



BACKTEST E SELEÇÃO DE MODELOS DE PROJEÇÃO PROBABILÍSTICA DE PREÇOS E ANÁLISE DE RISCO EM PROJETOS DE INVESTIMENTO

Marcus Vinicius de Oliveira¹

¹AlfaZeta Analytics and Consulting, Rio de Janeiro, Brasil (contato@alfazeta.com.br)

Resumo: A performance de projetos de investimento é influenciada por diversos fatores, entre os quais se destacam os preços de insumos e produtos que são negociados internacionalmente. As commodities, em particular, possuem uma volatilidade inerente que representa riscos significativos para esses projetos, exigindo uma modelagem cuidadosa para estimar adequadamente a exposição dos resultados. Este trabalho oferece uma visão abrangente do processo de avaliação retrospectiva de modelos de projeção probabilística de preços (backtest), que é uma etapa essencial na construção de modelos para avaliação de ativos e projetos, bem como na análise de riscos. O texto aborda tanto os fundamentos teóricos quanto alguns aspectos práticos relacionados ao backtest, destacando sua relevância e as dificuldades que podem surgir ao longo desse processo.

Palavras-chave: Backtest, Análise de Risco, Modelagem, Projetos de Investimento.

INTRODUÇÃO

Os preços de commodities desempenham um papel crucial nos resultados de projetos de investimento, especialmente em setores como petróleo, gás, mineração e agricultura. A volatilidade dos preços dessas commodities pode afetar diretamente a viabilidade econômica e a rentabilidade de um projeto. Quando os preços das commodities estão em alta, os projetos de investimento tendem a se beneficiar, pois os retornos financeiros aumentam. Isso pode resultar em maior atração de investidores, aumento do fluxo de caixa e, em muitos casos, a possibilidade de expandir as operações ou reinvestir os lucros em novos projetos. Por exemplo, um projeto de exploração de petróleo se torna mais lucrativo quando o preço do barril está elevado, pois os custos fixos são diluídos por receitas maiores.

Por outro lado, a queda nos preços das commodities pode gerar desafios significativos. Projetos que foram considerados viáveis em um ambiente de preços elevados podem se tornar financeiramente insustentáveis quando os preços caem, comprometendo inclusive a financiabilidade das empresas, a depender de seu porte. Isso pode levar a cortes de custos, adiamento de investimentos ou até mesmo cancelamento de projetos. A capacidade de uma empresa de gerenciar esses riscos de preço é essencial para manter a saúde financeira e a continuidade das operações.

Além disso, a variação nos preços das commodities também influencia as decisões de financiamento. Em períodos de preços elevados, as empresas podem obter

condições de financiamento mais favoráveis, enquanto em períodos de baixa, o acesso a capital pode se tornar mais restrito e caro.

Além disso, é importante considerar que a gestão de riscos associados à flutuação dos preços das commodities, como a utilização de contratos futuros e opções, pode ajudar as empresas a se protegerem contra a volatilidade e a garantir a estabilidade financeira de seus projetos de investimento. A capacidade de adaptar estratégias em resposta às condições de mercado é fundamental para maximizar os resultados e garantir a sustentabilidade dos projetos a longo prazo.

A simulação do preço do petróleo pode ser realizada usando modelagem de processo estocástico, conforme apresentado em (SCHWARTZ e SMITH, 2000), na qual o preço à vista é modelado por dois fatores, um de curto prazo e um de longo prazo. O fator de curto prazo (χ_t) tem dinâmica de reversão à média com coeficiente κ (velocidade de reversão à média) e volatilidade σ_χ , e o fator de longo prazo (ξ_t) tem dinâmica de passeio aleatório com drift μ_ξ e volatilidade σ_ξ . Os dois processos são correlacionados por meio dos incrementos dz_ξ e dz_χ , tais que $dz_\xi \cdot dz_\chi = \rho_{\chi,\xi}$ (coeficiente de correlação $\rho_{\chi,\xi}$). A equação (1) apresenta a formulação matemática do modelo conhecido como modelo de 2 fatores de Schwartz-Smith. Outras versões do modelo foram propostas na literatura, incluindo mais fatores ou modificando a dinâmica dos fatores, como permitir a reversão média do fator de longo prazo.

$$\begin{cases} St = \exp(\xi_t + \chi_t) \\ d\xi_t = \mu_\xi dt + \sigma_\xi dz_\xi \\ d\chi_t = -\kappa\chi_t dt + \sigma_\chi dz_\chi \end{cases} \quad (1)$$

A Figura 1 apresenta um exemplo de simulação realizada utilizando o modelo Schwartz-Smith de dois fatores para modelar o comportamento do Brent a longo prazo. Como se trata de uma simulação de longo prazo (cerca de 15 anos), a dispersão dos resultados é bastante significativa. Além disso, é importante destacar que nem sempre é necessário ou possível realizar um backtest com um horizonte tão longo, principalmente devido à escassez de dados. Por isso, frequentemente são realizados backtests com horizontes mais curtos, mesmo que os modelos precisem ser utilizados para horizontes maiores, simplesmente por questões de inviabilidade técnica.



Figura 1 Exemplo intervalo P10-P50 e mediana de uma simulação estocástica de preço de petróleo Brent de longo prazo utilizando-se o modelo Schwartz-Smith de dois fatores.

Apesar de muito eficientes e utilizados na simulação de séries de ativos financeiros, modelos estocásticos normalmente assumem premissas sobre a distribuição dos dados com base nas observações, o que usualmente leva a premissa de alguma normalidade nas distribuições de probabilidade. Outras técnicas como a geração de cenários para séries temporais utilizando GAN não precisam dessas premissas e podem gerar distribuições mais genéricas, como a metodologia apresentada em (YOON, 2019). Seja qual for a metodologia utilizada, é sempre necessário realizar um backtest rigoroso para garantir que o modelo esteja bem calibrado para representar fielmente a distribuição de probabilidade de preço e, conseqüentemente, suportar uma análise de risco robusta.

Apesar de constantemente a projeção de preços ser tratada como uma previsão (forecast), a previsão de preços é algo impossível, uma vez que depende de eventos futuros imprevisíveis. Portanto, empreender esforços tentando prever preços de ativos financeiros no futuro não é um bom uso de tempo e recursos. O melhor que se pode obter é uma estimativa da distribuição de probabilidade de preços dos ativos no

futuro, considerando-se o comportamento que já foi observado no futuro.

Modelos Estocásticos são criados para modelar o comportamento probabilístico dos preços dos ativos financeiros de forma a representar o comportamento histórico observado. Em (AIUBE, 2012) uma extensa análise de diversos aspectos relacionados à modelagem de processos estocásticos é apresentada. Além disso, mais recentemente, outras ferramentas baseadas em modelos de Machine Learning tem sido utilizadas para se obter estimativas probabilísticas de múltiplos cenários para séries temporais, incluindo aquelas baseadas em modelos generativos como apresentado em (GEISSLER et al, 2022) e (YOON et al, 2019).

Dessa forma, ao criar um modelo de projeção probabilística de preços, é necessário avaliar retrospectivamente a sua performance para verificar se a sua qualidade é satisfatória. Este procedimento é o que chamamos de backtest. A Figura 2 apresenta dois exemplos de janelas utilizadas durante o backtest. Os dados históricos são divididos em duas partes, o trecho sinalizado em verde é utilizado para ajustar o modelo (definir seus parâmetros) e o trecho azul é usado para calcular as métricas de qualidade do modelo. Não pode haver qualquer interseção entre os dados utilizados para ajustar (treinar) o modelo e avaliá-lo numa mesma janela.



Figura 2 Exemplo esquemático de janelas construídas para o backtest. Em cada janela, os dados são divididos em dois, sendo parte usada para ajustar o modelo e parte usada para avaliá-lo

Um aspecto importante a ser considerado quando se utiliza modelos para previsão probabilística de preços de ativos é o entendimento de que, mesmo que a performance atingida seja boa, é necessário lembrar que trata-se de uma análise retrospectiva e que não garante que de fato o comportamento futuro dos preços estará bem representado pela distribuição de probabilidades estimada. Um exemplo claro disso vem da própria indústria de petróleo. Embora modelos estocásticos sejam frequentemente utilizados para

simulação estocástica de preços de petróleo e derivados, não é possível garantir que o comportamento passado será replicado no futuro, sobretudo em virtude da transição energética em curso, que coloca dúvidas sobre a demanda por petróleo e derivados nos próximos anos.

Por isso, embora isso não seja tratado neste trabalho, é fundamental que, além de uma análise de performance dos modelos utilizados para a simulação dos preços, sejam também realizadas análises de sensibilidade e testes de stress para se ter uma melhor quantificação de impacto de cenários não cobertos ou com probabilidade subestimada por um modelo estocástico ajustado com dados históricos.

MATERIAL E MÉTODOS

Antes de fazer o backtest é necessário avaliar quando ele deve começar e qual é o passo entre as janelas. Quanto mais longo for o backtest, estatisticamente mais significativos serão seus resultados. No entanto, é preciso sempre avaliar o tamanho da séries temporais envolvidas. Por exemplo, imagine que você tenha uma série de 10 anos e que esteja pensando em fazer um backtest dos últimos 5 anos. A sua primeira janela de backtest considerará dados do ano 1 ao ano 5 para ajustar o modelo, o que pode ser insuficiente. Dessa forma, é importante sempre ponderar que quanto mais cedo começa seu backtest, menos dados se tem para ajustar o modelo nas primeiras janelas de backtest.

Outra decisão importante é o passo de uma janela de backtest para a próxima. Como queremos o máximo de janelas possível, o mais natural é dar o passo de apenas uma unidade de tempo para a construir a próxima janela. Por exemplo, imagine que seus dados são mensais e que sua janela 1 termine em 01/02/2010 (último dia de dado para ajustar o modelo), a próxima janela (janela 2) terminaria em 01/03/2010. Em alguns casos pode ser que isso leve a uma quantidade muito grande de janelas e se torne inviável em termos de capacidade computacional. No entanto, sempre que possível, mantenha o máximo de janelas de backtest que for viável, de modo que as métricas de performance tenham boa significância estatística. No caso do exemplo apresentado neste documento, foram considerados dados mensais e o início do backtest é em Jan/2010, totalizando **159 janelas**.

As etapas de um backtest podem ser sumarizadas da seguinte forma:

1. Definir o período de teste: Determine o intervalo de tempo para o qual você deseja realizar o backtest conforme discutido anteriormente.

2. Para cada janela, dividir os dados: Separe seus dados em conjuntos de treinamento e teste.
3. Ajustar (ou treinar) o modelo: Utilize o conjunto de treinamento para treinar o modelo de previsão.
4. Fazer previsões: Utilize o modelo treinado para fazer previsões no conjunto de teste.
5. Calcular as métricas
6. Verificar se as métricas são adequadas e comparar com métricas de outros modelos
7. Escolher o melhor modelo, combinando a análise comparativa das métricas com uma análise crítica, priorizando modelos mais simples em relação a modelos mais complexos.

As métricas podem variar a depender do estudo. Para esta análise, vamos utilizar as métricas de hit, perda quantílica e erro médio relativo. Vamos definir como foram calculadas cada uma dessas métricas a seguir.

Métricas de Frequência de Violação de Resultados

Para cada janela, conta-se a quantidade de vezes em que o preço fica acima da banda superior (P90) ou abaixo da banda inferior (P10) e divide-se pelo total pelo número de intervalos de tempo de projeção (no caso, 12, porque são 12 meses de projeção).

O HIT é muito útil porque é uma métrica de simples interpretação. Um bom modelo deve ter um HIT próximo de 20% para bandas P10 e P90, ou seja, deve violar as bandas aproximadamente 20% das vezes. Não deve ser utilizado, por exemplo, um modelo que tenha um HIT sistematicamente baixo (5%, por exemplo) uma vez que ele respeita o intervalo de 80% em 95% das projeções, assim como não é adequado um modelo que tem um HIT de 40%, uma vez que ele respeita o intervalo de 80% apenas 60% das vezes.

Métricas de Erro Absoluto e Relativo

Para cada janela de backtest, mensura o quanto a mediana (P50) da projeção se distancia relativamente do que foi realizado, tendo a formulação dada pela equação a seguir

$$EMR = \frac{1}{N_{projecao}} \sum_{k=1}^{N_{projecao}} \frac{|P50-realizado|}{realizado} \quad (2)$$

Trata-se de uma métrica útil para comparação entre os modelos. Quanto menor for o erro médio relativo, melhor será o modelo. Além disso, como foi mencionado anteriormente, é impossível prever preço, de forma que a mediana (P50) não pode ser interpretada como “valor esperado”, “valor mais provável” nem nada equivalente a isso, tratando-se apenas da mediana da distribuição de probabilidade de preço projetada pelo modelo.

Perda Quantílica

A perda quantílica (ou quantile loss em inglês) é uma métrica utilizada para avaliar o desempenho de modelos de previsão em relação à incerteza nas previsões. Ela é particularmente útil quando se deseja estimar intervalos de confiança para as previsões. Para um percentil α (0.9 se estivermos falando do P90 e 0.1 se for P10), a perda quantílica é definida pela equação a seguir.

$$qLoss = \sum_{k=1}^{N_{projecao}} \theta(k) \quad (3)$$

$$\theta(k) = \begin{cases} \alpha(\text{realizado} - \text{percentil}) & \text{se realizado} > \text{percentil} \\ (1 - \alpha)(\text{percentil} - \text{realizado}) & \text{se realizado} < \text{percentil} \end{cases}$$

Assim como o HIT, a perda quantílica quantifica violações dos percentis, penalizando mais modelos que violam essas bandas por uma maior distância que outros que violam com uma distância maior. A ideia é penalizar modelos que erram por muito (basicamente, é um complemento do HIT).

Métricas de Distância de Ponto e Distribuição

No contexto de análise de performance de modelos probabilísticos, existem as métricas de distância que podem ser utilizadas para medir a distância entre um ponto e uma distribuição. Neste trabalho vamos destacar a Distância de Mahalanobis, (MAHALANOBIS, 1936), (VENTURINI, 2015), que é uma métrica é especialmente útil em contextos multivariados, pois leva em consideração a correlação entre as variáveis. A distância de Mahalanobis entre um ponto x e uma distribuição com média μ e matriz de covariância Σ é dada por:

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \quad (4)$$

onde $x - \mu$ é o vetor de diferença entre o ponto e a média da distribuição, Σ^{-1} é a inversa da matriz de covariância.

A distância de Mahalanobis é particularmente útil para identificar outliers em dados multivariados, uma vez que considera não apenas a média, mas também a variabilidade e a correlação entre as variáveis.

RESULTADOS E DISCUSSÃO

A Figura 3 apresenta três janelas utilizadas no backtest. Nota-se que em alguns casos o modelo acerta mais as bandas e em outros menos. Naturalmente, como trata-se de uma previsão probabilística, acertar o preço não é o objetivo do modelo, o fato de em uma janela ou outra a P50 se aproximar do realizado não quer dizer nada, a qualidade do modelo tem que ser avaliada através das métricas, considerando não apenas o

quanto a realização se aproxima do P50 mas também, o que é mais importante, como a realização está com relação ao intervalo de confiança, P10-P90.

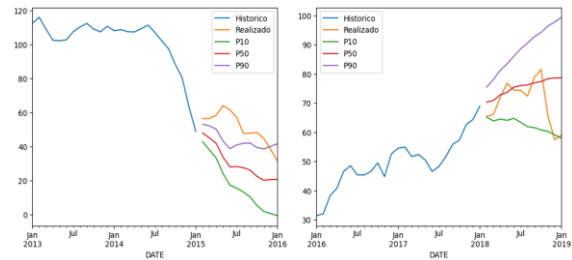


Figura 3 Exemplo de duas janelas, comparando os percentis obtidos pela simulação de um modelo AR(5) com dados históricos

Foram ajustados 7 modelos, para os quais foi realizado o backtest desde 2010 com step mensal, ou seja, considerando janelas com fim em todos os meses a partir de janeiro de 2010. Naturalmente, como é preciso ter 12 meses de dados para comparar com o resultado da simulação, a última janela deve terminar 13 meses antes do fim dos dados, de modo que ainda seja possível calcular as métricas para esta janela.

Nas Figuras Figura 4, Figura 5 e Figura 6 são apresentados os resultados das 3 métricas para os 7 modelos utilizados como exemplo de comparação de performance no formato de boxplot. A caixa correspondente a cada modelo representa a dispersão dos resultados da métrica em questão para o modelo (lembre-se que são 159 janelas). Olhando para o hit, nota-se que o AR(3) e AR(4) tem mediana de hit próximas a 0.2 (lembre-se de que 0.2 que seria o ideal, já que estamos usando os percentis P10 e P90 como banda inferior e superior, respectivamente). À medida em que aumentamos a ordem, maior fica a mediana do hit. O mesmo ocorre com o quartil superior, que também aumenta com o aumento da ordem.

Em relação à perda quantílica, o aumento da ordem leva a um aumento da dispersão, com caixas maiores. Há uma pequena variação da mediana da perda quantílica entre os modelos, mas chama mais atenção o fato de que o quartil superior para AR(1) é mais baixo que os demais, bem como a caixa é bem pequena. Quanto ao erro médio absoluto, os modelos são muito próximos, sugerindo que em termos de aproximação entre P50 e realização, nenhum dos modelos se destaca.

A seleção do modelo deve combinar a avaliação das métricas e o nível de complexidade. No caso em questão, não há qualquer motivo para utilizar um modelo de ordem mais alta que 5, na medida em que o próprio AR(1) é mais simples e tem uma performance similar, e até superior em alguns pontos, aos demais.

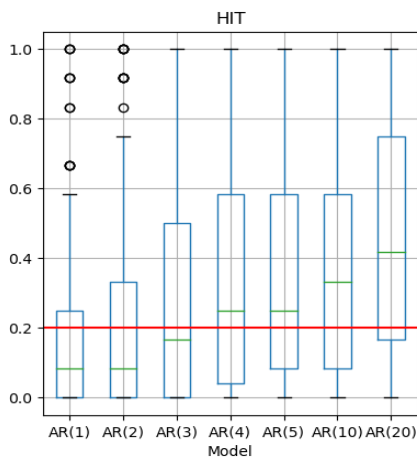


Figura 4 Comparação de modelos utilizando o HIT do intervalo P10-P90

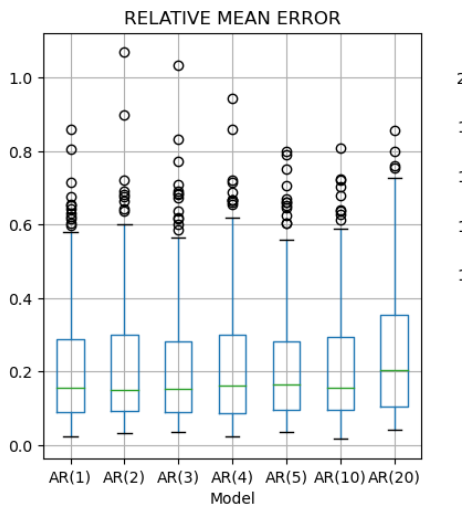


Figura 5 Comparação de o erro médio relativo

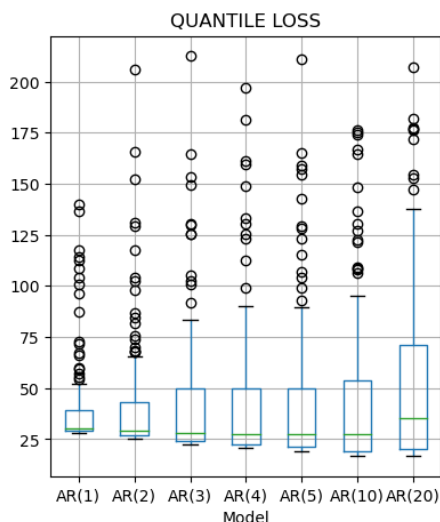


Figura 6 Comparação de modelos utilizando a perda quantílica

É sempre preferível ter modelos mais simples, por várias razões, entre as quais cabe destacar:

Interpretabilidade: Modelos simples são mais fáceis de entender e interpretar. Eles tendem a ter menos parâmetros e relações diretas entre as variáveis, o que facilita a compreensão das relações causais e a explicação do processo de tomada de decisão.

Evitar overfitting: Modelos complexos têm maior probabilidade de se ajustarem em excesso aos dados de treinamento, o que significa que eles podem se tornar muito sensíveis a pequenas variações nos dados de treinamento, resultando em um desempenho inferior em dados novos e desconhecidos. Modelos mais simples, por outro lado, têm menor tendência a overfitting e podem ter melhor capacidade de generalização.

Eficiência computacional: Modelos mais simples geralmente requerem menos recursos computacionais para treinar e simular. Isso pode ser importante em situações em que há limitações de tempo ou recursos.

Requisitos de dados: Modelos mais simples podem ser mais adequados quando há disponibilidade limitada de dados. Modelos complexos geralmente exigem mais dados para estimar com precisão os parâmetros e evitar overfitting.

Outro aspecto relevante é a estabilidade de parâmetros. Especialmente em modelos estocásticos, não se espera que mudanças bruscas ocorram nos parâmetros do modelo, como o da Equação (1) em virtude da adição de uma pequena quantidade de dados. Devido a problemas numéricos, pode-se ter problemas de estabilidade dos parâmetros, para os quais o analista deve estar atento. Muitas vezes, mesmo os resultados parecendo coerentes, podem estar ocorrendo variações muito grandes nos parâmetros dos modelos, o que é um sinal de alerta importante para a revisão do modelo e do processo de estimação.

No caso concreto deste estudo, é importante notar que ordens maiores não levam a melhores métricas e aumentam a complexidade do modelo. A Figura 7 exemplifica o impacto do aumento da ordem no formato dos percentis. Note que com o aumento de ordem, o preço depende de mais preços passados, o que faz com que as estimativas fiquem menos suaves, conforme pode ser verificado.

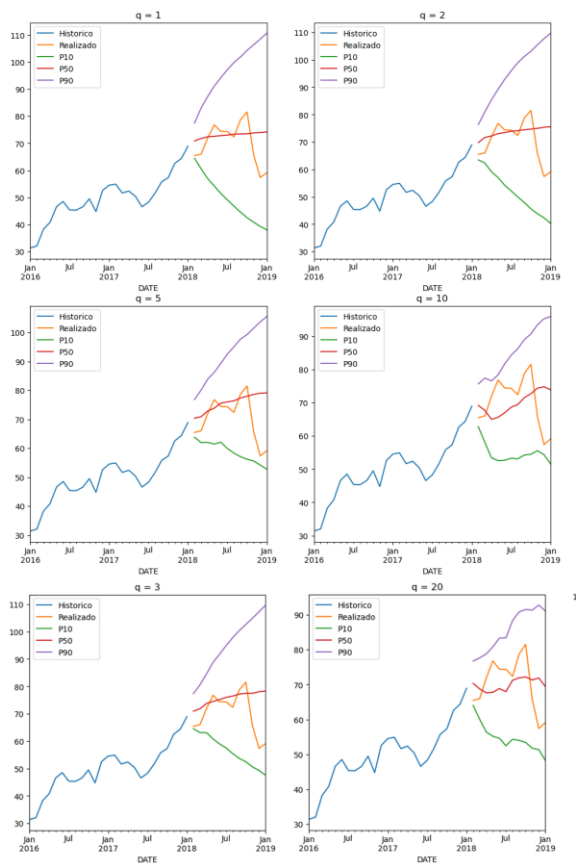


Figura 7 Exemplo de variação do formato das projeções a depender da ordem

Note que esse aumento de complexidade não se reflete em melhor qualidade da projeção, uma vez que as métricas do backtest não melhoram.

Dessa forma, embora a mediana de hit seja baixa (menos de 0.1), veja que o quartil superior é um pouco acima de 0.2. Dessa forma, pode-se concluir que provavelmente as bandas sejam mais largas que o necessário utilizando-se o AR(1). No entanto, por ser um modelo mais simples e não haver grandes diferenças de performance entre AR(1), AR(2) e AR(3), a escolha de AR(1) é justificável.

Não se justifica, por exemplo, escolher o AR(20), que é mais complexo e tem uma performance pior quando olhamos as métricas de backtest.

CONCLUSÕES

Este artigo apresenta o problema da avaliação de performance de modelos utilizados para a previsão probabilística de preços de ativos, principalmente commodities, que são utilizados na análise de risco em projetos de investimento. Por se tratar de produtos com preços negociados em mercados internacionais e de alta volatilidade, projetos que sofrem grande influência dos preços de commodities, como os projetos de desenvolvimento da produção na área de

Óleo e Gás, precisam constantemente gerar simulações estocásticas para suportar a análise de riscos. Uma vez que é necessário garantir que o modelo esteja bem calibrado para representar a dinâmica de preços, é importante que sejam realizadas análises de performance periódicas desses modelos.

No texto, apresentamos alguns detalhes de uma metodologia de backtest para modelos de projeção de preços de ativos financeiros, exemplificando com um caso prático através da modelagem do petróleo Brent utilizando sete modelos autorregressivos. O backtest é essencial para escolher o modelo mais adequado para realizar as projeções probabilísticas. Foram sugeridas algumas métricas de performance, havendo espaço para muitas outras.

Além dos resultados em relação às métricas de backtest, é essencial que a complexidade dos modelos seja avaliada antes da seleção, optando-se, sempre que possível, por modelos mais simples em detrimento de outros mais complexos. Modelos de Machine Learning, que tipicamente têm muitos parâmetros, devem ser utilizados somente se sua performance for superior à de modelos mais simples, como os apresentados neste documento.

Outro aspecto importante a ser considerado é o fato de que uma performance boa em backtest indica apenas que o modelo representa bem o comportamento probabilístico dos preços no passado. Portanto, para que uma análise de riscos seja completa, é de fundamental importância que sejam realizadas análises de sensibilidade quanto a mudanças estruturais que possam ocorrer. No setor de Óleo e Gás, por exemplo, a transição energética pode impor mudanças significativas na demanda, e conseqüentemente no preço, tanto do petróleo cru quanto nos crack spreads dos derivados. Razão pela qual é fundamental realizar uma análise crítica dos resultados obtidos a luz de análises fundamentalistas do futuro.

A complexidade dos modelos é também um fator relevante para a seleção de modelos, sendo preferível os modelos com menos parâmetros e menos complexo. Modelos mais simples só devem ser substituídos por modelos mais complexos se a performance retrospectiva compensar. O mesmo deve ser observado com a utilização de modelos de Machine Learning, que normalmente possuem mais parâmetros que os modelos tradicionais. Em algumas situações observa-se uma preferência por utilizar modelos de Machine Learning simplesmente por questões de promoção da técnica e pela roupagem de estado da arte que o uso dessas técnicas traz. No entanto, o analista deve sempre preferir modelos mais simples e consolidados, migrando para versões mais sofisticadas apenas se tiverem melhor performance e generalidade.



REFERÊNCIAS

- AIUBE, F. A. L. Modelos Quantitativos em Finanças: Com Enfoque em Commodities, Bookman, 2012.
- AIUBE, F. A. L., BAIDYA, T. K. N. and TITO, E. A. H. Analysis of commodity prices with the particle filter, *Energy Economics* 30(2): 597–605. URL:<https://www.sciencedirect.com/science/article/pii/S0140988306000880>, 2008.
- CHO, K. et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724-1734). Association for Computational Linguistics, 2014.
- LU, J. and DING, D. A hybrid approach on conditional gan for portfolio analysis, 2022.
- MAHALANOBIS, P. On the generalized distance in statistics, *Proceedings of the National Institute of Sciences (Calcutta)*, 1936.
- MOHAMMED, S. Research on financial risk prevention and control methods based on big data, 2019.
- SCHWARTZ E. S. and SMITH, J. E. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911, 2000.
- VENTURINI, M. and ALEJANDRO, G. Statistical distances and probability metrics for multivariate data, ensembles and probability distributions. URL:<https://api.semanticscholar.org/CorpusID:124313030>, 2015.
- GEISSLER, C., MORIZET, N., RIZZATO, M. and WALLART, J. Generative adversarial networks applied to synthetic financial scenarios generation. GEISSLER, C., MORIZET, N., RIZZATO, M. and WALLART, J. (2022b). Generative adversarial networks applied to synthetic financial scenarios Generation, 2022.
- YOON, J., JARRETT, D. and VAN DER SCHAAAR, M. Time-series generative adversarial networks, 32, 2019.