

Harnessing artificial intelligence for monitoring financial markets

Matteo Aquilina, Douglas Araujo, Gaston Gelos, Taejin Park and Fernando Pérez-Cruz
Bank for International Settlements

This talk represents my views and not necessarily those of the BIS.

Can market dysfunction and stressed conditions be predicted in advance?

- A policy-relevant issue:
  - Episodes of market dysfunctions, from periods of sustained illiquidity to fully-fledged financial crises can have substantial economic effects.

- If supervisors could forecast when these episode are most likely to materialize, they could intervene in advance and reduce their severity (or even stop them from happening).

The task is complex

- The ability of econometric models to predict dysfunction and crises out of sample is limited (Fouliard et al, 2021)

- Financial markets are deeply interconnected among themselves and with the real economy
  - Risks can materialise in different ways
  - The system responds to changes implying that propagation channels evolve and are difficult to predict

- Observed episodes of severe financial stress are (thankfully) rare

What characteristics should a predictive tool have?

- Two important features:

    1. Ability to predict a wide range of dysfunctions in advance.

    2. Help authorities identify the potential source of future dysfunction
        - Combine ML models with generative AI (LLMs)

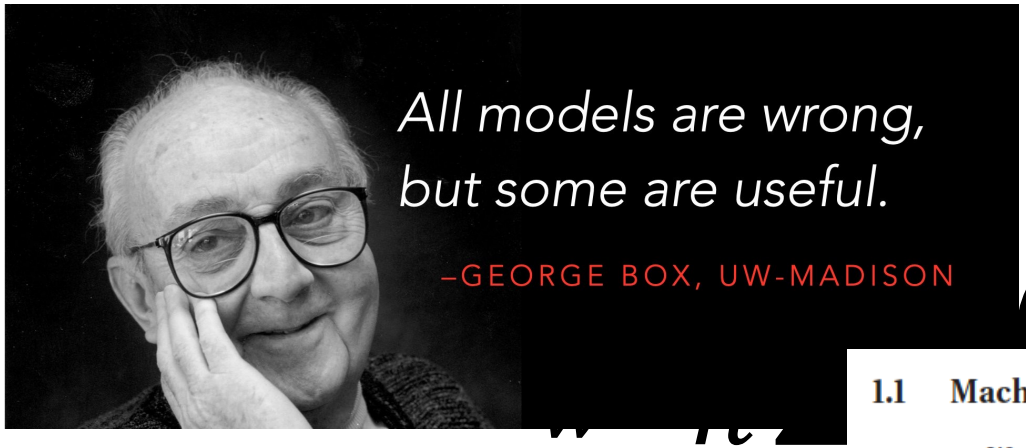
Why ML is not Statistics?

$$\min_{w} \frac{1}{n} \sum_{i=1}^{n} L\left(y_i, f_w(\boldsymbol{x}_i)\right)$$

$$\{(\boldsymbol{x}_i, y_i)\}_{i=1\ldots n} \underset{iid}{\sim} p_{X,Y}(\boldsymbol{x}, y)$$

Machine Learners and Statisticians solved the **same** problem

Why ML is not Statistics?



$$\{(x_i, y_i)\}_{i=1}$$

Machine Learners anc

1.1 Machine learning: what and why?

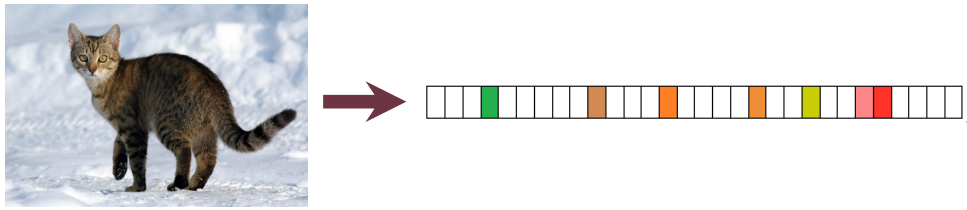
We are drowning in information and starving for knowledge. — John Naisbitt.

We are entering the era of **big data**. For example, there are about 1 trillion web pages[1]; one hour of video is uploaded to YouTube every second, amounting to 10 years of content every day[2]; the genomes of 1000s of people, each of which has a length of 3.8×10^9 base pairs, have been sequenced by various labs; Walmart handles more than 1M transactions per hour and has databases containing more than 2.5 petabytes (2.5×10^{15}) of information (Cukier 2010); and so on.

This deluge of data calls for automated methods of data analysis, which is what **machine learning** provides. In particular, we define machine learning as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!).

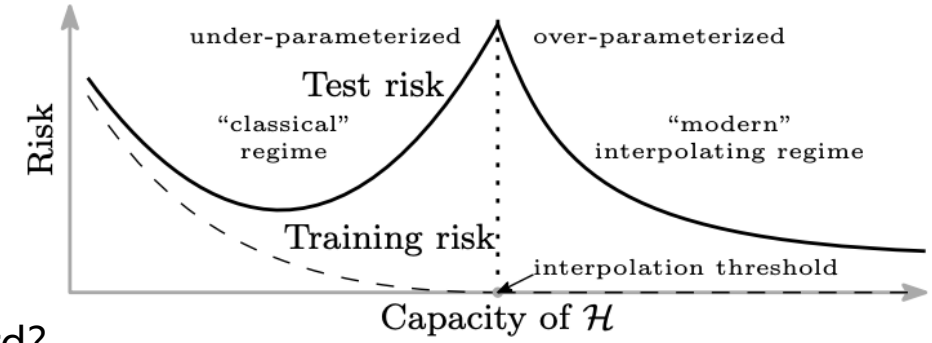
The key aspects that separate Machine Learning from Statistics

- Embeddings:
  - Transform non vectorial data into vectors.



- Over-parametrization:
  - x100 or x1000 more parameters than data.
- Data:
  - The more the merrier.
  - No diminishing returns.
- Self-supervised Learning:
  - Labels are expensive.
  - Data is not.

} Predict the next word?



- Zero-shot learning:
  - From prediction to generation.

Zero-shot

Zero-shot Behaviors We'd like to better understand why language model pre-training of transformers is effective. A hypothesis is that the underlying generative model learns to perform many of the tasks we evaluate on in order to improve its language modeling capability and that the more structured attentional memory of the transformer assists in transfer compared to LSTMs. We designed a series of heuristic solutions that use the underlying generative model to perform tasks without supervised finetuning. We visualize the effectiveness of these heuristic solutions over the course of generative pre-training in Fig 2(right). We observe the performance of these heuristics is stable and steadily increases over training suggesting that generative pretraining supports the learning of a wide variety of task relevant functionality. We also observe the LSTM exhibits higher variance in its zero-shot performance suggesting that the inductive bias of the Transformer architecture assists in transfer.

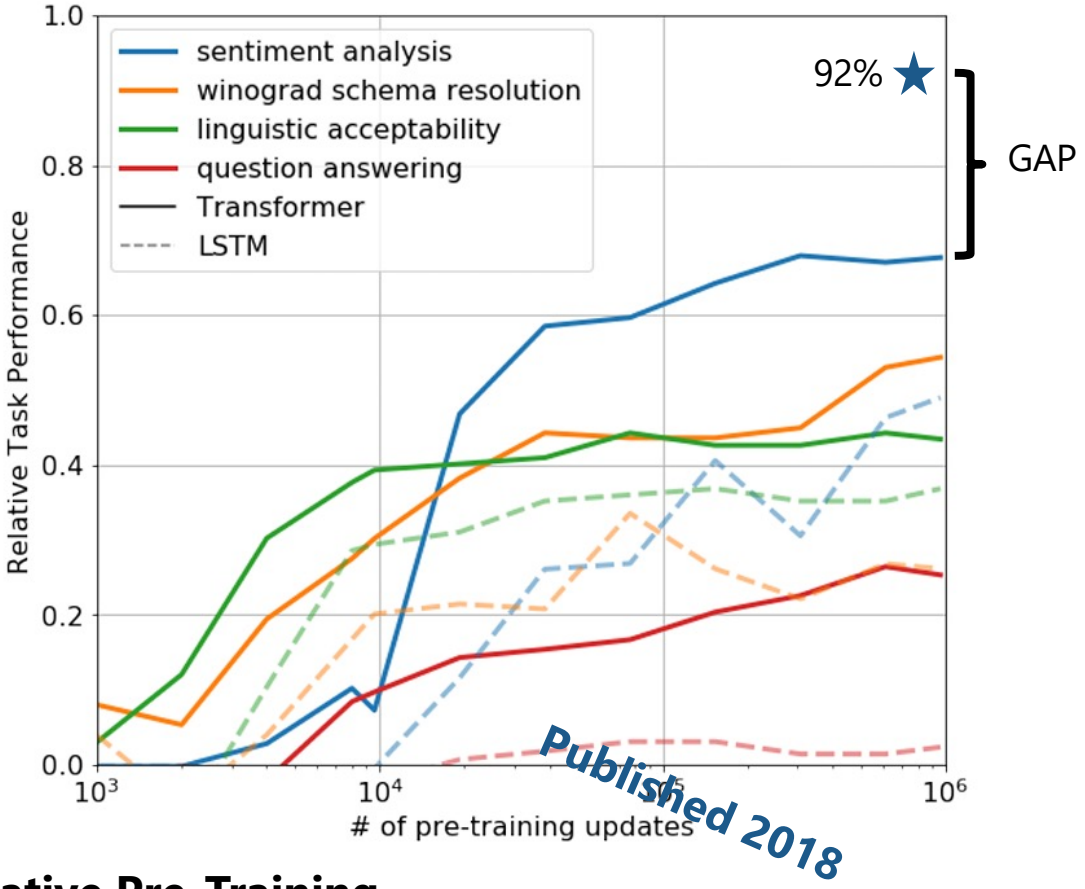
Published 2018

Improving Language Understanding by Generative Pre-Training

Zero-shot

Zero-shot Behaviors We'd like to better unders
ers is effective. A hypothesis is that the underlyir
tasks we evaluate on in order to improve its langua

attentional memory of the transformer assists in t
of heuristic solutions that use the underlying gen
finetuning. We visualize the effectiveness of thes
pre-training in Fig 2(right). We observe the perf
increases over training suggesting that generative
of task relevant functionality. We also observe t
performance suggesting that the inductive bias of



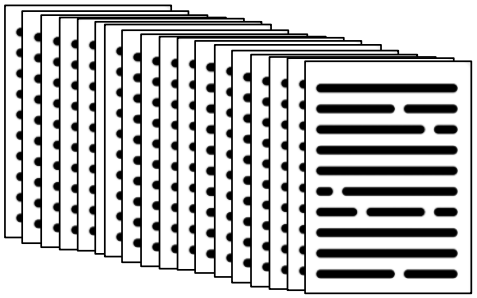
Improving Language Understanding by Generative Pre-Training

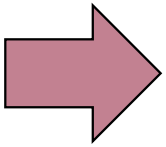
How can Generative AI/LLM help financial supervisors?

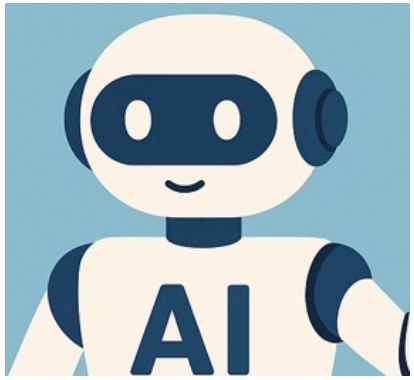


"Summarise the most critical news for a financial supervisory institute to monitor any potential market dysfunction (eg shortage of liquidity) in the coming days."

News articles







AI and ML have significant advantages over traditional econometric models…

- Can easily incorporate nonlinear dynamics
- Use large sets of explanatory variables
- Better-suited to capture rare events
- One drawback is lack of “explainability”…what variable is driving the forecast?

Literature still relatively small & largely focused on asset prices, but growing

- Kelly et al (2023): 'complex' ML model outperform 'simple' econometric ones for stock return predictions. Bali et al (2023): similar result for options.
  - Sirignano and Cont (2019) : deep learning to document that price formation may be driven by similar mechanisms across different types of assets.

- Beutel et al (2019): AI/ML models have poor record of predicting banking crises, while Bluwstein et al (2020) and Tölö (2020) find more positive results for financial crises defined more widely.

Innovation to RNN architecture: variable weights to explain prediction

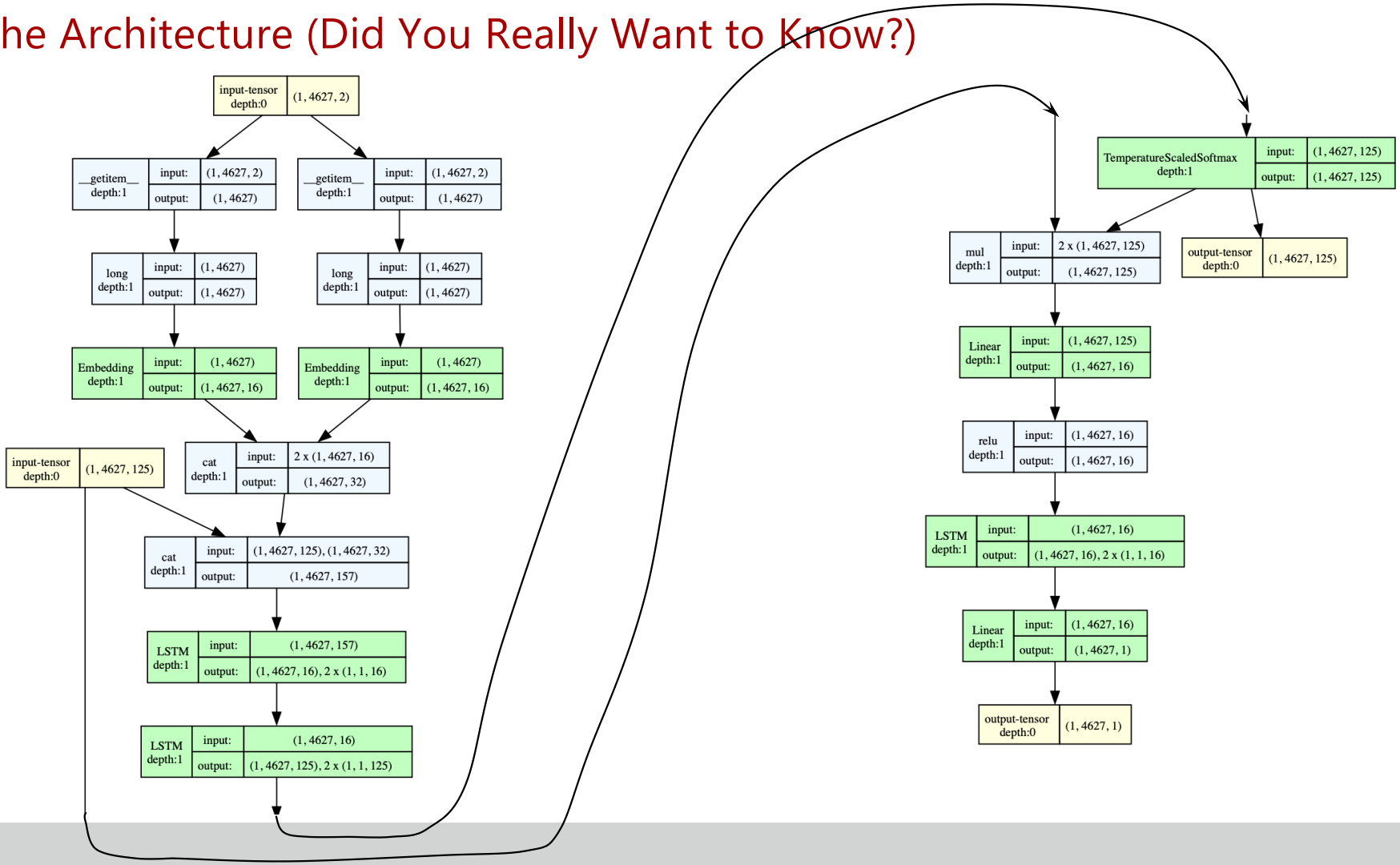
- Add a module to the model that:
  - Uses all the prediction variables
  - Learns the importance of each predictor for the final prediction
  - Updates previous weights to assess current importance of different predictors.

- Useful feature for supervisors and regulators as it gives hints of what is driving the prediction

Data

- TAP deviation (22 variables):
  - EUR-AUD, EUR-CAD, EUR-CHF, EUR-GBP, EUR-JPY, EUR-MXN, EUR-ZAR, JPY-AUD, JPY-CAD, JPY-CHF and JPY-GBP (2 stats for each)
- Foreign exchange bid-ask spreads (5)
- Equity market returns (6)
- Input data: Volatility indices (6)
- Spot exchange rates (5)
- Forward exchange rates (24)
- 25 delta risk reversals (30)
- Government bond/T-Bill yields (27)
- Temporal dummy variables (2 -> 16+16)

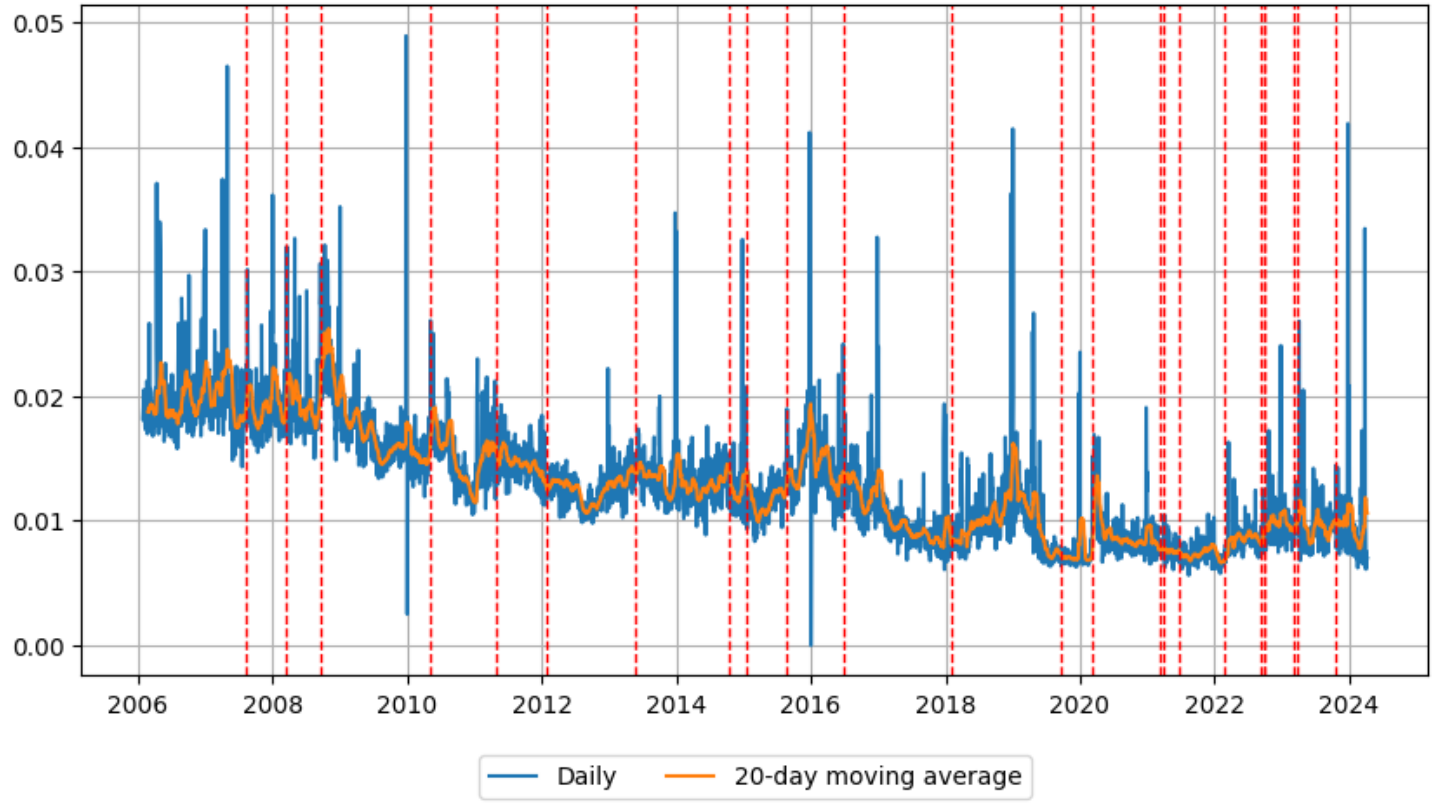
The Architecture (Did You Really Want to Know?)



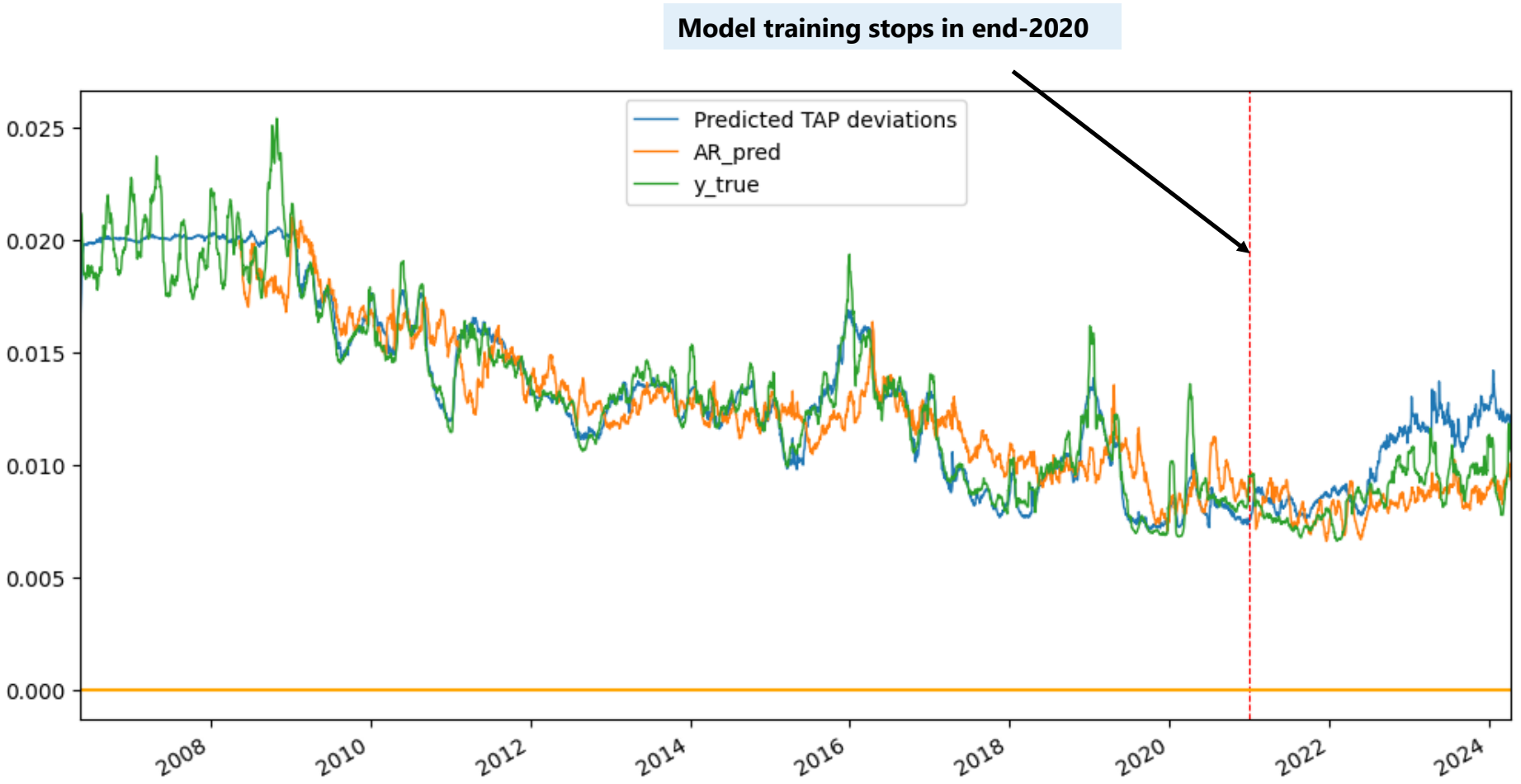
Violations of triangular arbitrage parity (TAP) as a proxy for market dysfunction

- Example: EUR/JPY = EUR/USD × USD/JPY
  - USD is the "vehicle" currency
- TAP should be easily arbitraged (Lyons and Moore (2009), Pasquariello (2014))
- Focus on TAP violations, ie 'actual' minus 'theoretical' FX rate
- Our measurement:
  - minute-by-minute TAP violations from LSEG: EUR/JPY – (EUR/USD × USD/JPY)
  - aggregated into daily standard deviation

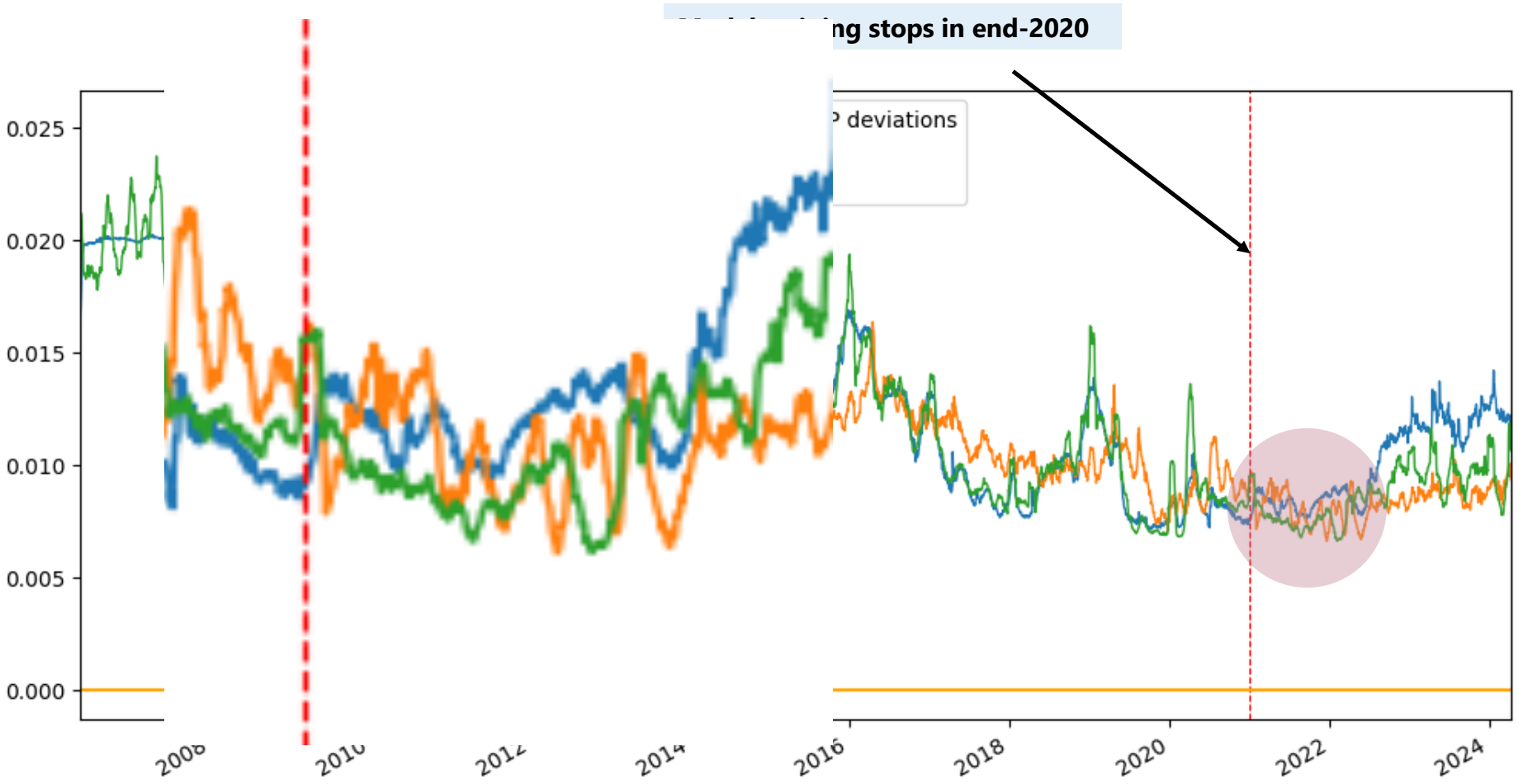
Daily standard deviation of Euro-Yen TAP violations



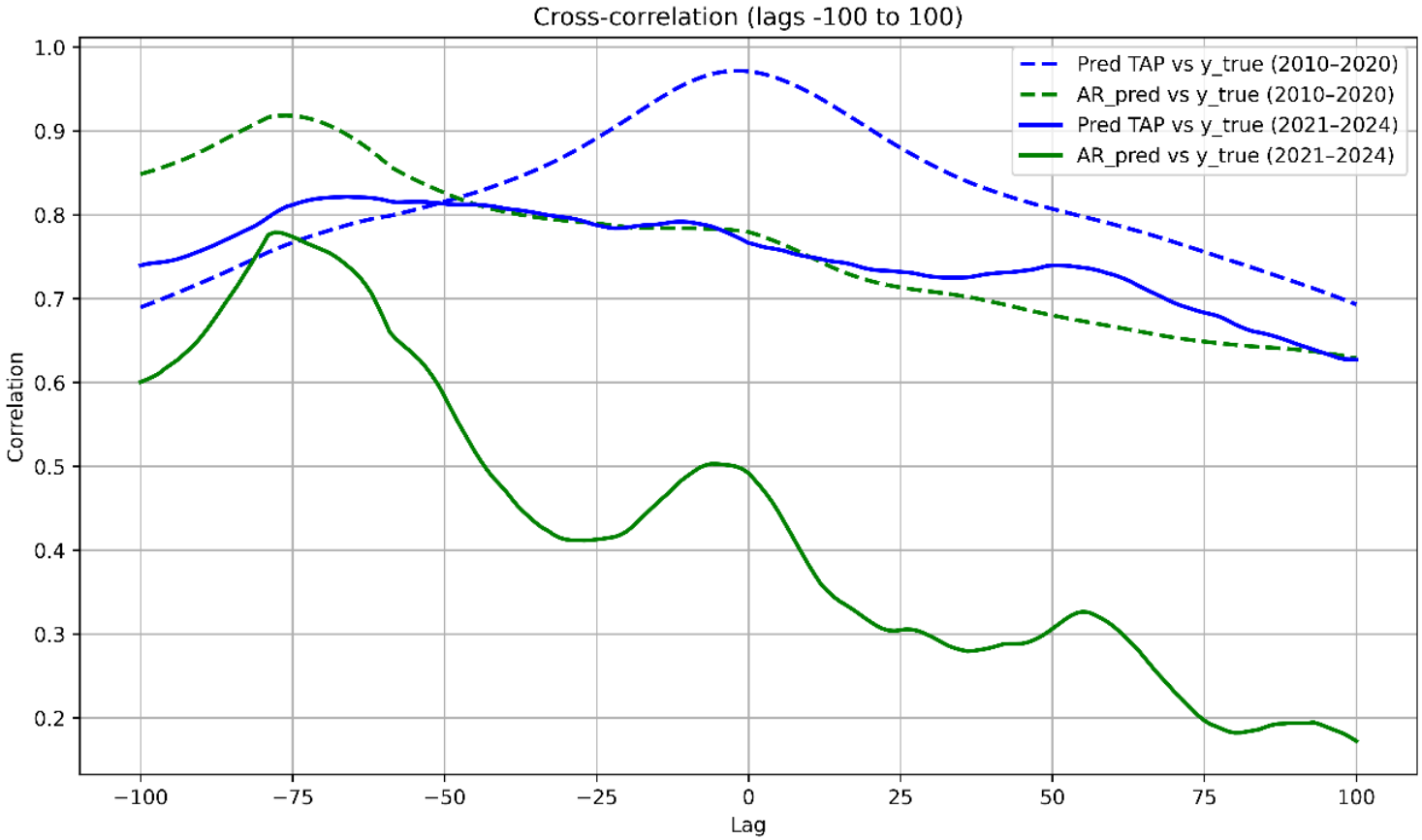
Predicting variance of triangular parity deviations in the EUR-YEN exchange rate



Predicting variance of triangular parity deviations in the EUR-YEN exchange rate

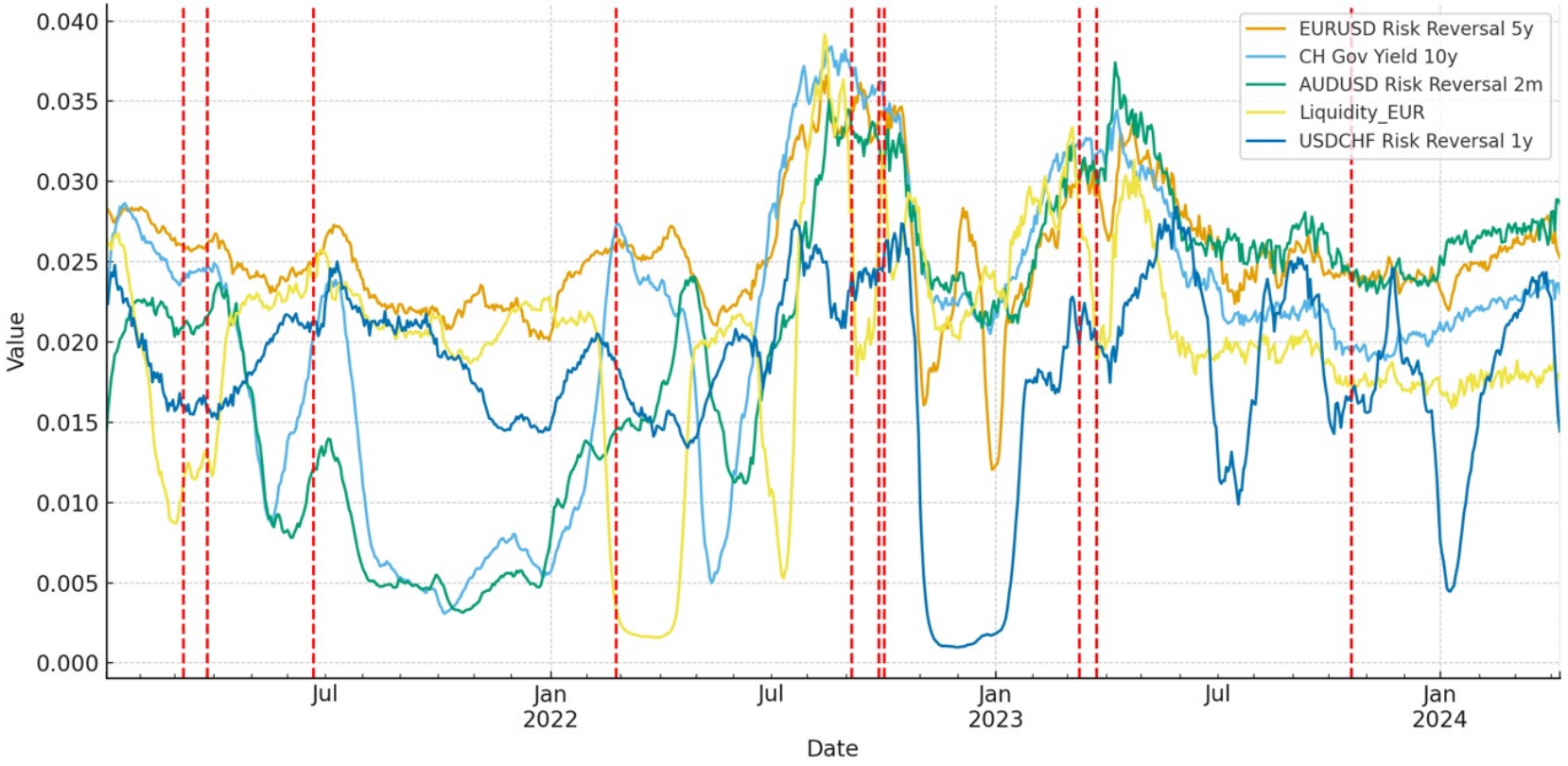


Cross Correlation

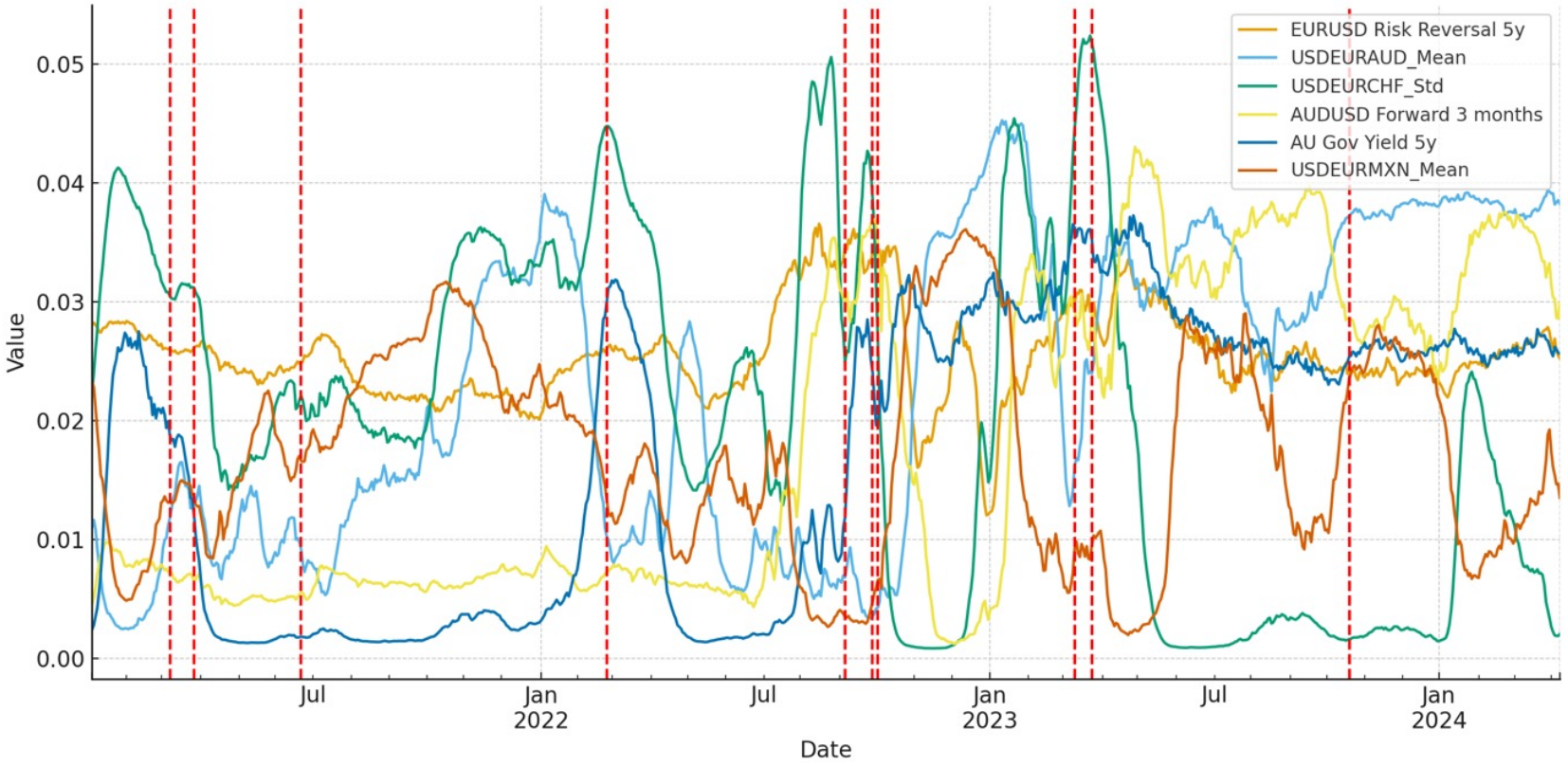


Which variables are driving the forecast?

More significant weights



More significant weights



Example: JPY turmoil of 5 August 2024

1 August 2024	2 August 2024
1. **US Treasury Yields and Basis Trade Concerns:** U.S. Treasury yields experienced a significant drop, with two -year yields falling 17.8 basis points—the largest daily decline since December 2023. This movement is linked to the Fed's potential interest rate cuts and a record volume of basis trades involvi ng Treasury futures and overnight repo markets. The extreme positions in these markets, valued at over $1 trillion each, pose a substantial risk if unwound suddenly, potentially leading to severe liquidity shortages and market disruptions. The highly leveraged nature of these trades could trigger instability in the broader financial system if market conditions shift abruptly, necessitating close monitoring by supervisory authorities. 2. **Bank of Japan's Hawkish Stance and Yen Volatility:** The Bank of Japan's (BOJ) shift towards a more hawkish policy stance, including potential further rate hikes, has already led to significant volatility in the yen, impacting global carry trades. The BOJ's actions could lead to further strengthening of the yen, disrupting global markets that have heavily relied on the yen as a funding currency. The yen's volatility could trigger a rapid unwinding of carry trades, leading to substantial market fluctuations and liquidity challenges, particularly in emerging markets and other highly leveraged financial sectors. 3. **Bank of England's Rate Cut and Potential Policy Reversal:** The Bank of England (BoE) recently cut interest rates from a 16-year high, signalling a cautious approach to future cuts. This decision reflects growing concerns about the UK economy's resilience and inflation dynamics. The BoE's actions could influence market expectations and create volatility in UK financial assets, with potential spillo ver effects on global markets. The BoE's policy decisions may impact investor confidence and market stability, especially if future rate cuts or reversals are perceived as insufficient to address underlying economic weaknesses. Monitoring the BoE's communications and market reactions wi ll be crucial.	1. **Sharp Decline in U.S. Treasury Yields and Aggressive Rate Cut Expectations:** U.S. Treasury yields, particularly the two -year and ten -year, experienced significant drops, with the two -year yield falling below 4% for the first time since May 2023. This decline is driven by weaker -than-expected employment data and increasing market expectations of aggressive interest rate cuts by the Federal Reserve. The market is now pricing in a 120 basis point cut by the end of the year, which could lead to increased volatility and liquidity concerns in the bond market. Sudden shifts in yield curves and rate expectations can cause liquidity strains and lead to disorderly market conditions, particularly if leveraged positions in Treasury futures and overnight repo markets begin to unwind. 2. **Emerging Markets and Asian Equity Market Volatility:** Emerging Asian markets, particularly South Korea and Taiwan, saw significant declines, with indices dropping 4.4% and 3.7% respectively. This sell-off was triggered by fears of a U.S. economic slowdown and was exacerbated by the Fed's potential end to its restrictive policy stance. The broader impact included a significant decline in the MSCI International All Asia Pacific Index. Market volatility in these regions can lead to capital outflows, currency instability, and broader financial instability, which could spill over into other emerging markets and potentially affect global financial stability. 3. **Increased Bearish Sentiment Among Hedge Funds:** Global hedge funds have been increasingly adding bearish bets, particularly in equity markets, driven by fears of a faster-than-anticipated U.S. economic slowdown. This sentiment was reinforced by recent weak economic data, leading to a correction in the N asdaq Composite and broader equity markets. Hedge funds reduced exposure across multiple sectors, signaling a shift towards a more defensive posture. A continued shift towards bearish sentiment among major financial players could lead to increased market stress and a tightening of liquidity, particularly in riskier asset classes. Supervisory bodies need to monitor these developments to mitigate potentia l systemic risks.

- Hundreds of daily financial news
- Ask LLM to find most critical news for supervisors *before turmoil*
- LLM successfully identified critical and relevant news

Gemini 2.5 (Trained until early 2023)

- Prompt: Attached is a collection of financial-market news in the first half of July 2023. We search for predictive signals across various daily financial market data, and our prediction model flagged unusual signals in USD/CHF 1-year risk reversal, AUD/USD 3-month forward points, and triangular-arbitrage parity (TAP) deviations in USD–EUR–AUD and USD–EUR–MXN. Informed by that, identify and summarize from your corpus the top three developments a financial supervisor should monitor for potential market dysfunction in the coming months. Output: exactly three bullets, ordered by importance/relevance. For each bullet, 1–2 sentences on why it matters. Use only the attached articles. Keep it concise.

Gemini 2.5 says:



Based on the provided news articles from July 2023, here are the top three developments a financial supervisor should monitor for potential market dysfunction:

- **Diverging views on the U.S. Federal Reserve's monetary policy.** Markets are increasingly pricing in an end to the Fed's rate-hike cycle due to softening inflation data. However, Fed officials have indicated they may favor more rate increases this year, creating a significant gap between market expectations and central bank guidance that could trigger volatility.
- **Broad and rapid depreciation of the U.S. dollar.** The dollar has fallen sharply, hitting a 15-month low against a basket of currencies and an eight-year low against the Swiss franc. This trend is fueled by the view that the Fed's tightening cycle is nearly over, and a continued, rapid repricing of the world's primary reserve currency could lead to broad market instability .
- **Sharp appreciation of emerging market currencies, particularly the Mexican peso.** The Mexican peso has strengthened to its highest level against the U.S. dollar since 2015. This signals potentially crowded carry trades into higher-yielding currencies, which are susceptible to abrupt unwinds if risk sentiment shifts.

Sources

Making AI/ML actionable for financial supervision

- **Real-time market monitoring with explanations**
  - Goal is not to build the best-performing crisis forecasting model
  - Demonstrate “practical” tools to enhance supervisory monitoring capabilities
  - Actionable timeframe for supervisory intervention

- **“Opening” the "Black Box“**
  - Adjust NN architecture to weight input vars dynamically for interpretability
  - When dysfunction predicted → identify which variables matter most
  - Feed these key variables to LLMs → retrieve relevant news and context

Conclusions

- AI and ML models highlight promising results for the prediction of episodes of market dysfunction and stress
- BIS analysis shows that ML models can forecast dysfunction in different markets, such as government bonds and FX
- 'Black box' nature of ML models can be mildly overcome by adjusting the architecture of the model to define weights for each variable
- Using LLMs can help supervisors and regulators sift through substantial amounts of news and highlight areas of potential concern
- Code will be available in **Gingado** with the next update: https://dkgaraujo.github.io/gingado/