

FRAMEWORK MULTIVARIADO PARA FOREX: COINTEGRAÇÃO, CLUSTERS, T² E APRENDIZADO DE MÁQUINA

HUGO A. A. J. PONTES – hugopontes@unifei.edu.br
PEDRO P. BALESTRASSI – pedro@unifei.br
MATHEUS BRENDON – matheus_brendon@unifei.edu.br

Palavras-chave: Forex, mercado de câmbio, inteligência artificial, previsão de séries temporais, redes neurais.

1. INTRODUÇÃO

Este trabalho, na modalidade “em desenvolvimento”, tem por objetivo propor e justificar uma linha de investigação para previsão e detecção de desequilíbrios no mercado de câmbio, motivada pela alta liquidez, volatilidade e mudanças de regime que desafiam abordagens convencionais. Insere-se na literatura de séries financeiras e aprendizado de máquina, buscando organizar um caminho de pesquisa reprodutível e útil à tomada de decisão quantitativa. Em termos gerais, delineia-se uma arquitetura integrada que combina fundamentos estatísticos, organização de ativos por similaridade e modelagem temporal baseada em aprendizado de máquina. Os fundamentos e os procedimentos são apresentados nas seções seguintes.

2. DESENVOLVIMENTO

2.1 Referencial teórico

A literatura de séries financeiras descreve fatos estilizados no câmbio — alta volatilidade, caudas pesadas, assimetria e mudanças de regime — que limitam modelos lineares simples e motivam abordagens robustas (MANDELBROT, 1963; HAMILTON, 1994; CONT, 2001; TSAY, 2010). A previsibilidade, quando existente, é tipicamente condicional e sensível ao horizonte e ao desenho de validação, no contexto do debate sobre eficiência informacional (FAMA, 1970; LO; MACKINLAY, 1999). A heteroscedasticidade condicionada levou ao uso de modelos de variância (ENGLE, 1982).

Para relações entre ativos, a cointegração oferece um arcabouço para identificar equilíbrios de longo prazo entre séries integradas, dos quais podem emergir desequilíbrios transitórios (HAMILTON, 1994). A organização dos ativos por similaridade (clusters) ajuda a estruturar universos coerentes. Para monitorar desvios em múltiplas variáveis, a estatística T^2 de Hotelling — uma medida multivariada baseada na distribuição F — considera covariâncias e costuma ser superior a limiares univariados (BOX et al., 2015; BALESTRASSI et al., 2025).

No campo de modelagem, redes neurais recorrentes (RNNs, *Recurrent Neural Networks*), variantes de memória longa (LSTM, *Long Short-Term Memory*), unidades recorrentes simplificadas (GRU, *Gated Recurrent Units*), convoluções temporais (TCN, *Temporal Convolutional Networks*) e arquiteturas com mecanismos de atenção (Transformers) têm sido empregadas para capturar dependências temporais e não linearidades, com ênfase em validação temporal adequada e reprodutibilidade (HYNDMAN; ATHANASOPOULOS, 2018; TSAY, 2010).

2.2 Metodologia

Materiais e dados: séries temporais de pares/grupos de moedas do mercado de câmbio (Forex), com frequência diária (ou intradiária, quando disponível), abrangendo período suficientemente longo para estimar estabilidade de relações. As séries são obtidas de provedores públicos/mercado, padronizadas (timezone, moeda-

base), tratadas para valores ausentes e consolidadas em um repositório versionado com metadados de coleta.

Pré-processamento: cálculo de log-retornos e padronização por janela móvel; verificação de estacionariedade em retornos (testes ADF) e inspeção de dependências de curto prazo via funções de autocorrelação (ACF) e autocorrelação parcial (PACF) (TSAY, 2010; HYNDMAN; ATHANASOPOULOS, 2018). Identificação exploratória de regimes de volatilidade por quantis e modelos de Markov ocultos (HMM, Hidden Markov Models) para segmentar períodos com risco distinto (HAMILTON, 1994).

Relações de equilíbrio e seleção de spreads: triagem de cointegração em pares/grupos para formar spreads estacionários que representem equilíbrios de longo prazo (HAMILTON, 1994). Organização do universo de ativos por clusters (similaridade de dinâmica) para reduzir heterogeneidade e estruturar o espaço de busca.

Monitoramento multivariado: construção de vetores informativos (nível/retorno do spread, volatilidade local e variáveis auxiliares). Aplicação da estatística T^2 de Hotelling para detecção de desvios multivariados significativos com limites de controle derivados da distribuição F, mais adequado do que limiares univariados por incorporar covariâncias (BOX et al., 2015).

Modelagem temporal: estimação de modelos para prever retornos/deltas e/ou a magnitude esperada dos desvios, empregando redes recorrentes (LSTM/GRU), convolucionais causais (TCN) e Transformers com atenção causal para dependências de longo alcance. O alvo é definido em retorno ou delta percentual em horizonte de previsão H (ex.: 1–5 passos) para reduzir suavização média e melhorar acerto direcional (TSAY, 2010; HYNDMAN; ATHANASOPOULOS, 2018).

Validação e métricas: esquema walk-forward com janelas deslizantes — treino e validação no passado, teste no trecho mais recente — evitando vazamento temporal (HYNDMAN; ATHANASOPOULOS, 2018). Relato de métricas ponto a ponto (Mean Absolute Error – MAE, Root Mean Squared Error – RMSE), direcionais (acerto direcional e coeficiente de correlação de Matthews – MCC) e econômicas (retorno, drawdown e índice de Sharpe). Comparação formal com baselines por teste de

diferença de previsões (Diebold–Mariano test) quando aplicável. Para detecção, comparação entre a estatística T^2 de Hotelling e métodos univariados (Z-score/Bollinger) sobre spreads cointegrados.

3. RESULTADOS E DISCUSSÃO

Nesta modalidade em fase de desenvolvimento, apresentam-se resultados esperados e como serão interpretados à luz dos objetivos do estudo.

Desempenho preditivo esperado: arquiteturas com atenção (Transformers) tendem a superar LSTM/GRU na captura de dependências de mais longo alcance, sobretudo quando o alvo é retorno/delta percentual e há variáveis exógenas, refletindo menores erros ponto-a-ponto (MAE/RMSE) e maior acerto direcional (HYNDMAN; ATHANASOPOULOS, 2018; TSAY, 2010).

Detecção multivariada vs. univariada: em spreads cointegrados, espera-se que a detecção baseada na estatística T^2 de Hotelling produza menor taxa de falsos alarmes e maior sensibilidade a desvios relevantes do equilíbrio, em comparação a métodos univariados (Z-score/Bollinger) (BOX et al., 2015; HAMILTON, 1994).

Robustez e estresse de cenários: diante de caudas pesadas, clustering de volatilidade e saltos — fatos estilizados do câmbio — a combinação “cointegração + T^2 + Transformer” deve manter desempenho estável em testes walk-forward e em cenários sintéticos com GARCH/saltos (CONT, 2001; HYNDMAN; ATHANASOPOULOS, 2018).

Métricas econômicas: espera-se melhoria em retorno ajustado ao risco (maior Sharpe, menor drawdown) quando os sinais de arbitragem forem disparados por desvios multivariados confirmados e por previsões consistentes do componente direcional.

Comparações e significância: diferenças entre modelos/sinais serão discutidas com testes formais (p.ex., Diebold-Mariano para previsões) e análise de sensibilidade a janelas, horizontes e atributos (HYNDMAN; ATHANASOPOULOS, 2018).

Ameaças à validade e limitações: riscos de sobreajuste, instabilidade de cointegração, custos de transação, latência/execução e dependência de dados serão tratados com validação temporal, backtests conservadores e documentação reprodutível.

4. CONSIDERAÇÕES FINAIS

Este trabalho, em fase de desenvolvimento, sintetizou a motivação e o enquadramento teórico para investigar previsão e detecção de desequilíbrios no mercado de câmbio, com ênfase em reprodutibilidade e avaliação adequada. Propõe-se um caminho de pesquisa que integra relações de equilíbrio de longo prazo, monitoramento multivariado e modelagem temporal com aprendizado de máquina, mantendo o foco na utilidade prática para a tomada de decisão quantitativa.

Como resultados esperados, antecipa-se que a combinação entre seleção de pares/grupos cointegrados, detecção de desvios via estatística T^2 de Hotelling e modelos com atenção (Transformers), comparados a baselines usuais, produza previsões mais fiéis à dinâmica do câmbio, melhor acerto direcional e sinais de arbitragem mais informativos em testes walk-forward. Limitações previstas — como mudanças estruturais, custos de transação e risco de sobreajuste — serão enfrentadas com validações temporais rigorosas, análise de sensibilidade e documentação completa para permitir replicação por outros pesquisadores.

AGRADECIMENTOS

Este trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) e desenvolvido no âmbito do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Itajubá (UNIFEI). Agradeço ao Instituto de Engenharia de Produção e Gestão (IEPG) pela promoção do SUMMIT e pelo incentivo à divulgação científica.

REFERÊNCIAS

- BOX, G. E. P. et al. Time series analysis: forecasting and control. 5. ed. Hoboken: Wiley, 2015.
- CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance, v. 1, n. 2, p. 223–236, 2001.
- DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. Journal of Business & Economic Statistics, v. 13, n. 3, p. 253–263, 1995.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, v. 50, n. 4, p. 987–1007, 1982.
- FAMA, E. F. Efficient capital markets: a review of theory and empirical work. The Journal of Finance, v. 25, n. 2, p. 383–417, 1970.
- HAMILTON, J. D. Time series analysis. Princeton: Princeton University Press, 1994.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. Forecasting: principles and practice. 2. ed. Melbourne: OTexts, 2018.
- LO, A. W.; MACKINLAY, A. C. A non-random walk down Wall Street. Princeton: Princeton University Press, 1999.
- MANDELBROT, B. The variation of certain speculative prices. The Journal of Business, v. 36, n. 4, p. 394–419, 1963.
- TSAY, R. S. Analysis of financial time series. 3. ed. Hoboken: Wiley, 2010.