

PREVISÃO FINANCEIRA CIENTE DE REGIMES: UM MODELO HÍBRIDO MARKOV-TRANSFORMER PARA PREDIÇÃO MULTI-ATIVOS DE RISCO E RETORNO

Rodrigo Stachlewski Palma ¹

Resumo: Este trabalho apresenta uma abordagem híbrida para previsão de séries temporais financeiras, geralmente não estacionárias, combinando modelos de regime de *Markov* com *Transformer* de aprendizado profundo (*Deep Learning*). Propomos o *Regime-Aware Transformer-HMM (RAT-HMM)* – trata-se de um modelo capaz de detectar os regimes de mercado usando Modelo de Markov Oculto (*HMM*) e, em sequência, utiliza-se dessa informação de regime dentro de uma arquitetura *Transformer* a fim de melhorar a previsão risco/retorno de múltiplos ativos financeiros (incluimos o *Bitcoin (BTC)*, *Ethereum (ETH)*, *S&P 500* e o Ouro). A motivação central é enfrentar a não estacionariedade dos mercados financeiros, onde ocorrem mudanças abruptas (“quebras de regime”) desafiando os modelos de previsão. Revisamos criticamente algumas abordagens de aprendizado profundo (*LSTM/GRU*). Em seguida, apresentamos o *RAT-HMM* e suas capacidades de modelar padrões sequenciais complexos das séries temporais. Em nossa avaliação empírica com dados reais, o *RAT-HMM* obteve alguns resultados muito significativos em especial com ativos com alta volatilidade.

Palavras-chave: previsão financeira, regimes de mercado, *transformer*, *hidden markov model (HMM)*

Abstract: This study presents a hybrid approach for forecasting financial time series, which are typically non-stationary, by combining regime-switching Markov models with deep learning-based Transformer architectures. We propose the Regime-Aware Transformer-HMM (RAT-HMM) - a model capable of detecting market regimes using a Hidden Markov Model (HMM) and subsequently incorporating this regime information into a Transformer architecture to enhance the multi-asset risk-return forecasting performance. Our target assets include Bitcoin (BTC), Ethereum (ETH), the S&P 500 Index, and Gold. The central motivation lies in addressing the non-stationarity of financial markets, where abrupt changes (“regime shifts”) pose significant challenges to predictive models. We critically review several deep learning baselines (LSTM/GRU) and then present the RAT-HMM and its capacity to model complex sequential patterns in time series data. In our empirical evaluation with real market data, the RAT-HMM demonstrates notably strong predictive performance, particularly for high-volatility assets.

Keywords: financial forecasting, market regimes, transformer, hidden Markov model (HMM)

¹ Pontifícia Universidade Católica do Rio Grande do Sul. E-mail: r.stachlewski@edu.pucrs.br.

1. INTRODUÇÃO

Nas últimas décadas, tentar prever o preço e riscos do mercado financeiro tem desafiado os pesquisadores devido à natureza não estacionária destes mercados. As estatísticas descritivas dos ativos financeiros frequentemente se modificam ao longo do tempo, refletindo as constantes mudanças dos cenários macroeconômicos, crises e momentos eufóricos (HAMILTON, 1989). Essa quebra de pressupostos de estacionariedade pode reduzir a acurácia dos modelos treinados em um determinado período quando o mercado entra em um novo regime (por exemplo, quando mercado saiu de um momento pessimista “*bear market*” para um otimista “*bull market*” (CHANG; CHOI; PARK, 2014). Tal comportamento desafia técnicas tradicionais e exige abordagens adaptáveis e cientes de mudanças de regime.

2. FUDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS

2.1 Modelos Ocultos de Markov

HMMs permitem representar processos estocásticos com mudanças discretas de estado não observáveis sendo amplamente utilizados em finanças para detecção de regimes de volatilidade e retorno (Hamilton, 1989).

2.2 Transformer em Séries Temporais

Desde sua introdução (Vaswani et al., 2017) *Transformers* revolucionaram a modelagem sequencial, com sucesso expressivo em previsão multivariada de séries financeiras (Zerveas et al., 2021). Ao contrário das Redes Neurais Recorrentes (*RNN*), as primeiras permitem paralelismo e captação de dependências de longo prazo via mecanismos de atenção.

2.3 Modelos Híbridos e Ciente de Regimes (*Regime-Aware*)

Modelos que integram múltiplas fontes de informação como o *Relation-Aware Transformer* (Li et al., 2020) mostram avanços ao explorar estruturas internas entre variáveis. O *RAT-HMM* estende essa ideia ao incorporar os conhecimentos sobre o regime de mercado como parte da representação aprendida, tornando-se especialmente útil com ativos com alta volatilidade estrutural, como, por exemplo, o Bitcoin.

3. O MODELO RAT-HMM

Propomos o Regime-Aware Transformer-HMM (RAT-HMM), um modelo híbrido que integra o Modelo de Markov Oculto com um Transformer sequencial para previsão financeira. Logo, *Hidden Markov Model (HMM)* analisa a série temporal e identifica os regimes ocultos de mercado. O *HMM* assume que os dados são gerados por um processo alternado de estados discretos (regimes) não observados diretamente (por exemplo, um estado pode representar um regime de alta volatilidade, retornos negativos). Portanto, o *HMM* infere a probabilidade de a série temporal estar em cada estado com suas características estatísticas de cada regime.

Já o *Transformer* realiza a previsão da série temporal utilizando-se de mecanismos de auto-atenção para capturar dependências complexas. No *RAT-HMM*, o *Transformer* recebe o input, além das variáveis exógenas convencionais (ex.: preços, retornos, volume e outros indicadores), informações sobre o regime de mercado atual inferido pelo *HMM*. Com isso o *Transformer* torna-se “ciente do regime”, podendo ajustar suas atenções e provisões conforme contexto predominante.

3.1 Estrutura Geral

O *RAT-HMM* combina dois componentes:

- *HMM* de 6 regimes – que aprende as distribuições de retornos e volatilidades, gerando uma série de estados latentes com finalidade de capturar as fases do mercado;
- *Transformer* – que recebe como input dos dados financeiros das séries temporais, bem como os features técnicos e a sequência de regimes detectados.

3.2 Pipeline de Dados

É composto por janelas deslizantes de 30 dias, com *Horizon target* de 5, 10 e 15 dias. As features são normalizadas por *rolling z-score* e as variáveis exógenas incluem regime, log-volatilidade e retorno acumulado.

3.3 Treinamento e Validação

Foi utilizado validação cruzada com divisão temporal (*TimeSeriesSplit*). E a função de perda utilizada foi a *MSE (Mean Squared Error)* - métrica que calcula a média dos quadrados

das diferenças entre valores previstos e os valores reais. Já para treinamento foi utilizado o PyTorch com *scheduler* de aprendizado e *early stopping*.

4 RESULTADOS

Foram conduzidos experimentos de previsão de retornos financeiros de quatro (4) ativos financeiros – Bitcoin (BTC), Ethereum (ETH), S&P 500 (SPX) e o Outro (GOLD) – em três horizontes de previsão, r5d, r10d e r15d, correspondentes aos retornos acumulados de 5, 10 e 15 dias, respectivamente. Para fins de avaliação de desempenho do modelo RAT-HMM foram utilizados os modelos de referência (LSTM, GRU e Transformer) para cada ativo e horizonte, como e sem a utilização do *HMM*. Amplamente documentados em nosso repositório no GitHub: <https://github.com/Rodrigo-Palma/RAT-HMM>.

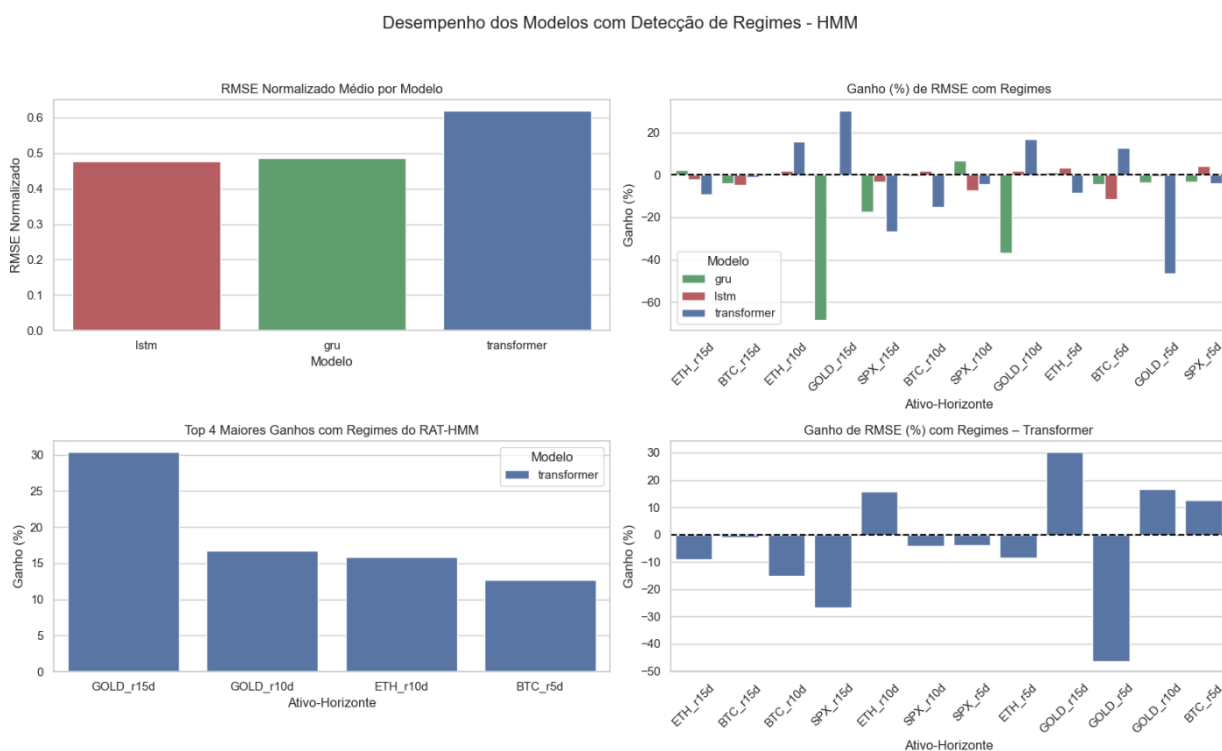


Figura 1

A Figura 1 resume os principais resultados dos experimentos com o modelo *RAT-HMM*, comparando com diferentes arquiteturas (*GRU*, *LSTM* e *Transformer*) com ou sem inferência de regimes provida pelo *HMM*. Os resultados foram analisados em termos do RMSE médio normalizado e estão em valores percentuais (%).

O painel superior esquerdo mostra o RMSE médio normalizado por tipo de modelo. Observa-se que o modelo LSTM obteve a menor média de erro ($\text{RMSE} \approx 0,48$), seguido do GRU ($\text{RMSE} \approx 0,49$). O Transformer apresentou o maior RMSE médio ($\approx 0,61$), porém, como será discutido, foi o modelo com os maiores ganhos proporcionados pela inclusão dos regimes latentes.

No painel superior direito, observa-se os ganhos e perdas dos três modelos ao utilizarem ou não a inferência de regimes através do *HMM*, em diferentes ativos e horizontes de retornos acumulados. Nota-se variações significativas de ganhos, tanto positivas quanto negativas, refletindo a sensibilidade da abordagem à exposição as séries temporais financeiras.

É importante introduzir o conceito de R^2 (coeficiente de determinação) é uma métrica estatística para avaliar desempenho de modelos de regressão. Ele varia entre 0 e 1, sendo 1 previsão perfeita (modelo explica 100% da variância) e 0 não explica nada da variância.

A seguir, apresentamos os melhores resultados de ganho percentual observados, majoritariamente, no modelo *RAT-HMM*.

GOLD_r15d (*RAT-HMM*): ganho de 30,38% com $R^2 = 0,78$;

GOLD_r10d (*RAT-HMM*): ganho de 16,73% com $R^2 = 0,66$;

ETH_r10d (*RAT-HMM*): ganho de 15,83% com $R^2 = 0,81$;

BTC_r5d (*RAT-HMM*): ganho de 12,73% com $R^2 = 0,50$;

SPX_r10d (*GRU*): ganho de 6,67% com $R^2 = 0,77$.

O destaque é para o ativo ouro (*GOLD*), indicando que o modelo é capaz de capturar transições de regime relevantes neste mercado, em especial em janelas mais longas, como 10 e 15 dias. Tal desempenho pode estar relacionado à natureza desse ativo e sua forte correlação com ciclos econômicos, padrões que o *HMM* visa capturar.

Por outro lado, alguns cenários apresentaram degradação do desempenho com a introdução da inferência de regimes do *HMM*, como evidenciado nas barras inferiores nos gráficos.

Destacamos os cinco piores casos:

GOLD_r15d (GRU): perda de -68,36% com $R^2 = 0,56$;

GOLD_r5d (RAT-HMM): perda de -46,35% com $R^2 = 0,04$;

GOLD_r10d (GRU): perda de -36,77% com $R^2 = 0,54$;

SPX_r15d (RAT-HMM): perda de -26,72% com $R^2 = 0,61$;

SPX_r15d (GRU): perda de -17,45% com $R^2 = 0,78$.

Esses resultados indicam que, embora o modelo *RAT-HMM* possa obter ganhos expressivos, ele também pode introduzir ruído preditivo em contextos onde a identificação dos regimes não contribui significativamente para a variabilidade do ativo. Em particular, os baixos valores em R^2 sugerem que a inferência de regimes nesses casos não capturou uma estrutura informativa útil para a previsão. Importante salientar que o modelo *Transformer* foi o único que realmente tirou proveito da inferência de regimes introduzida pelo *HMM*. O modelo LSTM não apresentou ganhos ou perdas significativas. E o GRU apresentou degradação significativa.

Abaixo, na figura 2, apresentamos uma comparação visual entre as previsões realizadas pelo modelo *Transformer* no horizonte de 15 dias para o ativo *GOLD* como e sem a utilização da inferência de regimes. Este representa o melhor caso observado em nossa avaliação empírica, com ganho percentual superior a 30% na redução do erro do modelo (*RMSE*).

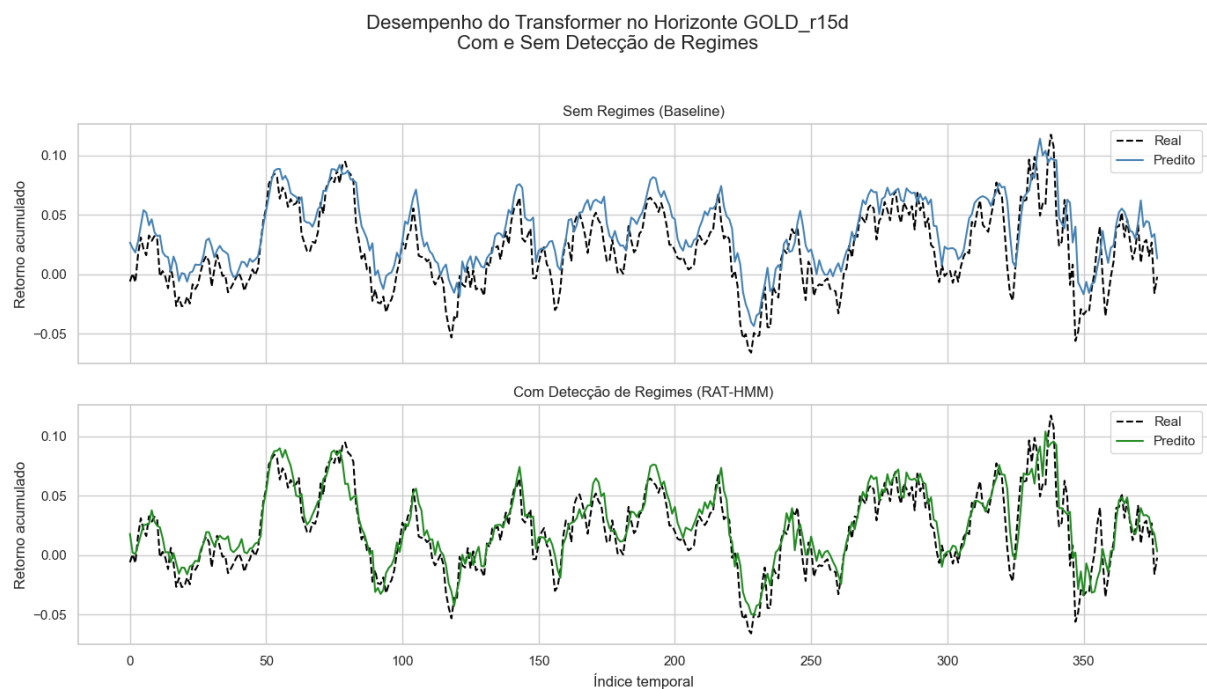


Figura 2.

Esse resultado evidencia o potencial do modelo híbrido *RAT-HMM*, em mudanças de contexto e alta volatilidade, sendo especialmente eficaz com ativos com dinâmicas complexas e sensíveis a ciclos econômicos, como é o caso do ouro.

5 DISCUSSÃO

Os resultados acima demonstram que a aplicação do modelo híbrido *RAT-HMM* evidencia sua eficácia preditiva em contextos complexos como do mercado financeiro caracterizados pela não estacionariedade e transações estruturais. A análise empírica conduzida sobre quatro ativos de natureza distinta – *Bitcoin (BTC)*, *Ethereum (ETH)*, *S&P 500 (SPX)* e *Ouro (GOLD)* – revela que a incorporação de regimes latentes via *Hidden Markov Models (HMM)* é capaz de melhorar significativamente a capacidade preditiva do modelo *Transformer*.

O *RAT-HMM* incorpora um mecanismo de *Transformer* com reconhecimento de regimes dos modelos de Markov Oculto (*Regime-Aware*) entre diferentes ativos. Essa capacidade de compreender correlações e interdependências se mostrou valiosa na melhoria das previsões.

Outro ponto a ser destacado é a robustez do *RAT-HMM* mesmo em diferentes condições de volatilidade. Observou-se que o modelo foi capaz de generalizar mesmo em condições de alta volatilidade e turbulência. Isso sugere que os estados latentes (regimes) do *HMM* conseguiram diferenciar momentos de forte estresse no mercado financeiro.

Cabe discutir também que o desempenho relativo entre os benchmarks dos modelos utilizados nos experimentos sem o *HMM*, obtiveram desempenho competitivo para alguns casos (particularmente em horizontes de curto prazo e médio prazo), por vezes, inclusive superando o *RAT-HMM*.

Adicionalmente, observa-se que o modelo *LSTM* apresentou estabilidade nos resultados, com pouca variação entre a versão com e sem regimes. Isso pode ser atribuído à sua arquitetura mais conservadora, que tende a suavizar variações abruptas, mas ao custo de menor sensibilidade a dinâmicas de mudança de regime. E que o modelo *GRU* não se beneficiou do uso da inferência dos regimes, inclusive prejudicando sua performance na maioria dos casos.

É importante considerar que nem todos os cenários apresentaram desempenho satisfatório. Casos como *GOLD_r15d (GRU)* e *GOLD_r5d (Transformer)* evidenciaram perdas significativas de desempenho (com reduções no *RMSE* superiores a -68% e -46%, respectivamente), sugerindo que a inferência de regimes pode introduzir ruído preditivo quando os regimes não capturam variações relevantes ou introduzem mudanças espúrias. O baixo valor de R^2 em alguns desses casos, como o *GOLD_r5d (Transformer)* com apenas 0,04, reforça a hipótese de que a estrutura latente inferida não trazia valor informativo para a tarefa preditiva naquele contexto.

Em resumo, a abordagem híbrida proposta mostra-se promissora, mas sua aplicação requer cuidados na modelagem dos regimes e na escolha do ativo e horizonte preditivo.

Do ponto de vista prático, o *RAT-HMM*, conseguiu extrair o melhor dos dois mundos a capacidade de aprender padrões complexos e relações interativas via *Transformer*, e a flexibilidade para mudanças estruturais via *HMM*. E, com isso, enfrentou de forma eficaz desafios típicos na previsão dos mercados financeiros, como a não-estacionariedade, as quebras de regimes e mudanças abruptas recorrentes em mercados financeiros.

6 CONCLUSÃO

Este trabalho apresentou o *RAT-HMM (Regime-Aware Transformer – Hidden Markov Model)*, um modelo híbrido inovador para previsão de risco e retorno em mercados financeiros. Através de uma análise empírica envolvendo diversas classes de ativos (criptomoedas, bolsa de valores, commodities) em múltiplos horizontes de tempo. Ainda, faz-se necessário, mais pesquisa, otimização. Contudo, o uso de modelos híbridos cientes de regime pode trazer redução significativa nos erros de previsão, potencializando os lucros dos investidores.

Outrossim, considerando que a arquitetura *Transformer* possui capacidades únicas em capturar as dependências de longo alcance em sequências de dados, escalável, de forma paralela – o tempo de treinamento e de inferência pode ser acelerado drasticamente, beneficiando o uso em conjuntos massivos de dados e onde o tempo é crucial para maximização de lucros potenciais como nos mercados financeiros.

Em conclusão, o *RAT-HMM* apresenta ganhos significativos em arquiteturas híbridas com utilização de *HMM*, não somente na arquitetura *Transformer*, mas também em outros

modelos de aprendizado profundo como *LSTM* e *GRU*. Mostrando-se resiliente, mesmo em diversos cenários e ativos, em especial os que possuem volatilidade maiores, bem como capacidade de realizar previsões em horizontes mais longos, como podemos verificar no caso dos retornos acumulados de quinze (15 dias). Em Trabalhos futuros, poderemos explorar a extensão do *RAT-HMM* para horizontes ainda mais amplos ou frequências de dados mais altas, bem como investigar a integração de outras variáveis, como indicadores macroeconômicos, mais indicadores técnicos e análise de sentimentos de notícias. Também seria interessante avaliar o desempenho deste modelo em outros cenários de séries temporais, além do mercado financeiro. De toda forma, os resultados obtidos até então contribuem para aprimorar estratégias quantitativas e gestão de riscos no domínio financeiro.

AGRADECIMENTOS

O presente artigo foi alcançado em cooperação com a **HP Brasil Indústria e Comércio de Equipamentos Eletrônicos LTDA.** e com recursos provenientes da Lei de Informática (Lei nº 8.248, de 1991).

REFERÊNCIAS

- HAMILTON, James D. *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*. *Econometrica*, v. 57, n. 2, p. 357–384, 1989. Disponível em: <https://users.ssc.wisc.edu/~bhansen/718/Hamilton1989.pdf>. Acesso em: 1 jul. 2025.
- CHANG, Y.; CHOI, Y.; PARK, J. Y. *Regime Switching Model with Endogenous Autoregressive Latent Factor*. Yale University Department of Economics Working Paper, 2014. Disponível em: <https://economics.yale.edu/sites/default/files/rsalf-2014-3-15.pdf>. Acesso em: 1 jul. 2025.
- LI, Zhe; ZHANG, Zhou; YU, Qin; ZHAO, Wanli; ZHANG, Qi. *Relation-Aware Transformer for Portfolio Policy Learning*. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI-20), 2020. Disponível em: <https://www.ijcai.org/proceedings/2020/0641.pdf>. Acesso em: 3 jul. 2025.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. *Attention is All You Need*. In: Advances in Neural Information Processing Systems (NeurIPS), 2017. Disponível em: https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 3 jul. 2025.
- ZERVEAS, G.; JAYARAMAN, S.; PATERSON, S.; BOGLE, J.; CONG, Y.; ANGELOV, P. *A Transformer-based Framework for Multivariate Time Series Representation Learning*. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD), 2021. Disponível em: <https://arxiv.org/abs/2010.02803>. Acesso em: 3 jul. 2025.