando este conjunto de dados, podemos notar que a série claramente não demonstra as características de uma série temporal estacionária e começamos a enxergar a possibilidade da série ser um passeio aleatório. Para confirmar isso, testamos pela presença de raiz unitária através do teste dickey-fuller. Os resultados indicam que a tese inicial está certa, mais especificamente, o valor da estatística de teste é 1,0265, portanto, rejeitando a hipótese nula de estacionariedade em favor da alternativa de uma raiz unitária aos níveis de significância de um, cinco e dez por cento. Realizando o mesmo teste porém com a série diferenciada, a estatística t de -65,54 indica que dessa vez rejeitamos fortemente a hipótese nula de existência de raiz unitária. Isso significa que alcançamos a estacionariedade com somente uma diferenciação.

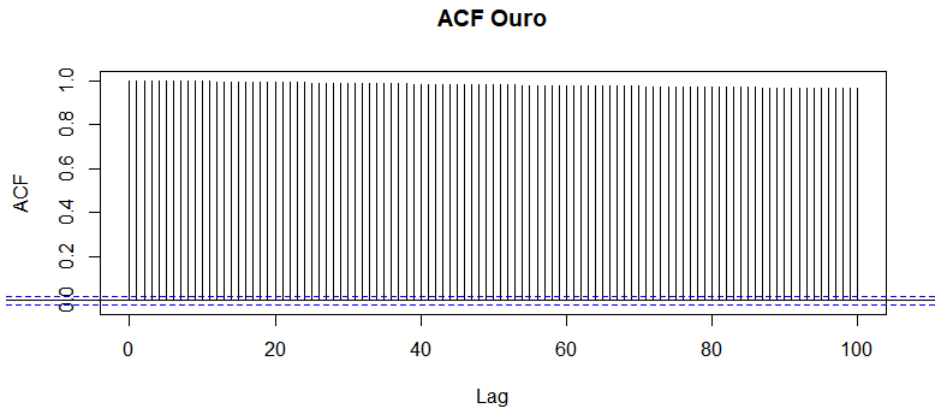

Figura 5: Gráfico da função de autocorrelação entre lags na série original do ouro.

Observando a função de autocorrelação da série original, vemos que existe autocorrelação em todos os lags, confirmando que a série é de fato não estacionária. Com essa confirmação, é importante analisar os gráficos ACF e PACF da função diferenciada, da série do ouro, pois como vimos, a série atinge estacionariedade em primeiras diferenças. Com isso podemos tomar uma decisão sobre qual é o modelo ARIMA ideal para o modelo.

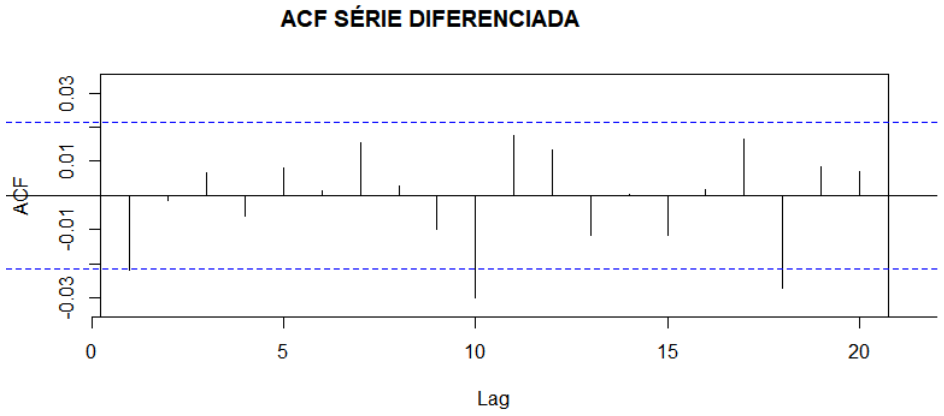

Figura 6: Gráfico da função de autocorrelação entre lags na série diferenciada do ouro.

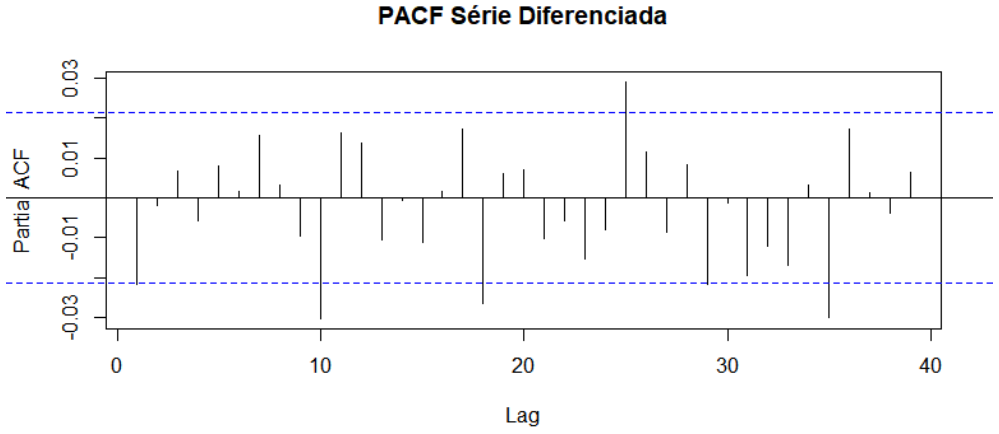

Figura 7: Gráfico da função parcial de autocorrelação entre lags na série diferenciada do ouro.

Após inspeção do gráfico ACF e PACF da série diferenciada, observamos que a autocorrelação no lag 1 é quase significativa e existe um decaimento logo em seguida. No entanto, a série não apresenta gráficos de autocorrelação limpos já que podemos ver alguns lags estatisticamente significativos na ordem superior. O lado bom, no entanto, é que não há um padrão discernível nessas defasagens. Se no gráfico ACF observássemos um decaimento geométrico suave e no PACF um corte no lag 1, usaríamos o modelo AR (1) puro. No entanto, dado que o ACF mostra um declínio após lag 1 que perdura até décimo,

construiremos um modelo MA (1) para ajustar os dados e vamos escolher d = 1 como nosso grau de diferenciação. Gerando portanto um modelo Arima (0,1,1).

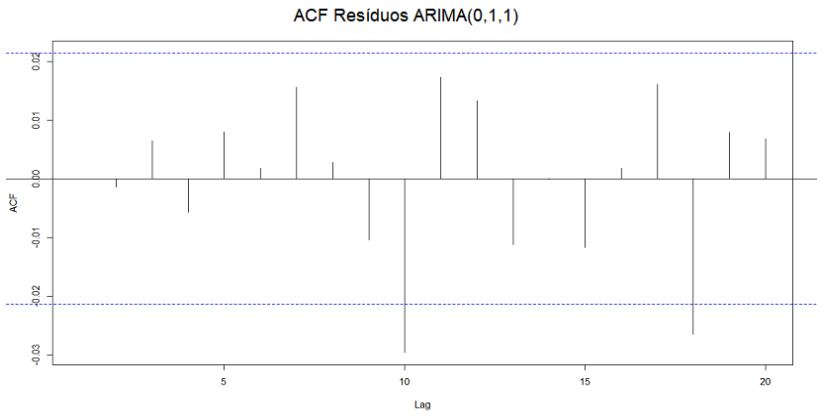

Figura 8: Gráfico da função parcial de autocorrelação entre lags nos resíduos do Modelo ARIMA(0,1,1)

Analisando o gráfico acima, vemos que a autocorrelação dos resíduos são baixas e somente o décimo e décimo oitavo lag apresentam autocorrelação significativa. Além disso, se observarmos o histograma abaixo, vemos que a frequência do tamanho dos resíduos se comporta como uma normal. Essa tese é confirmada através do teste jarque bera, teste do qual rejeitamos a hipótese nula e assim confirmamos a tese de normalidade dos resíduos da série.

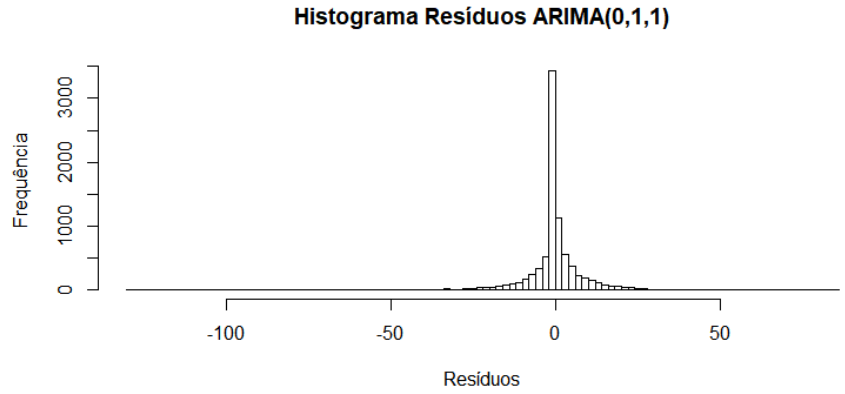

Figura 9: Histograma Resíduos do Modelo ARIMA(0,1,1)

Esses resultados são suficientes para escolhermos o modelo ARIMA da forma ARIMA(0,1,1). Também optei por adicionar a esse modelo o componente drift que permite incluir um termo constante no modelo ARIMA a fim de estimar uma tendência média diferente de zero.

7 Diagnóstico dos modelos nos dados de treino

7.1 Arima (0,1,1) com constante

$\hat{Y}_t = \mu + Y_{t-1} - \theta_1 e_{t-1}$
$\hat{Y}_t = 0{,}0118 + Y_{t-1} - 0{,}1031 e_{t-1}$

O modelo escolhido, portanto, foi um ARIMA (0,1,1) com uma constante, isso significa que o modelo tem um termo de média móvel, cujo coeficiente é 0,1031 e uma constante, 0,0188, que introduz tendência determinística no modelo e estende essa tendência indefinidamente no futuro.

7.2 Passeio Aleatório com constante

$\hat{Y}_t = \mu + Y_{t-1}$
$\hat{Y}_t = 0{,}1246 + Y_{t-1}$

O modelo de passeio aleatório por sua vez, também carrega essa constante, porém o valor que lhe é atribuído é significativamente maior, sendo 0,1246.

Como ambos os modelos acima são os únicos que não utilizam variáveis explicativas, optei por comparar o encaixe desses modelos nos dados de treino separadamente.

	Passeio Aleatório	**Arima (0,1,1)**
AIC	48270.97	51482.1
BIC	48284.59	51502.72
ME	5.279106e-05	0.0001230177
RMSE	8.83485	8.892919
MAE	4.361245	4.47937
MPE	-0.01150244	-0.01278333
MAPE	0.5444782	0.544146
MASE	1.008885	1.009985
ACF1	-0.0181254	1.887176e-05

Tabela 1: Métricas de precisão dos modelos Passeio Aleatório e ARIMA(0,1,1) nos dados de treino.

A respeito das métricas dos modelos para os dados de treino, notamos que o modelo ARIMA escolhido manualmente só supera o modelo de passeio aleatório em duas métricas, no coeficiente de autocorrelação de primeira ordem (ACF1) e no erro percentual médio absoluto (MAPE).

7.3 Ridge, Lasso, Lm e Random Forest

	Ridge	**Lasso**	**Lm**	**RF**
				Variable Imp %
lambda	0,4645257	0,2720736		
OIL	2,09333949	2,8906009	3,034	1,8718
DIX	-2,9541785	-2,1364098	-2,127	0,8856
VIX	1,52332327	-0,0835.205	-0,31	0
FED	2,63664752	10,1055983	5,705	21,0868
CPI	2,65573144	5,4755364	5,874	100
M1	0,18429957	0,2426799	0,2432	59,1502
SILVER	23,7115132	28,3173086	27,59	49,9888
US2Y	1,23139196	28,7033087	63,99	13,0279
US5Y	-16,764795	-22,20907	-87,31	8,5467
US10Y	-27,55818	-12,802488	26,42	2,4544
SPX	-0,0652.131	-0,2692554	-0,276	0,532
Cmdty_Ind	-0,5978847	-1,4441351	-1,54	1,563

Tabela 2: Coeficientes dos modelos Ridge, Lasso, Regressão Linear Multivariada e Random Forest nos dados de treino. Todos valores são estatisticamente significativos ao nível de 0,01%

Primeiramente nota-se que a regressão Ridge utiliza um valor maior que o Lasso para Lambda, a variável de penalização dos coeficientes. Isso significa que o modelo Ridge em comparação ao modelo Lasso é menos sensível a mudanças nas variáveis explicativas. À respeito dos coeficientes, vemos que existem muitas similaridades entre os modelos Ridge, Lasso e Lm, no entanto, o modelo Lm apresenta dois coeficientes com magnitudes muito elevadas, sendo eles US5Y e US2Y, com -87,31 e 63,99 respectivamente. Isso indica que o termo de penalização presente nos modelos Lasso e Ridge, de fato funcionam em controlar o problema de “overfitting” que talvez esteja presente no modelo linear.

8 Avaliação das previsões

8.1 Métricas de Precisão

	SSE	R^2	MSE	RMSE	MAPE	MASE
Ridge	4725838,39	0,73	1711036,32	1308,07	0,038	9.308835
Lasso	10005700,14	0,42	4443321,98	2107,92	0,056	13.90379
Lm	11286048,64	0,35	5699232,91	2387,31	0,060	14.83154
RF	54938655,18	-2,17	36990061,04	6081,95	0,118	30.58988
ARIMA	9697516,58	0,44	949264,25	974,30	0,049	41.12728
R Walk	9697390,88	0,44	949140,52	974,24	0,049	12.29579

Tabela 3: Métricas de precisão para as previsões feitas pelos modelos com destaque em verde para o primeiro lugar de cada coluna, em azul o segundo e em vermelho o último.

Primeiramente, podemos observar que o modelo Ridge tem melhor desempenho em 4 das 6 métricas, sendo superado somente pelo modelo de passeio aleatório nas métricas de erro quadrado médio (MSE) e raiz quadrática média (RMSE). Outro resultado marcante é o modelo Random Forest ser o pior modelo em 5 das 6 métricas e o antepenúltimo na outra (MASE).

Além disso, vemos que o modelo de passeio aleatório, o modelo mais simples para prever uma série não estacionária, claramente supera todos os modelos exceto o Ridge. Semelhantemente, outro modelo ARIMA, o ARIMA(0,1,1) também desempenha performance melhor que três dos quatro modelos que não são auto regressivos, Random Forest, Lasso, e a regressão linear multivariada.

Outro ponto curioso é a melhor performance do modelo Ridge em comparação ao modelo Lasso. Isso talvez seja decorrente do maior valor lambda do modelo Ridge, que causou uma maior penalização nos coeficientes e possivelmente reduziu a variância dos resíduos. Essa teoria que uma maior penalização acabou melhorando as nossas previsões é reforçada, uma vez que o modelo Lasso teve performance melhor que o modelo linear que não carrega nenhum tipo de penalização dos coeficientes.

8.2 Análise visual das previsões

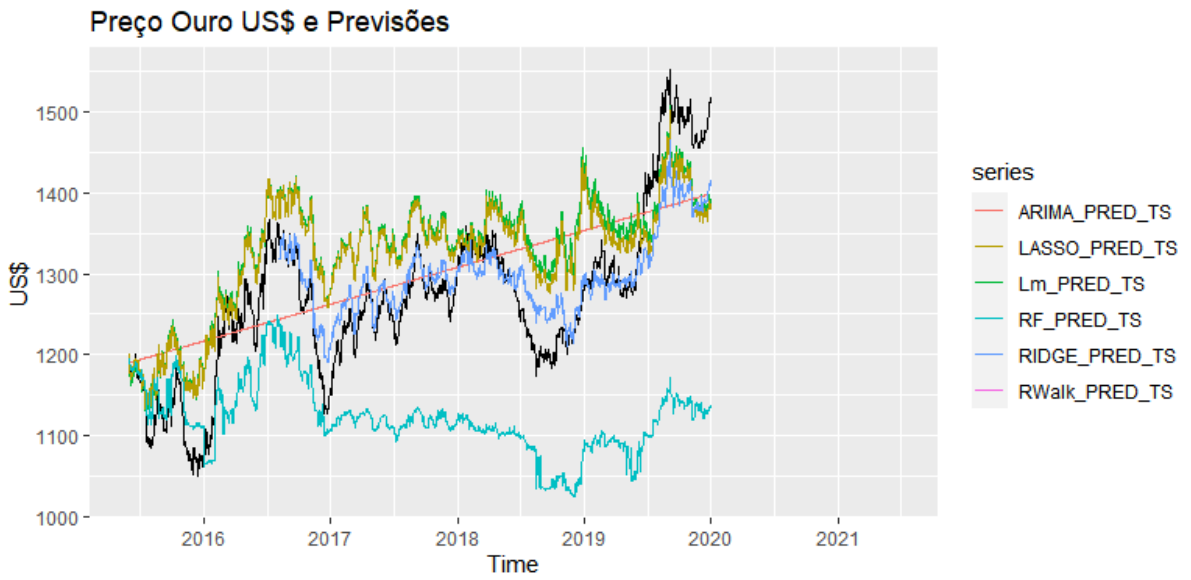

Figura 10: Previsões do modelos ao lado da série do Ouro

No gráfico acima comparo os dados de teste que são de "2015/05/30" até "2019/12/31" com as previsões fornecidas pelos modelos. Imediatamente observamos que o modelo Random Forest a partir de 2017 demonstra total desconexão com a série original, visto que essa última apresenta uma subida enquanto a primeira demonstra uma leva queda. Isso vai em linha com os resultados fornecidos na tabela acima que demonstram que esse modelo foi o mais impreciso. Além disso, o gráfico também confirma os dados da tabela que indicam que o modelo Ridge teve melhor desempenho, uma vez que a linha azul escura é a que melhor acompanha a série original traçada em preto.

Curiosamente, no início da série, as melhores previsões são feitas pelo modelo Random Forest, por ser o único capaz de capturar a magnitude da forte queda no final de 2015. A partir de 2016, o modelo Ridge apresenta previsões bem similares a série original, vale notar que os movimentos da série Lasso e Lm, também são semelhantes às previsões do modelo Ridge, no entanto, esses dois últimos modelos navegam em torno de 50 à 100 dólares acima dos valores corretos. No terceiro trimestre de 2016, a série tem uma forte queda reduzindo o preço para uma área entre 1130US$, porém, os modelos com variáveis explicativas são novamente incapazes de capturar a magnitude correta da queda. O modelo Random Forest é o que fornece uma previsão mais precisa do fundo formado no início de 2017, no entanto, a razão para isso foi a sua incapacidade de capturar a subida que ocorreu anteriormente, fazendo com que as previsões já navegassem em um patamar inferior. Excluindo os primeiros meses de 2017, o modelo Ridge apresenta previsões muito boas ao longo deste ano e isso perdura até o meio de 2018. Novamente, apesar de mostrar variações bem similares, o modelo linear e o lasso ainda seguem bem acima da série original e das previsões do modelo Ridge.

No segundo semestre de 2018 um outro fundo é formado e ele não consegue ser corretamente capturado pelas previsões. A partir desse momento, a série original mostra uma forte tendência de alta, em tal grau que após a metade de 2019, os modelos Lasso, Ridge, Passeio aleatório e linear ficam abaixo dos valores reais. Apesar de capturar corretamente as variações, os modelos Lasso, Ridge e Linear não alcançam o patamar correto para simular o pico feito pela série.

Com relação aos modelos ARIMA, o modelo de passeio aleatório e o Arima(0,1,1) demonstram previsões muito similares, de tal modo que no gráfico só conseguimos observar uma linha, pois uma está sobre a outra. Porém, apesar desse resultado, e considerando que as previsões dos modelos são linhas retas, os resultados apresentados foram extremamente satisfatórios. Inspecionando o gráfico, é nítido que os modelos corretamente prevêem a tendência a série, inclusive, conseguimos observar que as previsões dividem a série original de forma que os picos e os fundos se tornam similares aos movimentos de uma senóide.

Em virtude do que foi mencionado, podemos concluir que os modelos com melhor desempenho tiveram extrema dificuldade em capturar corretamente a volatilidade da série, prevendo valores acima dos fundos e abaixo dos picos. É perceptível também, que a penalização feita pelo modelo Ridge foi um fator diferencial, dado que esse mecanismo foi responsável por reduzir o nível geral das previsões. Por fim, como indicado nas métricas de precisão da tabela 3, o modelo Random Forest apresentou o pior desempenho e em certos momentos aparenta ter correlação negativa com a série original e as outras previsões.

8.3 Analisando o modelo Random Forest

Apesar do entendimento que as previsões do modelo Random Forest poderiam ser diferentes das previsões dos outros modelos, a sua baixa performance ainda é o resultado do estudo que mais surpreende, uma vez que esse modelo carrega as técnicas de machine learning mais complexas. Para tentar entender esse desempenho, faço uma análise abaixo para verificar se a baixa precisão das previsões foram consequências de alguma mudança de correlação entre a série do ouro e as variáveis explicativas ou se o modelo realmente não se adequou bem à série de treino.

Para fazer isso, analisarei as métricas de precisão dos modelos nos dados de treino para avaliar se o encaixe do modelo Random Forest à série original não foi adequado. Os modelos lasso e Ridge estão fora da comparação pois nos dados de treino, ambos usam o método de validação cruzada, e como ela é feita em uma variedade relativamente ampla de modelos e com estruturas variadas, não é possível obter os valores de encaixe do modelo.

	R^2	MSE	MAE
Lm	0.9891314	2547.06	38.43792
RF	0.9997711	38.72494	4.071099
ARIMA	0.999667	78.02867	4.362838
R Walk	0.9996669	78.05457	4.361245

Tabela 4: Métricas de precisão para o encaixe dos modelos nos dados de treino. Destaco em verde o primeiro lugar de cada coluna.

Na tabela acima, conseguimos ver que o modelo Random Forest teve o melhor encaixe a série original nos dados de treino, obtendo os melhores resultados e de forma bem superior ao outro modelo de regressão multivariada.

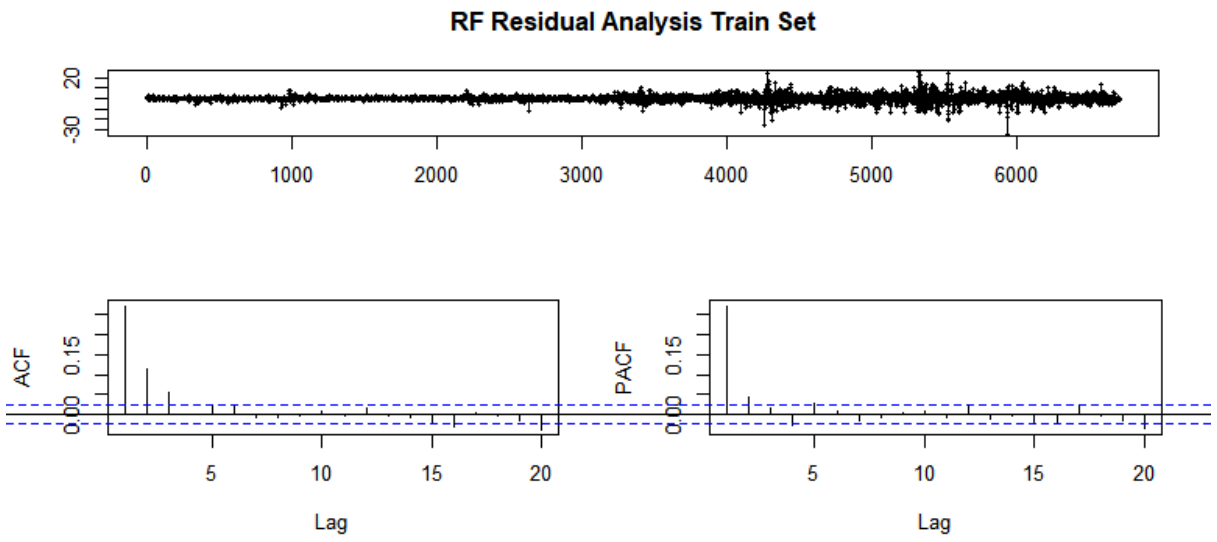

Figura 11: Análise dos resíduos do modelo Random Forest

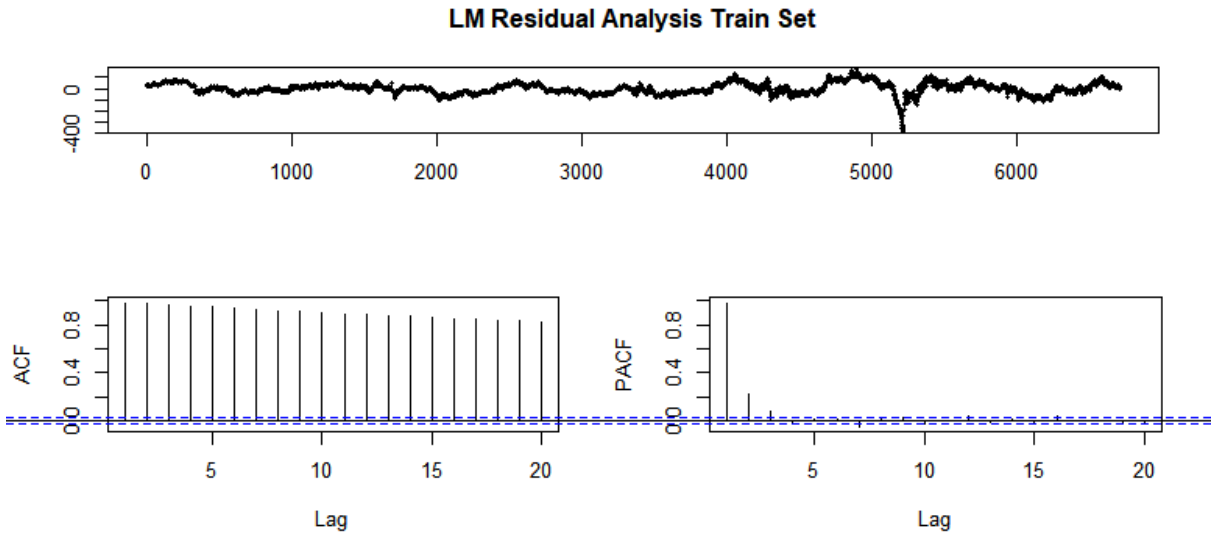

Figura 12: Análise dos resíduos do modelo de Regressão Linear Multivariada

Além disso, analisando os resíduos do modelo Random forest e comparando-os com o modelo linear, observamos que o modelo RF tem resíduos significativamente mais comportados e que apresentam autocorrelação menor. Isso nos leva a crer que os motivos para as previsões ruins do modelo RF são oriundos de alguma mudança na correlação do preço do ouro com as variáveis explicativas durante os dados de teste. Como vimos antes, as variáveis de maior relevância para o modelo são inflação (CPI) com 100% de importância, oferta monetária (M1) com 59,2% e Prata com 50%. Portanto, observei como essas variáveis explicativas se comportaram individualmente com a série do ouro.

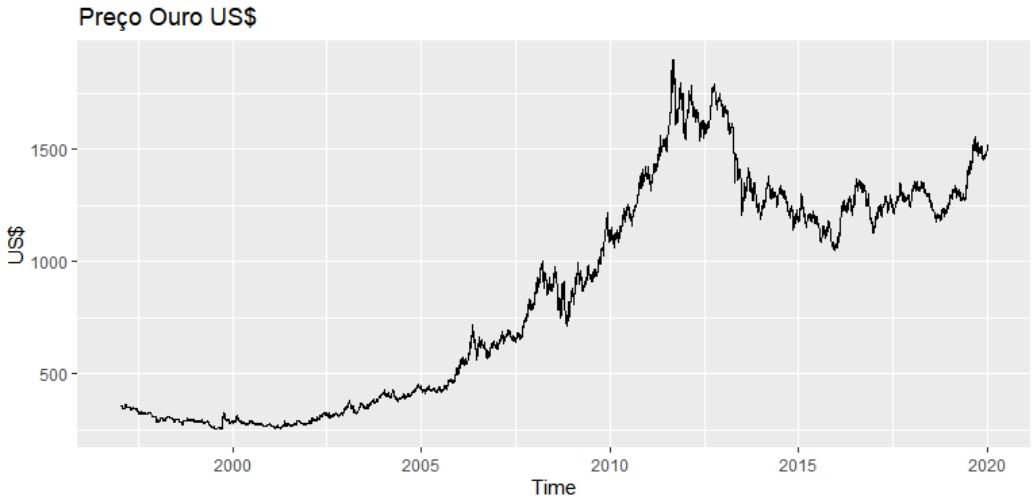

Figura 13: Preço do Ouro em dólar entre 15 de Janeiro de 1997 até 31 de Dezembro de 2019

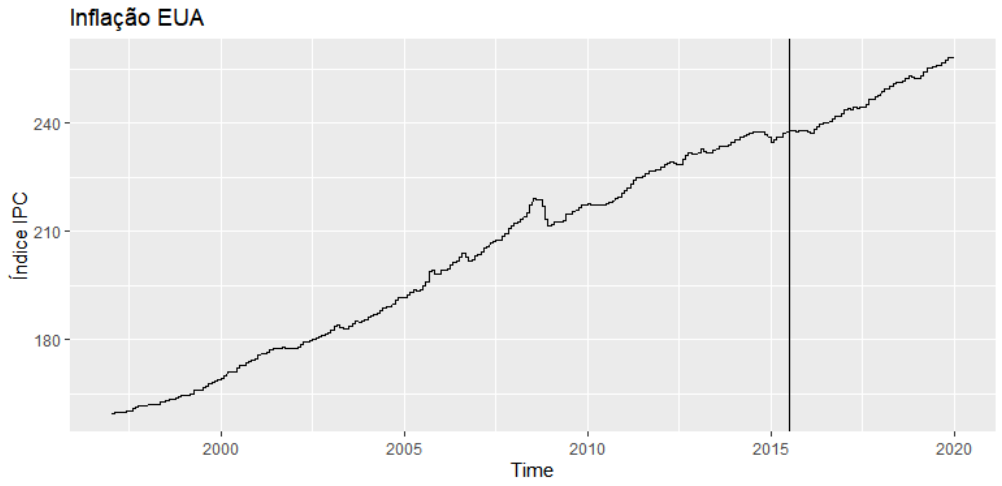

Figura 14: Índice de Preço ao Consumidor

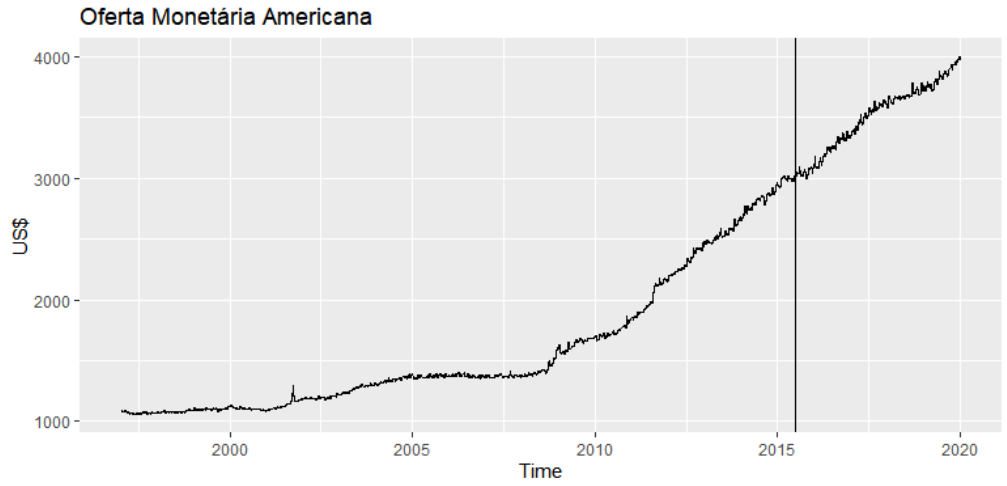

Figura 15: Oferta Monetária americana em dólares

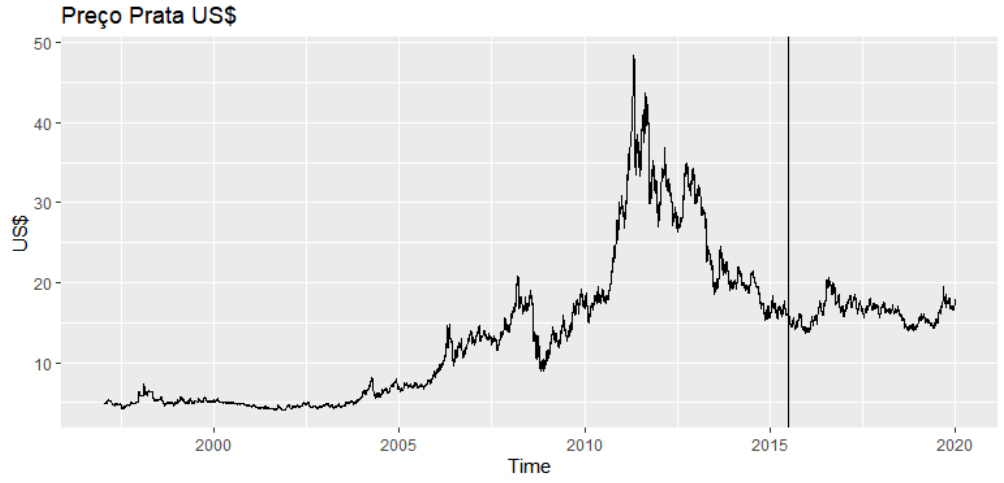

Figura 16: Preço do Prata em dólares

Observando os gráficos acima, vemos que o gráfico da inflação e da oferta monetária sobem de forma constante e acompanham a tendência de alta da série do ouro até o início de 2012, após isso, a série original inicia um movimento de descida enquanto as outras continuam subindo. Esse período de tempo entre 2012 e meados de 2015, exibe uma mudança na correlação entre as variáveis explicativas e a série do ouro, mais especificamente, as covariâncias mudam de positivo para negativo. Quando o período de teste se inicia, o gráfico da oferta monetária e da inflação permanecem subindo, a série original também apresenta este movimento, entretanto, de forma bem mais volátil, com subidas e quedas abruptas.

O gráfico da prata por sua vez é muito similar à série do ouro nos dados de teste, ambas apresentam uma período de estabilidade de 1997 - 2005, e após isso iniciam um forte movimento de alta até ela ser contida pela crise de 2008 - 2009. No entanto, as duas séries se recuperam atingem uma alta histórica entre 2011 e 2013 e depois caem. Porém, em 2015, momento em que os dados de teste começam, a série da prata se move lateralmente enquanto a do ouro inicia outro movimento de alta.

Após essas análises, podemos observar que a série do ouro no momento em que começam os dados de teste, se movimenta de forma diferente em relação às outras variáveis se comparado ao período dos dados de treino. Esse fenômeno pode ter impactado fortemente as suas previsões, e ser a razão pela qual o modelo mais sofisticado do estudo apresentou os piores resultados.

9 Conclusão

Em conclusão, esta monografia compara a precisão de métodos econométricos concorrentes usados para construir previsões para o preço futuro do ouro. O conjunto de métodos selecionados incluem, técnicas novas como árvores de regressão, abordagens tradicionais usando penalizações e modelos econométricos padrão.

Este estudo considerou dados diários para os preços do ouro durante o período de 15 de janeiro de 1997 a 31 de dezembro de 2019 para comparar as seguintes técnicas de previsão: ARIMA, Lasso, Ridge, Random Forest e Regressão Linear Múltipla). Além disso, informações diárias de 12 variáveis macroeconômicas importantes foram utilizadas para compor os modelos dependentes de variáveis explicativas. Para avaliar esses modelos, as métricas MSE, R2, SSE, RMSE, MAPE e MASE foram usadas para analisar a precisão das previsões dos modelos.

Para construir nossos modelos de machine learning, mais especificamente Ridge, Lasso e Random Forest, a validação cruzada de 10 folhas foi usada para produzir os parâmetros de ajuste dos modelos. Com relação ao modelo ARIMA, foi necessário um estudo mais aprofundado da estacionariedade e dos resíduos das séries. Para analisar isso, realizamos o teste dickey fuller, para verificar a presença de raiz unitária, e diagnosticamos gráficos ACF, PCF e histograma para avaliar a qualidade dos resíduos.

Os resultados do estudo mostraram que, para o período de teste escolhido, o modelo Ridge apresentou as previsões mais precisas. Por sua vez, o modelo Lasso, que é muito semelhante ao modelo citado acima, veio atrás de ambos os modelos ARIMA, o passeio aleatório e o ARIMA (0,1,1). Curiosamente, o modelo tecnicamente mais sofisticado, o Random Forest, foi o modelo que apresentou de longe as piores previsões. Um estudo mais aprofundado sobre as razões por trás disso mostraram que o fraco desempenho do modelo se deve a uma mudança na correlação entre a série do ouro e as variáveis explicativas de maior importância para o modelo.

Pesquisas futuras sobre esse tópico devem se concentrar em duas etapas principais. Em primeiro lugar, avaliando a dinâmica caótica dos preços do ouro usando conjuntos de dados relativamente pequenos, talvez em uma base anual para analisar períodos de crise, e em segundo lugar, integrando modelos híbridos ou métodos de aprendizagem profunda para produzir uma previsão ainda mais precisa.

Em suma, os métodos de previsão aqui aplicados para resolver um importante problema de previsão econômica podem ser úteis para ajudar a melhorar o conjunto de ferramentas atualmente utilizadas por acadêmicos e agentes de mercado para prever os preços do ouro, oferecendo assim uma contribuição valiosa para o campo da previsão macroeconômica.

10 Fonte de Dados

	Classificação	Nome	Fonte
1	Dinheiro e Crédito	M1 Money Stock	FRED - MD
2	Juros e taxas de câmbio	Effective Fed Funds Rate	FRED - MD
3	Juros e taxas de câmbio	5 year treasury rate	FRED - MD
4	Índice	CPI Index	Tradingview
5	Mercado financeiro	S&P 500	FRED - MD
6	Mercado financeiro	US - Dollar index	TradingView
7	Preços	Preço brent oil real	Tradingview
8	Financial Indicators	VIX	TradingView
9	Índice	Commodity Index	TradingView
10	Preços	Preço Prata	Tradingview
11	Juros e taxas de câmbio	2 year treasury rate	FRED - MD
12	Juros e taxas de câmbio	10 year treasury rate	FRED - MD

11 Bibliografia

TIBSHIRANI, R. (2013). An Introduction to Statistical Learning with Applications in R

BONNET R. COSTAY, A. et al (2021). Machine Learning and Oil Price Point and Density Forecasting

HASSANI, H et al (2015). Forecasting the Price of Gold, https://core.ac.uk/download/pdf/42141942.pdf

Colwell, W et al (2015). Forecasting the Price of Gold, https://rstudio-pubs-static.s3.amazonaws.com/73785_8a3efc7ee2bf452ea05f7a6f063de9ef.html#fn3

SHANKARI, S (2015). An empirical investigation of random walks in gold price movement, https://www.longdom.org/articles/an-empirical-investigation-of-random-walks-in-gold-price-movement.pdf

Collins G. Ntim et al (2015) On the efficiency of the global gold markets,

LiLi . L (2013). Research of the Influence of Macro-Economic Factors on the Price of Gold

Peter A. Abken (1980). The economics of gold price movements, Federal reserve bank of Richmond

Baur, D.G & Lucey, B.M.(2010) Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold,

Kunkler, M &,MacDonald, R. (2016) THE PRICE OF GOLD AS A HEDGE AGAINST THE US DOLLAR,

C.A.Tapia Cortez (2018), Alternative techniques for forecasting mineral commodity prices

Jian Chai (2021), Structural analysis and forecast of gold price returns

Apoorv Gupta (2018), Do VaR exceptions have seasonality? An empirical study on Indian commodity spot prices

Kausik Gangopadhyay (2016), Forecasting the price of gold: An error correction approach

DonBredin (2021), Forecasting WTI crude oil futures returns: Does the term structure help?

Edel Tully (2006), A power GARCH examination of the gold market

GillianDooley (2005), An assessment of time series methods in metal price forecasting

Adegbenga Olayiwola (2016), Forecasting Copper Spot Prices: A Knowledge-Discovery Approach