Harnessing artificial intelligence for monitoring financial markets

Matteo Aquilina[1,2], Douglas Araujo[3], Gaston Gelos[1,4], Taejin Park[1] Fernando Pérez-Cruz[1]

[1] Bank for International Settlements [2] Macquarie University [3] Central Bank of Brazil [4] CEPR

Abstract

Predicting financial market stress has long proven to be a largely elusive goal. Advances in artificial intelligence and machine learning offer new possibilities to tackle this problem, given their ability to handle large datasets and unearth hidden nonlinear patterns. In this paper, we develop a new approach based on a combination of a recurrent neural network (RNN) and a large language model. Focusing on deviations from triangular arbitrage parity (TAP) in the Euro-Yen currency pair, our RNN produces interpretable daily forecasts of market dysfunction 60 business days ahead. To address the "black box" limitations of RNNs, our model assigns data-driven, time-varying weights to the input variables, making its decision process transparent. These weights serve a dual purpose. First, their evolution in and of itself provides early signals of latent changes in market dynamics. Second, when the network forecasts a higher probability of market dysfunction, these variable-specific weights help identify relevant market variables that we use to prompt an LLM to search for relevant information about potential market stress drivers.

Keywords: Market dysfunction, liquidity, arbitrage, artificial intelligence, financial stability

JEL classification: G14, G15, G17

Introduction

Predicting financial market stress has long been a highly desirable yet elusive goal for researchers, policymakers and market participants. Despite advances in econometric modelling, anticipating episodes of market dysfunction remains extraordinarily challenging. The rarity of severe stress events limits the training data available for statistical models (Giglio et al (2016)), while the unique and often nonlinear transmission mechanisms across markets through which risks materialise make it difficult to identify common patterns across different crisis episodes (Brunnermeier and Oehmke (2013)). Traditional early warning systems have shown mixed success, often suffering from high false positive rates or failing to capture novel sources of systemic risk (Alessi and Detken (2011)). These fundamental challenges have motivated researchers to explore alternative approaches, including machine learning techniques that can process high-dimensional data and capture complex nonlinear relationships.

Recent advances in artificial intelligence (AI) and machine learning (ML) have opened new possibilities for financial stability monitoring. Large language models (LLMs) excel at processing vast amounts of textual data, extracting relevant signals, and synthesizing complex information into actionable insights for supervisors and regulators. Their ability to identify patterns in economic and financial narratives could help detect emerging risks before they fully materialise by using LLMs.[1] However, for LLMs to effectively support real-time risk monitoring, they must be directed to focus on the most relevant information, which requires knowing which market indicators and themes deserve attention at any given moment.

This paper introduces a novel two-stage approach that combines recurrent neural networks (RNNs, Goodfellow et al (2016)) with LLMs to both forecast and explain episodes of financial market stress. Our method addresses a critical gap: while quantitative models can identify statistical patterns predictive of stress, they often cannot explain why certain indicators matter or what underlying economic narratives drive the risks. Conversely, while LLMs can synthesize vast amounts of textual information about market conditions, they need guidance on which topics and variables to prioritize. We bridge this gap by developing an adjusted RNN architecture that dynamically learns to weight input variables, automatically identifying which indicators matter most for each forecast period. These time-varying weights then serve as a focusing mechanism for LLMs, directing them to search for and analyse news and commentary about the specific variables that our model identifies as critical for impending market dysfunction episodes.[2]

We apply this framework to predict stress in foreign exchange markets, focusing on deviations from triangular arbitrage parity (TAP) in the Euro-Yen currency pair. Specifically, we analyse daily standard deviations of minute-by-minute discrepancies in the parity, traded directly and indirectly with the US dollar as a vehicle currency. Such deviations can serve as a "canary in the coal mine" for broader market dysfunction, since they indicate frictions in one of the world's most liquid markets where arbitrage opportunities should typically be eliminated within seconds (Pasquariello (2014), Drehmann and Sushko (2022)). Using over one hundred daily financial variables as inputs, our RNN forecasts 20-day averages of TAP deviations with a 60-day horizon—essentially predicting market conditions three months ahead based on current information.

Our key findings demonstrate the practical value of this integrated approach. In rigorous out-of-sample testing covering 3.5 years of data (2021-2024), the model successfully identifies periods of heightened market stress, with the dynamic weights highlighting variables subsequently known to be

[1] See Appendix A.1 for more details.

[2] The code for the estimation and use of this adjusted RNN is planned to be made available through the BIS *gingado* open source library (Araujo (2023)).

associated with actual market events, including the March 2023 banking turmoil. Most importantly, when our RNN signals potential future dislocations and identifies which variables have surged in importance, we can deploy LLMs to search contemporaneous financial news and analysis for context about precisely these indicators. To illustrate this technique, we offer a case study showing that timestamped LLM queries about high-weight variables were able to uncover discussions of key drivers of financial stability episodes around three months before they occurred. This targeted use of LLMs—guided by our RNN's variable weights—transforms opaque statistical predictions into narrative explanations that policymakers can understand and act upon.

The rest of the paper is organised along the following structure: the next section provides context for our work. Section 3 describes the data and Section 4 presents the methodology we use. Section 5 presents the results and the last section concludes.

2. Market dysfunctions: drivers, dynamics, and forecasting approaches

2.1 Characteristics of market dysfunctions

Market dysfunctions, ie episodes of significant impairments in market liquidity or substantial price dislocations, are a recurring phenomenon in financial markets. Understanding the drivers of such dysfunctions and using the appropriate tools to forecast their occurrence are critical for regulators and risk managers. In this section, we explore the nature and causes of market dysfunctions, and review the approaches taken by the literature to forecast these episodes across various asset classes.

These episodes are important because are indicative of the presence of significant frictions in financial markets which can persist and eventually lead to significant market distortions and, in the extreme fully-fledged crises. Pasquariello (2014) conducts a comprehensive study of such dislocations across foreign exchange, interest rate, and equity markets and shows that these dislocations significantly increase in proximity of financial stress episodes, such as the Asian crisis in 1997, the LTCM episode in 1998, and the global financial crisis in 2008. More recently, Aldasoro et al (2022a) show that summary measures of market conditions (capturing impairments to price formation and the breakdown of no-arbitrage conditions) for three key market segments are associated with significant periods of stress.

Low market liquidity can act as both a cause and a symptom of market dysfunction. It can lead to mispricing or signal that markets are not functioning efficiently or orderly. In extreme cases, a lack of liquidity can cause the trading process to break down entirely, resulting in what Persaud (2003) terms "liquidity black holes" – episodes where the trading process itself becomes destabilizing, often triggering significant stress. The literature identifies multiple underlying causes for these phenomena. For instance, Brunnermeier and Pedersen (2009) highlight the interaction between funding liquidity and market liquidity, while Tirole (2011) attributes such episodes to adverse selection or shortages of financial capacity among intermediaries.

Having explored the characteristics and drivers of market dysfunctions, we now turn to the question of how the literature has attempted to forecast them through the use of statistical models.

2.2 Approaches to forecast market dysfunctions

A substantial body of literature has focused on understanding the potential drivers of market dysfunctions, financial stress and financial crises (Fouliard et al (2021)). The approaches taken vary depending on the specific question asked. One focus has been on distinguishing between tranquil and stressed regimes.

Another one has been on identifying variables that help forecast periods of dysfunction or low liquidity. Studies have largely centred on stock markets, where data has been most readily available. But over the past two decades, this research has expanded to include other markets—especially government bonds and foreign exchange—and to highlight the commonality of liquidity across them.

Markov (regime) switching models (Hamilton (1989)) are the most widely used tool to model the transition between tranquil and stressed periods. Regime switching models estimate two types of parameters: the variables of interest in each of the regimes and the probability of switching from one regime to the other. While in the early incarnations of these models such switching probabilities were fixed, more recent developments have incorporated time-varying ones. Guidolin (2011) provides a summary of the available literature and highlights the ability of these methods to forecast regime changes. However, the focus has been mostly on low-frequency data, typically at quarterly or monthly frequency.

By contrast, a vast literature describes forecasting dysfunction or lack of liquidity directly. It largely focuses on stock markets, where most of the data have historically been available. Karolyi et al (2012) examine liquidity commonality in stock markets across countries and over time and highlight that demand factors, and in particular correlated investor flows (mainly mutual funds), are more important for stock market liquidity. Pástor and Stambaugh (2003) and Acharya and Pedersen (2005) show that investors demand a premium for bearing liquidity risk. But the determinants of liquidity are not related simply to a single asset class. For example, Chordia et al (2005) analyse the interaction between liquidity in sovereign bonds and in stocks in the United States. They find that changes in volatility and liquidity within and across asset classes are associated with future liquidity levels, suggesting both market interconnections and an autoregressive component in liquidity determination.

Foreign exchange markets have been studied significantly less than other markets. Karnaukh et al (2015) and Ranaldo and Santucci de Magistris (2022) are among the most comprehensive studies of liquidity determinants in foreign exchange markets. Both papers emphasize that liquidity deteriorates when intermediaries are less able to respond to shocks, whether due to operational reasons (eg, reduced liquidity outside London hours and during holidays) or risk management issues (eg, higher funding costs or volatility). They find economically significant effects: a one standard deviation increase in the VIX is associated with a 17% rise in liquidity costs in foreign exchange markets. Additionally, currencies more exposed to systemic risk experience higher liquidity commonality. A key finding is the high degree of liquidity commonality across foreign exchange, bond, and stock markets, with liquidity in one market deteriorating when it is lower in another. Gradojević et al (2020) look more specifically at currency TAP deviations at an ultra-high frequency definition (less than one second) and find that a denoised version can be forecasted at that frequency.

In a recent study, Aliyev et al (2024) examine the resilience of liquidity in foreign exchange, sovereign bond, and stock markets over the past 25 years across various geographical areas. They find that while the mean and standard deviation of bid-ask spreads have declined across all markets studied, higher moments of the distribution, especially skewness, show different patterns. Skewness has increased significantly in equity and sovereign bond markets, while it exhibits an inverted-U shape in currency markets. This suggests that episodes of illiquidity are now more common in sovereign bonds and stock markets. The authors also find that structural breaks in mean spreads correlate with macroeconomic shocks, while breaks in skewness are more closely associated with changes in regulation and market microstructure. Overall, modern markets may have traded off smaller average spreads for higher liquidity skewness.

Recent advances in AI/ML have expanded the frontier of financial forecasting by enabling the analysis of high-dimensional and complex datasets. These methods have been particularly effective in capturing nonlinear relationships and high-frequency dynamics. For example, Bali et al. (2023) highlight the role of non-linearity in forecasting option returns, while Tölö (2020) demonstrates the superiority of

RNNs in predicting banking crises. Similarly, Aldasoro et al. (2025) show that ML models outperform traditional benchmarks in forecasting the future distribution of market conditions. Sirignano and Cont (2019) use RNNs to document universal features of asset prices, and Vuletić et al (2024) combine RNNs with a generative adversarial network (GAN) model, to successfully forecast financial series. Ward (2017), Beutel et al (2019) and Bluwstein et al (2020) exploit AI/ML methods to predict banking and financial crises.

Recent developments in large language models (LLMs) have further transformed financial market prediction and risk monitoring. These models enable the systematic extraction and synthesis of latent signals from vast textual datasets, encompassing market commentary, corporate disclosures, regulatory filings, and social media postings, thereby expanding the analytical frontier beyond conventional structured data sources. While promising, their probabilistic nature and reliance on historical patterns limit their ability to provide causal insights, making them more suitable as complementary tools for human decision-making rather than standalone forecasting systems. (Cao et al. (2024); Ding et al. (2024)).

3. Measuring market dysfunction

We study TAP violations in a representative currency pair. We focus on TAP violations as a testing ground, but it is important to note that these violations are associated with periods of significant market dysfunction and other indicators of strain in global financial markets, such as increases in cross-currency swap spreads. Hence, while our approach in this study is limited to a single indicator, it is representative of market-wide financial conditions.

We focus on the EUR-JPY exchange rate with the USD as the vehicle currency; this pair is markedly more volatile than others, even though the two currencies are widely traded (see Table 1). Defining the FX rate between two currencies i and j, with $i \neq j$, as $S_{i/j}$, this TAP deviation is:

$$\dot{y} = S_{\frac{EUR}{JPY}} - \left(S_{\frac{EUR}{USD}} \times S_{\frac{USD}{JPY}} \right), (1)$$

In our application, we calculate $\dot{y}$ for each 1-minute window. Our variable of interest is the daily standard deviation of these 1-minute TAP deviations:

$$y_t = \sqrt{\sum_{t(\dot{y}) \in t} \left(\dot{y} - \frac{\sum_{t(\dot{y}) \in t} \dot{y}}{|\dot{y}|_t} \right)^2 \Big/ |\dot{y}|_t}, (2)$$

where t represents the day, $|\dot{y}|_t$ is the number of valid measurements in day t, and $t(\dot{y}) \in t$ means that the operator is only valid for the measurements in day t. Using the standard deviation, and not the average, ensures the resulting daily series is more sensitive to differences in intraday values of TAP deviations, while capturing deviations in either direction of the trade, since both directions equally offer arbitrage opportunities. Furthermore, the daily mean of $\dot{y}$ is not a good metric because it would become attenuated if TAP violations occurred in different directions during the same day, as one violation would offset the other, resulting in a false sense that deviations are smaller than they are.

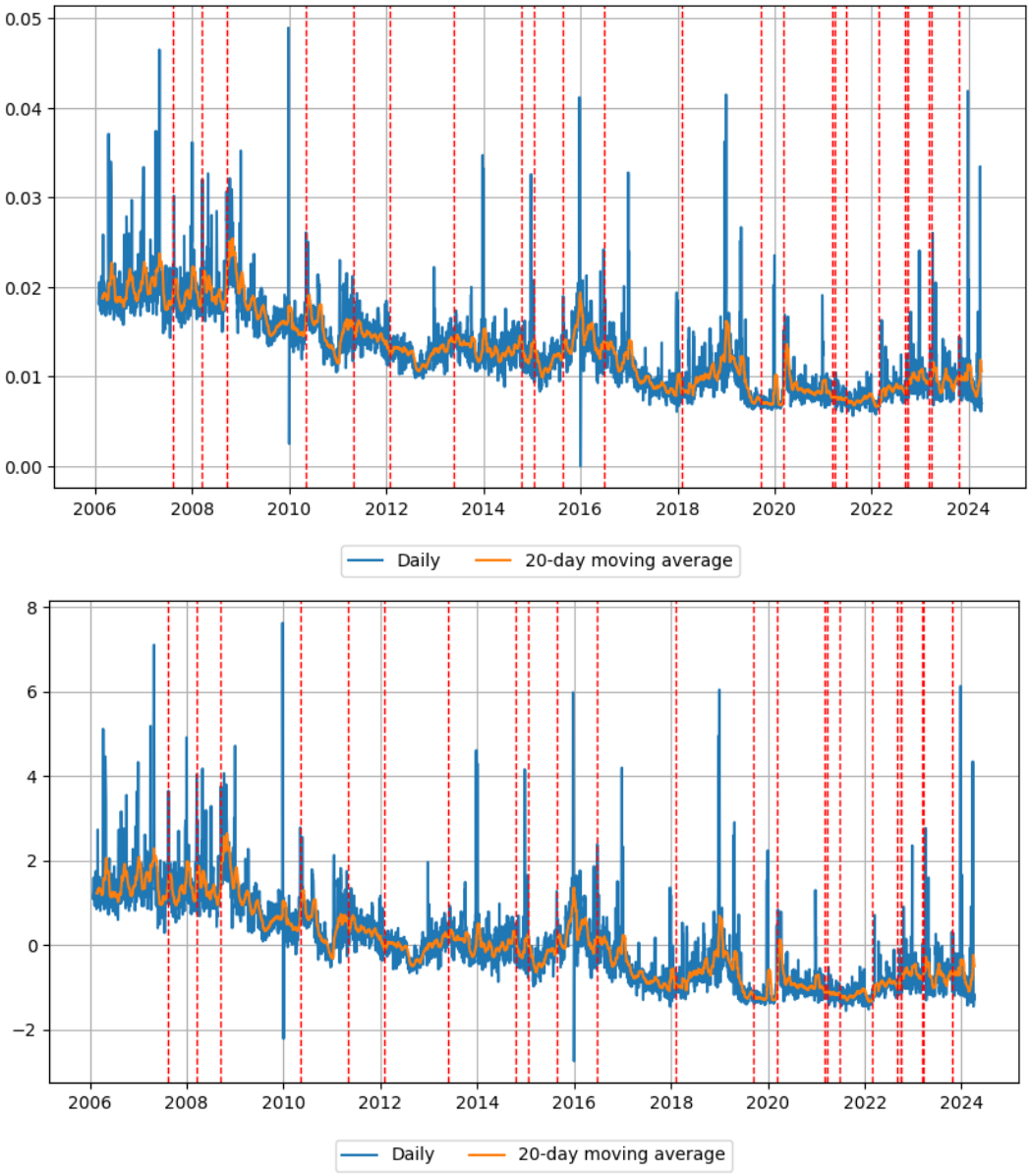
Since the daily data are noisy and forecasting exact dates is not the focus of financial stability analyses, we set out to forecast a smoothed version of the data. For simplicity, we use as dependent variable a moving average with a window of w=20 days, roughly equivalent to a calendar month:

$$\bar{y}_t = \frac{1}{w} \sum_{p=0}^{w-1} y_{t-p}. (3)$$

Graph 1 gives an intuitive picture of how TAP deviations ebb and flow over time and highlights when markets came under stress.

Standard deviation of Euro-Yen TAP violations

Graph 1



Top panel: Raw standard deviation. Bottom panel: Standardised value (zero mean and unit variance). The vertical red dashed lines represent market stress events.

Source: Authors' calculations.

4. Methods and Data

This section presents the neural network machinery, describes the formal tests applied to these models, and discusses the input variables used to forecast the TAP deviations.

4.1 Recurrent neural networks

Of all myriad possibilities for nonlinear and nonparametric modelling of TAP violations, we focus on RNNs, since they combine all key characteristics of these processes documented in the literature. RNNs nest autoregressive models, an important feature since market liquidity innovations help explain future liquidity (Chordia et al (2005)). RNNs also flexibly model regime-switching behaviour with time-varying dynamics, addressing empirical evidence of different persistent regimes (Hauptmann et al (2014), Duprey and Klaus (2017), Chuliá et al (2021)). Another desirable characteristic is their ability to efficiently model signal from large datasets, reducing model risk from sparse representations (Giannone et al (2021)).

In addition, the empirical use of RNNs is very flexible. We take advantage of this flexibility and design a neural network whose prediction architecture can at the same time extract meaningful signals and inform the econometrician as to which variables these signals came from, at each point in time. This and other design decisions are discussed in Appendix A.2.

The essence of our model is described by

$$\hat{\bar{y}}_{t+w+h} = f(h_{t-1}, \tilde{x}_t)(4.i)$$

$$h_{t+1} = g(h_{t-1}, \tilde{x}_t), (4.ii)$$

where both $f(.)$ and $g(.)$ are nonlinear functions of the relevant information from past periods, h_{t-1}, input data $\tilde{x}_t$. $\hat{\bar{y}}_{t+w+h}$ represents the estimate of the average TAP deviation for the period between $t+h+1$ and $t+w+h$ days ahead. The input variable $\tilde{x}_t$ is described below:

$$\tilde{x}_t = x_t \odot \frac{w_t}{\max(w_t)}, (4.iii)$$

$$w_t = v(w_{t-1}, x_t, \xi_{t+w+h}), (4.iv)$$

where x_t are the explanatory variables for our prediction, covered in Subsection 4.3, $\odot$ stands for the elementwise (or Hadamard) multiplication, and w_t are importance weights that indicate in each period how relevant each variable is for predicting $\bar{y}_{t+w+h}$. We divide the weights w_t by the highest weight at each period so that the variable carrying the most signal is not dampened by the model. All input variables are standardised individually to zero-mean, unit-variance series, considering only the training period mean and standard deviation. This ensures that $v(.)$ is not influenced by the scale of the variables, only by the relevance of their signal.

The weights are determined by $v(.)$, a nonlinear function of the past value of the weights, current input data, and temporal dummies ξ_{t+w+h} marking the quarter of the year for the forecasted day and whether the forecasted day is within a 10-day period centred at quarter-ends. These temporal dummies help determine the weight given to each variable, consistent with literature findings that seasonality and/or quarter-end timing matter (Chordia et al (2005), Aldasoro et al (2022b)), but are themselves not input variables in the main equation (4.i). The weight of each variable m at date t is always non-negative, and in each period, all weights add up to one: $w_{m,t} \geq 0$, $\sum_m w_{m,t} = 1 \forall t$. Hence, at each period the different variables compete for importance in the prediction, with our model "learning" how to optimise forecasting performance by allocating the weights of each variable based on current and past observations of all variables and seasonal markers. Once the model estimates the weight of each variable for a particular

date, that variable is multiplied by its weight scaled by the largest weight of that date. This ensures that the highest-weighted variable is multiplied by one.[3]

The system (4.i)-(4.iv) comprises an iterative map $G_{h,w}: (x_t, \xi_{t+w+h}) \rightarrow \mathbb{R}$ that directly forecasts $\bar{y}_{t+w+h}$, the 20-day average of the TAP deviation measure starting after each forecast horizon $h = 60$ days. In other words, each forecast uses data up to day t to forecast the 20-day window that starts 60 days later. Our forecasts are direct, with data at time t used to forecast the w-day average values for the period ending at time $t + w + h$; rather than iterated, where time t-data forecasts $t + w + 1$, and $t + 1$ data forecasts $t + w + 2$, and so on all the way until $t + w + h$. Empirically, we structure $G_{h,w}$ as an RNN, specifically a neural network containing Long-Short Term Memory (LSTM) layers and estimate it via gradient descent. In Appendix A.1 we present the general architecture for LSTMs and in A.2 we detail our specific implementation.[4] The weights w_t are themselves also computed with another RNN building block that is estimated jointly as part of the larger model (ie, $v(.)$ is parameterised as an RNN; details can be found in Appendix A.2). These weights are a valuable byproduct of the model, because they inform us about which input variables are relevant for predicting the TAP in each time instance.

As a baseline, we compare our model to the challenging AR model prediction, namely:

$$\hat{\bar{y}}^{AR}_{t+w+h} = \alpha + \sum_{i=0}^{l} \beta_i \bar{y}_{t-i}, (5)$$

where l represents the size of the context used to train RNN models, h the direct forecasting horizon and w the number of days over which the dependent variable is averaged, all measured in days. The values of α and β are estimated by OLS.

Many macroeconomic time series are highly persistent and have low signal-to-noise ratios. A first-order autoregressive process, AR (1)—of which the random walk is a special case—captures persistence with a single parameter, enabling competitive forecasts without incurring the variance and overfitting penalties that undermine richer models (Nelson and Plosser (1982)). Across numerous classic forecasting horse races, particularly for inflation, more sophisticated models do not reliably outperform them out of sample (Atkeson and Ohanian (2001); Stock and Watson (2007)).

4.2 Data

Our main source of data is LSEG Tick History. Tick History provides historical information at different levels of aggregation (from tick-by-tick to daily) for several asset classes. Its coverage goes as far back as 1996 for many times series and is widely used by industry practitioners. The dataset consists of recorded trades and quotes from a number of several real-time feeds across more than 500 trading venues including all types of participants (ie, not just dealers).

We obtain 1-minute snapshots of bid and ask quotes and calculate the midpoint, as well as the bid-ask spread and its standard deviation. We then average these high frequency observations to obtain daily observations. We clean the data by only using local trading hours (9:30 to 16:00) and removing

[3] For more details on this dynamic variable weighting algorithm and alternative setups, see Araujo (2025).

[4] Another neural network architecture that could perform the same task is the so-called transformer architecture (Vaswani et al (2017)). Instead of a recurrent model that applies to each time step, learning to retain a "memory" that is relevant to predict the future, like LSTMs, transformers directly estimate the dependence (self-attention) of a time step on the others. Indeed, Snellman (2025) uses a transformer model to construct a market fragility index. Other widely used ML methods based on completely different algorithms than neural networks, such as elastic net (Zou and Hastie (2005)) and XGBoost (Chen and Guestrin (2016)), are not used as they do not hold a stateful representation of the time series as RNNs do.

weekends.[5] We also eliminate data when the order book is crossed (i.e., ask price is lower than or equal to the bid price). To predict TAP deviations, we use a wide-ranging set of financial market variables.

The daily mean and standard deviation of the minute-by-minute TAP deviations of other FX pairs of relevance, and lagged values of the pair of choice, are used as explanatory variables (Table 1). These are, namely: EUR-AUD, EUR-CAD, EUR-CHF, EUR-GBP, EUR-JPY, EUR-MXN, EUR-ZAR, JPY-AUD, JPY-CHF and JPY-GBP with the USD as the vehicle currency. Note that these pairs include both advanced and emerging market economies, two currencies that are usually considered safe havens (CHF and JPY) and also advanced economy currencies that are usually associated with commodities (AUD and CAD).

Input variables: TAP deviations[1] Table 1

	Full sample							Training set	Test set
Currency pair	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
EURAUD (Mean)	1.7E-05	4.0E-05	-6.3E-04	2.8E-06	1.5E-05	3.3E-05	3.9E-04	1.7E-05	1.9E-05
EURAUD (S.D.)	2.0E-04	9.5E-05	7.6E-05	1.2E-04	1.8E-04	2.5E-04	9.1E-04	2.2E-04	1.1E-04
EURCAD (Mean)	4.1E-05	5.8E-05	-4.1E-04	1.3E-06	3.9E-05	7.7E-05	3.1E-04	4.9E-05	1.4E-06
EURCAD (S.D.)	1.8E-04	6.7E-05	5.9E-05	1.5E-04	1.8E-04	2.1E-04	4.5E-04	2.0E-04	1.1E-04
EURCHF (Mean)	-9.9E-06	3.3E-05	-2.8E-04	-1.6E-05	-5.8E-06	2.8E-06	4.2E-04	-1.0E-05	-7.8E-06
EURCHF (S.D.)	1.4E-04	5.6E-05	0.0E+00	8.9E-05	1.3E-04	1.8E-04	4.9E-04	1.5E-04	8.3E-05
EURGBP (Mean)	6.7E-06	1.6E-05	-1.2E-04	2.1E-06	7.5E-06	1.3E-05	3.2E-04	6.0E-06	1.0E-05
EURGBP (S.D.)	8.4E-05	2.3E-05	0.0E+00	6.8E-05	8.1E-05	9.4E-05	3.1E-04	8.8E-05	6.4E-05
EURJPY (Mean)	-8.2E-04	6.2E-03	-1.0E-01	-2.8E-03	-9.6E-04	8.0E-04	8.4E-02	-1.1E-03	2.7E-04
EURJPY (S.D.)	1.3E-02	4.7E-03	0.0E+00	9.1E-03	1.2E-02	1.6E-02	4.9E-02	1.4E-02	8.7E-03
EURMXN (Mean)	-1.2E-03	4.2E-03	-2.3E-02	-2.7E-03	-3.2E-05	1.2E-03	2.1E-02	7.5E-05	-7.3E-03
EURMXN (S.D.)	3.6E-03	1.6E-03	0.0E+00	2.7E-03	3.3E-03	4.3E-03	1.5E-02	3.6E-03	3.8E-03
EURZAR (Mean)	-5.2E-04	1.7E-03	-1.8E-02	-8.3E-04	-1.4E-04	1.2E-04	4.9E-02	-6.5E-04	7.0E-05
EURZAR (S.D.)	3.9E-03	2.1E-03	0.0E+00	2.6E-03	3.3E-03	4.5E-03	2.7E-02	4.1E-03	3.0E-03
JPYAUD (Mean)	1.1E-07	6.0E-07	-2.3E-06	-3.5E-08	1.1E-07	2.5E-07	8.2E-06	2.0E-07	-2.9E-07
JPYAUD (S.D.)	1.7E-06	7.3E-07	0.0E+00	1.2E-06	1.5E-06	2.0E-06	1.0E-05	1.7E-06	1.8E-06
JPYCAD (Mean)	1.7E-07	1.1E-06	-6.1E-06	-1.4E-07	1.9E-07	5.5E-07	6.0E-06	3.8E-08	7.9E-07
JPYCAD (S.D.)	1.3E-06	6.3E-07	0.0E+00	1.1E-06	1.3E-06	1.7E-06	3.9E-06	1.5E-06	4.4E-07
JPYCHF (Mean)	-1.8E-07	5.7E-07	-5.7E-06	-3.0E-07	-5.7E-08	8.1E-08	4.4E-06	-1.0E-07	-5.3E-07
JPYCHF (S.D.)	1.2E-06	5.1E-07	0.0E+00	7.5E-07	1.2E-06	1.5E-06	3.6E-06	1.3E-06	8.7E-07
JPYGBP (Mean)	5.9E-08	2.0E-07	-2.9E-06	2.1E-08	5.8E-08	9.7E-08	3.1E-06	5.2E-08	9.1E-08
JPYGBP (S.D.)	5.6E-07	2.4E-07	0.0E+00	3.8E-07	4.8E-07	6.9E-07	1.7E-06	6.0E-07	3.5E-07

[1] Daily mean (Mean) and standard deviation (S.D.) of the minute-by-minute difference between the currency pair and the indirect price using the US dollar as a vehicle currency.

Sources: LSEG Tick History; authors' calculations.

In addition, our selection of potential predictors also includes variables that have proven to be useful when used in machine learning models for prediction of returns (Gu et al (2020)). Specifically, we include as explanatory variables bid-ask spreads on large currency pairs as proxies of liquidity, equity

market implied volatility (as estimated from options markets), foreign exchange risk reversals (the difference in implied volatility between comparable call and put options), forward FX quotes to capture forward-looking measures of market expectation. In contrast, while the volatility of bid-ask spreads over a rolling window of time and momentum variables for both the TAP deviations and liquidity are known to be able to forecast financial variables, we did not include them as the recurrent nature of the model should have the ability to approximate these time-wise moments from the data if they are indeed relevant. The input data further include equity benchmark returns for major financial markets, the MOVE bond volatility index and a range of government bond yields.

In addition, we include quarter dummies and a dummy marking the ten-day period around quarter-ends. These variables capture variations in liquidity conditions that are seasonal or periodic, as well as any factors around quarter-ends that could create or amplify liquidity shortages (eg, simple seasonal festivities or window-dressing (Garcia et al (2021)).

A complete list of the input data is in Table 2 below. Descriptive statistics for the other input data that are not TAP deviations are found in Appendix 4.

Input variables: sources

Table 2

Variable	Source
TAP deviations	LSEG Tick History; authors' calculations.
Foreign exchange bid-ask spreads	LSEG Tick History; authors' calculations.
Equity indices	Bloomberg.
Spot exchange rates against the US dollar	Bloomberg.
Forward exchange rates against the US dollar	Bloomberg.
Equity market volatility indices	Bloomberg.
MOVE index	Bloomberg.
Government bond yields	Bloomberg; LSEG Datastream.
Risk reversals	Bloomberg.

5. Findings

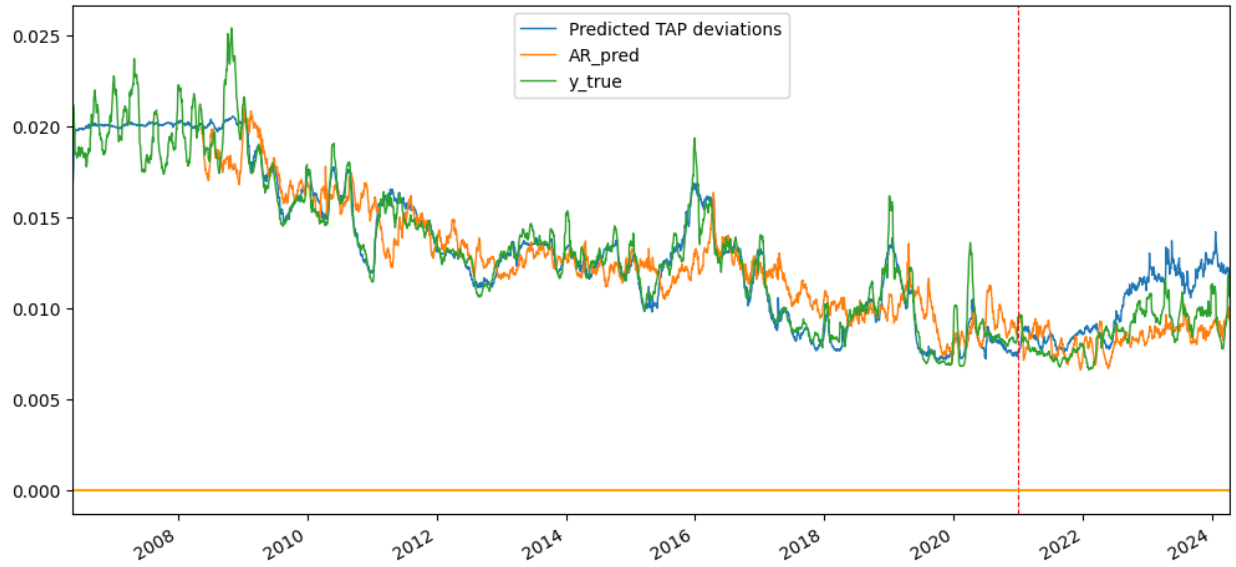
This section introduces the findings from the forecasting exercise, the identification of liquidity periods of interest, and the use of LLMs for summarisation of text content that could be related to events of interest.

5.1 Market dysfunction prediction

Relatively simple LSTM models, regularised by the dynamic variable selection feature, perform competitively with respect to an AR benchmark, although slightly different configurations along the dimension size perform better at different horizons. Graph 2 presents the results for the 20-day average of Euro-Yen TAP violations, forecast 60 business days ahead. We discuss the performance of the model more formally below. However, a few interesting insights emerge from a visual inspection of the graph.

Performance of LSTM model

Graph 2



True data: 20-day average of the daily Euro-Yen triangular arbitrage parity difference with the US dollar as vehicle currency, calculated on a minute-by-minute basis. The vertical red dashed line represents the end of the training period, end-2020; everything to the right of this line is considered pseudo out-of-sample.

Source: Authors' calculations.

First, TAP deviations are indeed forecastable to some extent, and exploring the nonlinearities across both the cross-sectional and temporal dimensions might be key to extract this signal. It is plausible that a wider search along the parameter space of these models might yield even better performances, although that remains an empirical question.

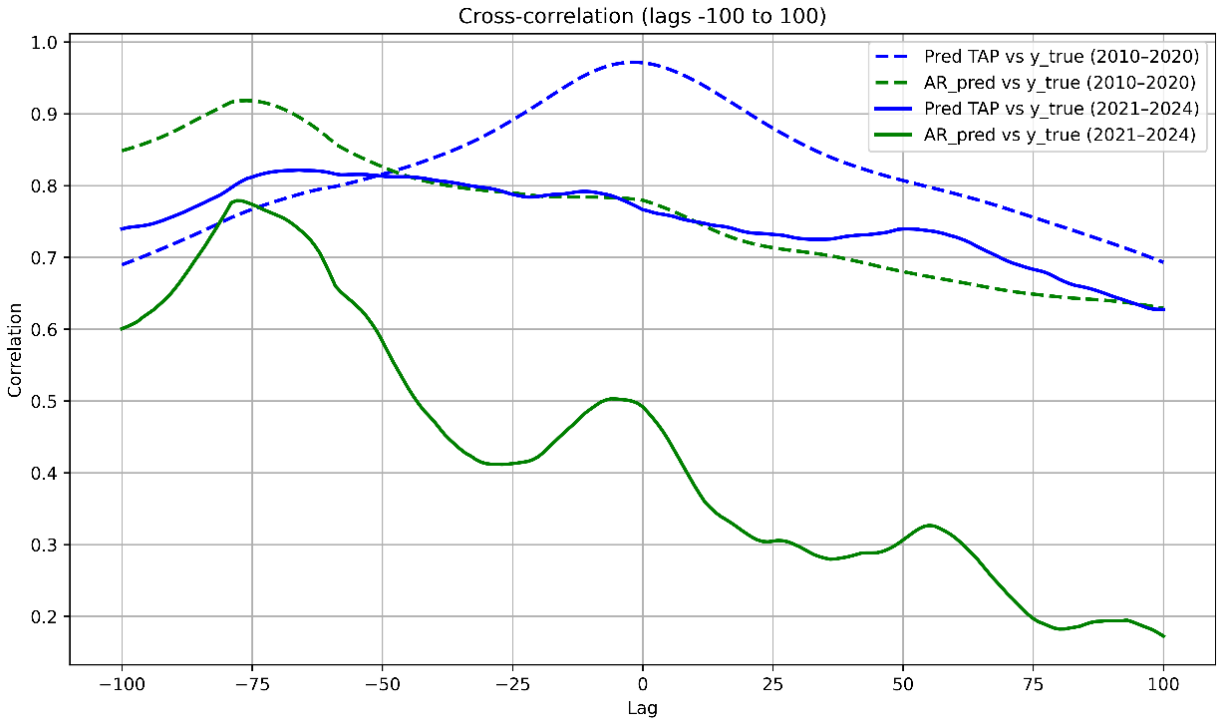
Second, the RNN model has a particularly useful feature, namely that it is less noisy than either the actual realisations of the target variable and of the predictions based on the autoregressive model. This implies that the RNN model does particularly well at forecasting instances in which dysfunctions will take place in the future rather than forecasting the realisations of the target variable in general. As a result, the user can be reasonably confident that – when the model is forecasting dysfunction – it is very likely to materialise.

The out-of-sample performance in key recent periods is also satisfactory. There are signs that these models were picking up a higher probability of market dysfunction in the lead up to September 2022 LDI event in the UK, and especially in the lead up to the March 2023 banking stresses. This performance occurs even though this model was not fed data after end-2020. In contrast, periods such as the onset of Covid-19 were essentially not forecasted by the model, consistent with the fact that the origins of the event were completely outside the financial system.

To better understand the performance of both estimators, we compute the Pearson correlation (Pearson (1895)) between $\hat{\bar{y}}^{AR}_{t+w+h}$ and $\bar{y}_t$, and $\hat{\bar{y}}_{t+w+h}$ and $\bar{y}_t$ for $h = 60$ and $w = 20$ (Graph 3). The correlation between the LSTM predictions and the actual signal peaks at lag 0 of the training sequence. For the test sequence the cross-correlation is flat from -70 to 0 lags at 0.8, indicating that the model is most predictive from 80 to 0 days lags in advance. However, the predictive performance decreases in the test set compared to the training set. This decline is typical of overparameterized machine learning models, suggesting that a longer training sequence would be necessary to improve performance.

In contrast, the correlation between the AR model and the true signal peaks at lag -80. Even when using an AR model with $l = 500$ to match the memory capacity of the RNN model, it still mostly behaves like an AR (1) model, where the best prediction for 60+20 days ahead is today's value. During the test period, where the signal remains stable for most of the sample, this explains why the AR model performs similarly to the RNN model.

Pearson Correlation for the RNN and AR models — Graph 3



The solid lines represent the correlation in the test period (post 1st January 2021) and the dashed lines the correlation prior. The blue lines denote the LSTM model and the green lines the AR model

Source: Authors' calculations.

5.2 Statistical tests of the forecasts

The ability of these models to predict the TAP violations for the Euro-Yen pair is formally evaluated in four complementary ways, all using only out-of-sample data. The first evaluation is a Diebold-Mariano test (Diebold and Mariano (1995), Diebold (2015)) that compares the forecast loss of our neural network to the

forecast loss of an equivalent AR model, using the root mean squared error (RMSE) as a loss function. The second evaluation is also a Diebold-Mariano test, but executed only on dates of actual financial market stress events. The third and most important evaluation directly checks whether the RNN predictions are contemporaneously related to the target variable: if yes, the former can help predict the latter even if it does not perfectly match its scale or contains some bias. In fact, this second evaluation addresses the cases where the neural network produces forecasts that match the future *dynamics* of the true variable, even if it does not provide a precise point estimate. These tests are described in the following paragraphs.

The Diebold-Mariano (DM) test checks whether the pseudo out-of-sample forecast losses of the RNN models are statistically the same as the forecast losses of a corresponding AR model. Due to the random walk nature of financial market variables, this is a reasonable benchmark. The AR is challenging to beat when predicting currency movements (Rossi (2013)), especially at this horizon. The DM test is agnostic to the forecast loss function used. We evaluate these forecasts with a symmetric loss function, namely the RMSE, to assess the null hypothesis that these two forecasts are equal. Since the outcome variable is a 20-day moving average, we select one of every 20 data points when doing this test to avoid autocorrelation. Empirically, the DM test is implemented as a regression of the difference between the RMSE of the RNN out-of-sample forecasts and that of the AR on a constant, while using heteroskedasticity and autocorrelation (HAC) robust standard errors.

These tests show that the RNN forecast losses are systematically higher than those of the AR, both in the complete out-of-sample period (column 1 of Table 3) and on specific dates associated with market stress events (column 2). However, the magnitude suggests this difference is small: the forecast loss difference of 0.0007 is only 39% of the standard deviation of the out-of-sample TAP deviation. In other words, the out-of-sample RMSE of the RNN model (4.i)-(4.iv) is higher but close to that of the AR model (5) in the test set.

However, given that AR predictions essentially carry forward by 60 days the last known value, a more informative test directly checks whether the RNN forecasts are relevant predictors of the true value. These forecast relevance tests take the following form:

$$\bar{y}_{t+w+h} = \hat{\rho} + \hat{\mu}\hat{\bar{y}}_{t+w+h} + \eta_{t+w+h}. \quad (6)$$

Armed with this regression, we inspect both the constant $\hat{\rho}$ and $\hat{\mu}$, the coefficient on the forecast, in both cases using HAC standard errors. The constant identifies whether the forecasts are unconditionally biased in any direction. Ideally, $\hat{\rho}$ would be statistically zero. When it is not zero, $\hat{\rho}$ identifies the direction and magnitude of the forecast bias. The real litmus test is the value of $\hat{\mu}$: controlling for any potential bias, a statistically significant *positive* loading on the contemporaneous forecasted value serves to confirm the relevance of the estimates, especially if this coefficient approaches one, ie the forecast can be, in expectation, mapped one-to-one with the actual variable.

The RNN forecasts serve as statistically significant approximators for the target variable. The results in column 3 of Table 3 demonstrate that the loading of the future TAP deviation on the prediction is not only statistically significant, but also explains a meaningful portion of the former (R-squared of 53.2%). An additional encouraging sign of the relevance of the RNN forecasts comes from an extension of equation (6) to also include the AR forecast as a covariate. As seen in column 4 of Table 3, the RNN forecast continues to be a relevant predictor of the true value, even when considering the AR forecast. If anything, the addition of the latter takes away the significance of $\hat{\rho}$.

All in all, these analyses confirm the relevance of the predictions obtained with RNNs, while noting that this particular model is a useful indicator of the dynamics of the true value 60 days ahead, but not so much of its value (given the systematic bias).

Out-of-sample forecast tests [1] Table 3

	Diebold-Mariano tests		Forecast relevance tests[3]	
	(I)[2]	(II)[2]		
RNN forecast loss minus AR forecast loss	0.0007***	0.0007**		
	(0.0002)	(0.0003)		
Constant			0.0039***	0.0024
			(0.001)	(0.002)
RNN forecast			0.4690***	0.4180***
			(0.069)	(0.082)
AR forecast				0.2471
				(0.218)
Observations	43	10	43	43
R-squared			0.532	0.546

[1] ***/**/* indicates statistical significance at the 1/5/10% level. Standard errors are heteroscedasticity and autocorrelation robust (HAC). [2] Test (I) is executed on every 20th day, since the dependent variable is a 20-day moving average. Test (II) is based on forecasts for dates of market stress only, namely 8 March 2021 (Greensill), 26 March 2021 (Archegos), 22 June 2021 (Evergrande downgrade), 24 February 2022 (Russian invasion of Ukraine), 6 September 2022 (massive margin calls in energy markets), 28 September 2022 (LDI crisis), 1 October 2022 (viral tweet questioning viability of a G-SIB, widely understood to mean Credit Suisse), 10 March 2023 (Silicon Valley Bank collapse), 24 March 2023 (UBS merges with Credit Suisse), 19 October 2023 (peak US Treasury yield, "Treasury Tantrum") A positive (negative) value for the difference between the RNN forecast loss and the AR forecast loss means that the RNN performs worse (better) than the AR. When the value is statistically similar to zero, there is no evidence to reject the null hypothesis that the forecast losses are equal. [3] The dependent variable is the true value of the TAP deviation. A statistically-significant positive loading on the RNN forecast demonstrates its relevance throughout the out-of-sample period, because it shows that an analyst that has only the RNN forecast is able to systematically anticipate the future value of the TAP deviation.

Source: Authors' calculations.

5.3 Variable importance

One of the advantages of our proposed model is that the weights assigned to each variable in (4.iv) provide insight into the importance of each variable daily, and hence what could be driving the predictions of the

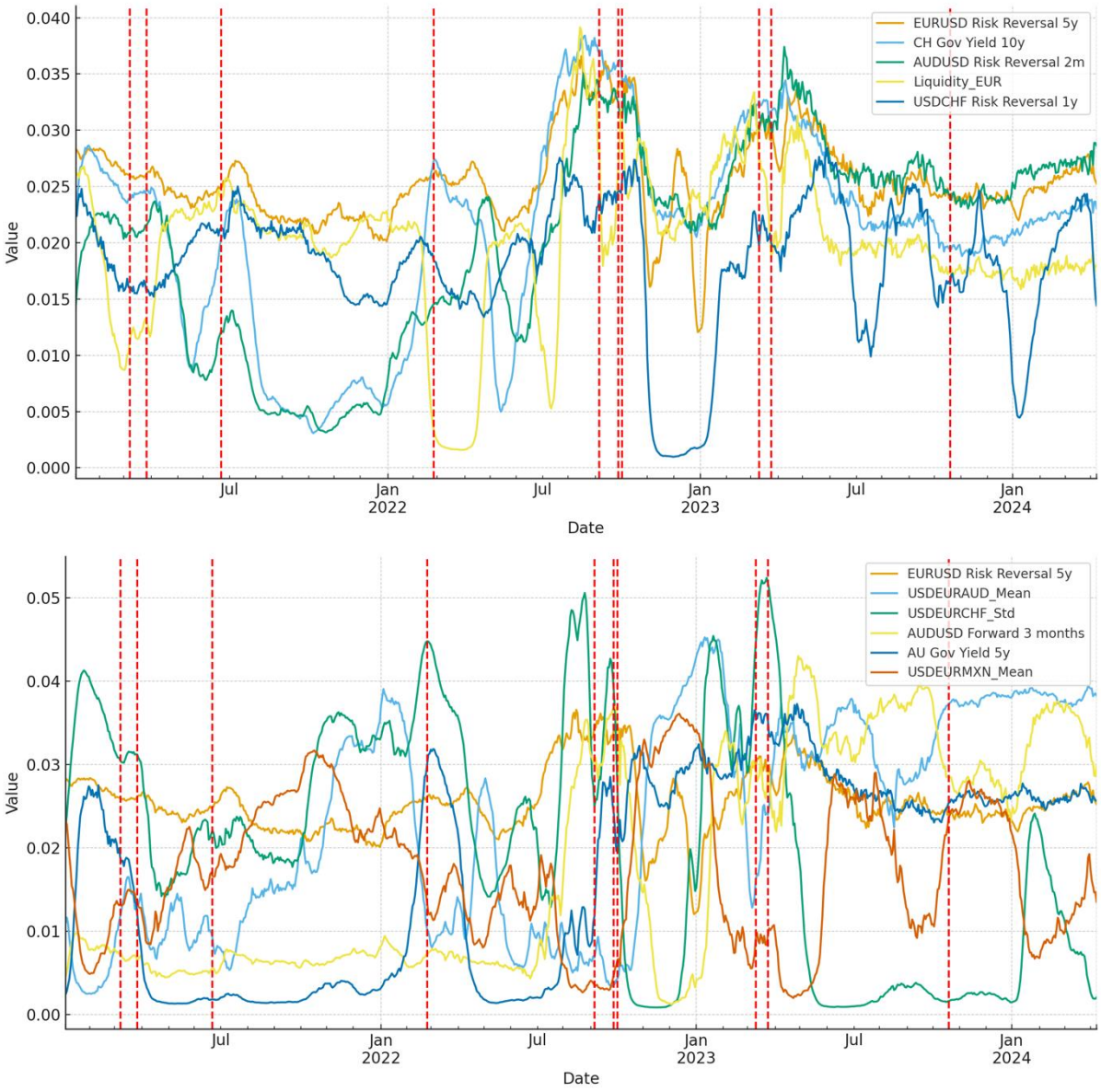
LSTM model. In other words, the model can select at each date the variables that contain a more predictive signal, and this in turn opens up the possibility of inspecting the model's "attention." In Graph 4, we represent the ten series with the largest overall weight during the test period. The EURUSD Risk reversal 5y is plotted in both for comparison, as it possesses the largest mean weight.

One key observation is that changes in the weights of the factors do not always result in changes to the predicted values, and that some changes in the variables do not correspond to identified market stress events. For example, the expected value remains relatively stable up to mid-2022, even though the weights associated with many variables change significantly. For example, the Euro Liquidity changes considerably after the Russian invasion of Ukraine (24 February 2022), on massive margin calls in the energy market (6 September 2022), the LDI crisis (28 September 2022), a Tweet questioning the viability of a G-SIB (1 October 2022), the downfall of Silicon Valley Bank (10 March 2023) and the UBS acquisition of Credit Suisse (CS) (24 March 2023). It does not move significantly for the events in 2021 or the last one in 2024. In all the events, there is a considerable change in at least one of the selected variables, except for the Evergrande downgrade (June 22, 2021). In advance of the March 2023 (USB acquisition of CS) episode, the model increases the weights of two variables: Euro Liquidity and TAP deviations of USD–EUR–CHF (Graph 4), but the 10-year CHF bond yield and the 1-year USDCHF risk reversal remain mostly unchanged.

The model's ability to maintain a consistent prediction while altering the weights of the predictors is a captivating feature. For the specific case of monitoring financial market distress, given our goals, it is interesting that the model predicted market stress in March 2023 by adjusting the weights of different variables. This is particularly noteworthy since the stress originated in the banking sector, leading to widespread market disruption, despite none of the included variables being directly related to the banking sector.

Weights for selected variables during the out-of-sample period

Graph 4



Both charts include the 5-year EUR/USD Risk Reversal for comparison purposes, as it possesses the largest mean weight. The remaining nine series represent those with the highest overall weights. The series depicted in the upper graph exhibit a cosine similarity exceeding 0.94 with the 5-year EUR/USD Risk Reversal, while the series shown in the lower graph demonstrate a cosine similarity below 0.84.

Source: Authors' calculations.

5.4 LLM news search

In this section, we illustrate how to use the insights provided by the variable-specific weights, as seen in the previous subsection, to inform LLM-powered searches. For this use case, we use a sophisticated language model, namely Google's Gemini 2.5 pro. We choose this model because, among the key LLMs available, this one was trained with data that went only up to early 2023, which affords us the possibility to prototype its use in an out-of-sample way for one financial market stress in our study period, namely the "Treasury Tantrum" of October 2023 (Diercks and Asnani (2024)). Other LLMs were trained with more recent data, and therefore any result could be plausibly attributed to the data it had seen before.

We collect a large number of financial news from the first half of July 2023 to mimic the information set available at the time of forecast. Then, we prompt the model to read the news and identify and summarise developments that financial supervisors should monitor, knowing that the RNN forecast assigned particularly high weights to a small set of variables. Based on the context from the financial news and the knowledge about which financial market variables provide more signal, the model lists three candidates for monitoring (Graph 5). Interestingly, the first such item relates to the Treasury Tantrum episode.

To summarise the exercise: suppose the analyst is in July 2023, and sees that the RNN forecasts higher values for TAP deviations for October 2023, as seen in Graph 2. Then, the analyst can use an LLM to sift through current information (financial market news as in this example, or other relevant documents) while knowing which financial market variables contributed more to the forecast. The result is a more targeted search process that increases the odds that the LLM will provide meaningful responses to help the process of monitoring financial markets.

Response of a large language model to a targeted prompt

Prompt based on RNN-provided variable weights Graph 5



Response of Gemini 2.5 Pro (trained with data up to early 2023) to the following prompt, itself informed by the variables identified in Graph 4: "Attached is a collection of financial-market news in the first half of July 2023. We search for predictive signal 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." We provided the model circa 1,000 different financial news from the first half of July 2023 as context.

Source: Authors' elaboration.

6. Conclusions

Our findings show that neural networks with recurrent layers are able to successfully forecast market dysfunctions out-of-sample. In fact, the RNN predictions covary with the target variable even while being smoother. This suggests that these types of models can provide insight into potential looming periods of illiquidity or other adverse market situations.

One novel aspect of the RNNs we present in this paper is their ability to learn how to weight each input variable such that the network can combine their signals to optimise each period's forecast. These

weights are directly visible to the econometrician, providing valuable insights to understand periods forecasted to be adverse, as well as calmer periods.

Building on these findings, we also demonstrate how LLMs can further support financial market monitoring. Armed with these time-varying weights for each variable, the econometrician can select the most important variable in periods of interest and use LLMs to sift through large text datasets, such as financial news (or supervisory texts, internal bank documents etc), to help identify "needles in a haystack" related to periods of forecasted illiquidity. This can further support financial stability monitoring by providing signals to authorities that might otherwise miss.

The performance of our RNN model is tested rigorously but can likely improve further. Finding model configurations that perform well in practice is an empirical exercise: the model parameter space is wide and multi-dimensional, and there is scant guidance on configurations that work well in practice in financial data. In addition, the architecture of neural networks, ie how their layers are set up and connected to each other, is usually very flexible; choices around data transformations are also plentiful. This requires the econometrician to test out a large number of configurations. We hope that our work will contribute to inform this exercise, by offering insights into an architecture and data transformations that works in practice, providing a basis for further improvements.

References

Acharya, V and L Pedersen (2005): "Asset pricing with liquidity risk", *Journal of Financial Economics*, vol 77, no 2, pp 375–410.

Aldasoro, I, P Hördahl and S Zhu (2022a): "Under pressure: market conditions and stress", *BIS Quarterly Review*, September, pp 31–45.

Aldasoro, I, T Ehlers and E Eren (2022b): "Global banks, dollar funding, and regulation", *Journal of International Economics*, vol 137, 103609.

Aldasoro, I, P Hördahl, A Schrimpf and S Zhu (2025): "Predicting financial market stress with machine learning", *BIS Working Papers*, no 1250.

Alessi, L and C Detken (2011): "Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity", *European Journal of Political Economy*, vol 27, no 3, pp 520–533.

Aliyev, N, M Aquilina, K Rzayev and S Zhu (2024): "Through stormy seas: how fragile is liquidity across asset classes and time?", *BIS Working Paper*, no 1229.

Almosova, A and N Andresen (2023): "Nonlinear inflation forecasting with recurrent neural networks", *Journal of Forecasting*, vol 42, pp 240–259.

Araujo, D K G (2023): "gingado: a machine learning library focused on economics and finance", *BIS Working Papers*, no 1122.

——— (2025): "Dynamic nonlinear variable selection in time series", mimeo.

Aquilina, M, M Lombardi, A Schrimpf and V Sushko (2024): "The market turbulence and carry trade unwind of August 2024", *BIS Bulletin*, no 90.

Atkeson, A and L Ohanian (2001): "Are Phillips curves useful for forecasting inflation?", *Federal Reserve Bank of Minneapolis Quarterly Review*.

Bali, T, H Beckmeyer, M Mörke and F Weigert (2023): "Option return predictability with machine learning and big data", *Review of Financial Studies*, vol 36, pp 3548–3602.

Beutel, J, S List and G von Schweinitz (2019): "Does machine learning help us predict banking crises?", *Journal of Financial Stability*, vol 45, 100693.

Bluwstein, K, M Buckmann, A Joseph, M Kang, S Kapadia and O Simsek (2020): "Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach", *Bank of England, Working Paper*, no 848.

Brunnermeier, M and L Pedersen (2009): "Market liquidity and funding liquidity", *Review of Financial Studies*, vol 22, no 6, pp 2201–2238.

Brunnermeier, M and M Oehmke (2013): "Chapter 18 - Bubbles, Financial Crises, and Systemic Risk", *Handbook of the Economics of Finance*, vol 2, part B, pp 1221–1288.

Cao, Y, Z Chen, Q Pei, F Dimino, L Ausiello, P Kumar, K Subbalakshmi and P Ndiaye (2024): "RiskLabs: predicting financial risk using large language model based on multi-sources data", *arXiv:2404.07452* [q-fin.RM].

Chen, T and C Guestrin (2016): "XGBoost: a scalable tree boosting system", KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Chordia, T, A Sarkar and A Subrahmanyam (2005): "An empirical analysis of stock and bond market liquidity", *Review of Financial Studies*, vol 18, no 1, pp 85–129.

Diebold, F X (2015): "Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of Diebold–Mariano tests", *Journal of Business and Economic Statistics*, vol 33(1), 1.

Diebold, F X, & R S Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253-263.

Diercks, A M and D Asnani (2024): "The Treasury tantrum of 2023", *FEDS Notes*, 3 September.

Ding, H, Y Li, J Wang and H Chen (2024): "Large language model agent in financial trading: a survey", *arXiv:2408.06361* [q-fin.TR].

Drehmann, M and V Sushko (2022): "The global foreign exchange market in a higher volatility environment", *BIS Quarterly Review*, December, pp 33–48.

Duprey, T and B Klaus (2017): "How to predict financial stress? An assessment of Markov switching models", *ECB Working Paper*, no 2057.

Fouliard, J, M Howell, H Rey and V Stavrakeva (2021): "Answering the Queen: machine learning and financial crises", *NBER Working Paper*, no 28302.

Garcia, L, U Lewrick and T Sečnik (2021): "Is window dressing by banks systemically important?", *BIS Working Papers*, no 960.

Gelb, A (ed) (1974): *Applied optimal estimation*, MIT Press.

Giannone, D, M Lenza and G Primiceri (2021): "Economic Predictions With Big Data: The Illusion of Sparsity", *Econometrica*, vol 89, pp 2409-2437.

Giglio, S, B Kelly and S Pruitt (2016): "Systemic risk and the macroeconomy: an empirical evaluation", *Journal of Financial Economics*, vol 119, no 3, pp 457–471.

Goodfellow, I, Y Bengio and A Courville (2016): *Deep learning*, MIT Press.

Gradojević, N, D Erdemlioglu and R Gençay (2020): "A new wavelet-based ultra-high-frequency analysis of triangular currency arbitrage", *Economic Modelling*, vol 85, pp 57–73.

Graves, A, S Fernández, F Gomez and J Schmidhuber (2006): "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks", *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp 369–376.

Chuliá, H, C Koser and J Uribe (2021): "Analyzing the nonlinear pricing of liquidity risk according to the market state", *Finance Research Letters*, vol 38, 10515.

Gu, S, B Kelly and D Xiu (2020): "Empirical asset pricing via machine learning", *Review of Financial Studies*, vol 33, no 5, May, pp 2223–2273.

Guidolin, M (2011): "Markov Switching Models in Empirical Finance", In *Missing Data Methods: Time-Series Methods and Applications*; Drukker, D.M., Ed.; Advances in Econometrics; Emerald Group Publishing Limited: Leeds, UK, vol 27, pp. 1–86.

Hamilton, J D (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica*, vol 57, no 2, pp 357–384.

Hauptmann, J, A Hoppenkamps, A Min, F Ramsauer and R Zagst (2014): "Forecasting market turbulence using regime-switching models", *Financial Markets and Portfolio Management*, vol 28, pp 139–164.

He, Zhiguo and Arvind Krishnamurthy (2013): "Intermediary asset pricing", *American Economic Review*, vol 103, no 2, pp 732–770.

——— (2018): "Intermediary asset pricing and the financial crisis", *Annual Review of Financial Economics*, vol 10, pp 173–197.

Hochreiter, S and J Schmidhuber (1997): "Long short-term memory", *Neural Computation*, vol 9, no 8, pp 1735–1780.

Karnaukh, N, A Ranaldo and P Söderlind (2015): "Understanding FX liquidity", *Review of Financial Studies*, vol 28, no 11, pp 3073–3108.

Karolyi, G A, K H Lee and M A Van Dijk (2012): "Understanding commonality in liquidity around the world", *Journal of Financial Economics*, vol 105, no 1, pp 82–112.

Kelly, B T and D Xiu (2023): "Financial machine learning", *University of Chicago, Becker Friedman Institute for Economics Working Paper,* no 2023-100.

Kelly, B T, S Malamud and K Zhou (2024): "The virtue of complexity in return prediction", *Journal of Finance*, vol 79, no 1, pp 459–503.

Lewis, P, E Perez, A Piktus, F Petroni, V Karpukhin, N Goyal, H Küttler, M Lewis, Wen-tau Yih, T Rocktäschel, S Riedel and D Kiela (2020): "Retrieval-augmented generation for knowledge-intensive NLP tasks", *Advances in Neural Information Processing Systems*, vol 33, pp 9459–9474.

Nelson, C R and C I Plosser (1982): "Trends and random walks in macroeconomic time series", *Journal of Monetary Economics*, vol 10, pp 139–162.

Olah, C (2015): "Understanding LSTM Networks", https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Pástor, L and R F Stambaugh (2003): "Liquidity risk and expected stock returns", *Journal of Political Economy*, vol 111, pp 642–685.

Park, T (2024): "Enhancing anomaly detection in financial markets with an LLM-based multi-agent framework", *arXiv:2403.19735* [q-fin.RM].

Pearson, K (1895): "Notes on regression and inheritance in the case of two parents", *Proceedings of the Royal Society of London*, vol 58, pp 240–242.

Pasquariello, P (2014): "Financial market dislocations", *Review of Financial Studies*, vol 27, no 6, pp 1868–1914.

Persaud, A (2003): "Liquidity Black Holes", Risk Books.

Petropoulos, F and S Makridakis (2020): "The M4 competition: bigger. stronger. better", *International Journal of Forecasting*, vol 36, no 1, pp 3–6.

Ranaldo, A and P Santucci de Magistris (2022): "Liquidity in the global currency market", *Journal of Financial Economics*, vol 146, no 3, pp 859–883.

Rossi, B (2013): "Exchange rate predictability", *Journal of Economic Literature*, vol 51, no 4, pp 1063–1119.

Schularick, M and A M Taylor (2012): "Credit booms gone bust: monetary policy, leverage cycles and financial crises, 1870–2008", *American Economic Review*, vol 102, no 2, pp 1029–1061.

Sirignano, J and R Cont (2019): "Universal features of price formation in financial markets: perspectives from deep learning", *Quantitative Finance*, vol 19, no 9, pp 1449–1459.

Smyl, S (2020): "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting", *International Journal of Forecasting*, vol 36, no 1, pp 75–85.

Snellman, O (2025): "Using a Transformer Network to Measure Fragility in the Financial System". Available at SSRN: https://ssrn.com/abstract=4802476

Stock, J H and M W Watson (2007): "Why has US inflation become harder to forecast?", *Journal of Money, Credit and Banking*, vol 39, issue s1, pp 3–33.

Tirole, J (2011): "Illiquidity and all its friends", *Journal of Economic Literature*, vol 49, no 2, pp 287–325.

Tölö, E (2020): "Predicting systemic financial crises with recurrent neural networks", *Journal of Financial Stability*, vol 49, 100746.

Vaswani, A, N Shazeer, N Parmar, J Uszkoreit, L Jones, A N Gomez, Ł Kaiser and I Polosukhin (2017): "Attention is all you need", *Advances in Neural Information Processing Systems*, vol 30, pp 5998–6008. Vuletić, M, F Prenzel and M Cucuringu (2024): "Fin-GAN: forecasting and classifying financial time series via generative adversarial networks", *Quantitative Finance*, vol 24, no 2, pp 177–199.

Ward, F (2017): "Spotting the danger zone: forecasting financial crises with classification tree ensembles and many predictors", *Journal of Applied Econometrics*, vol 32, pp 359–378.

Zalla, R (2017): "Economic policy uncertainty in Ireland", *Atlantic Economic Journal*, vol 45, no 2, pp 269–271.

Zou, H and T Hastie (2005): "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol 67, no 2, pp 301–320.

Appendix

A.1 Generative AI as an interpretative and 'early warning' tool

Generative AI and LLMs can be used to assist supervisors and regulators in validating and interpreting the anomalies predicted using AI and ML models. This process also aids in assessing the potential causes of future dislocations, based on a real-time search of financial news and other relevant information.

Identifying probable predictors of market dislocations is useful, even without an assessment of what are the potential causes of such future dislocations. Supervisors could use this information to prepare for ex post intervention, preventing more severe consequences. However, an evaluation of the likely underlying causes of such dislocations would enhance the tool's usefulness even further. Supervisors would then be able to address the causes of such dislocations, preventing situations from deteriorating before they occur. By leveraging the capabilities of LLMs to sift news in real time and aggregate information, supervisors could have access to potential explanations and decide on an appropriate course of action.

To illustrate this potential, we tested the capacity of generative LLMs using the financial market disruption on 5 August 2024 (Aquilina et al (2024)). The goal was to determine whether generative LLMs can identify critical news concisely, ensuring that financial supervisors are aware of emerging risk signals. For this purpose, we compiled 200-300 financial market news articles per day for the two business days leading up to the disruption. The generative LLM was then tasked with summarising the most critical news items that a financial supervisory institute should monitor to detect potential upcoming market dysfunctions, such as liquidity shortages.[6] The LLM was instructed to condense this information into three bullet points, including the reasons and relevance for each. Table A1.1 below summarises the results.

This example illustrates how the generative LLM can identify critical and relevant news that, based on information available before the event, are now known to have triggered the disruption. For instance, the LLM pinpointed an aggressive shift in policy stance by major central banks in response to indications of a US economic slowdown, rising market volatilities and bearish market sentiments. Moreover, it accurately explained that the Bank of Japan's hawkish stance could lead to a further strengthening of the yen, potentially disrupting global markets that heavily rely on the yen as a funding currency. This, in turn, could trigger a rapid unwinding of carry trades—a central issue widely discussed in the aftermath of the disruption. Notably, this task is accomplished without any prior training (ie zero-shot learning).

While we highlight that this example serves to explore and illustrate rather than formally test and confirm the ability of LLMs to correctly explain events before or as they unfold, much more can be done. For example, with an information retrieval system, the LLM is able to combine its extensive knowledge base with the external data source of news articles, a process known as retrieval-augmented generation (RAG). In addition to using LLMs to sift news sources to highlight potential causes of future dislocations, one can also use a multi-agent framework in the spirit of Park (2024) to harness a collaborative network of AI agents and coordinate their answers in the pursuit of the common objective to explain the potential cause of future dislocations. The important component is that this neural network architecture enables a targeted inspection, based on the variables identified in Graph 4.

[6] The prompt is: "Attached is a collection of financial market news for a specific day. Summarise the most critical news for a financial supervisory institute to monitor any potential market dysfunction (eg shortage of liquidity) in the coming days. Summarise in three bullets according to the importance and relevance. Briefly explain why they are important."

Prompt and response to large language model about August 2024 Yen turbulence — Table A1.1

1 August 2024	2 August 2024
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 involving 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. **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. **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 spillover 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 will be crucial.	**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. **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. **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 Nasdaq 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 potential systemic risks.

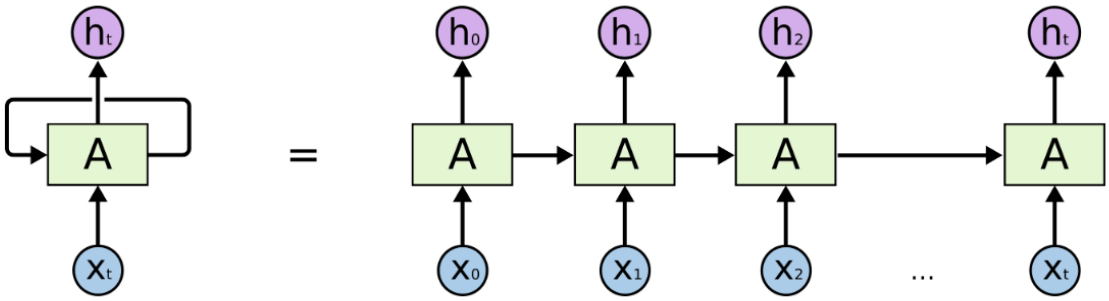
Source: Authors.

A.2 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are an architecture for deep neural networks that operate with sequential data, such as time series, speech, or text (Goodfellow et al. (2016)). RNNs take a new input, update their state (also called "memory"), and produce an output at each time step (illustrated in Graph A2.1). The output of an RNN – the target variable - depends solely on the current input data and on the stored memory. The memory efficiently stores all relevant information from previous inputs. Economic uses of RNNs include inflation forecasting (Almosova and Andresen (2023)) and well-performing submissions to forecasting competitions. In fact, the winner of a prominent forecasting competition (Petropoulos and Makridakis (2020)) combines RNNs with exponential smoothing (Smyl (2020)).

RNNs function similarly to Kalman filters (Gelb (1974)) on state-space models, with the main difference being that the relationships between inputs, outputs, and the memory update functions are nonlinear. Further, the nonlinearities are flexibly learned from the data, as opposed to modelled explicitly by the analyst. For example, in traditional econometrics a prototypical nonlinear model is the inclusion of a squared regressor in the estimation. In contrast, all a neural network requires is the set of regressors and the corresponding dependent variables to model, with sufficient examples, the (potentially much more complex) nonlinear form that best fits the data.

Scheme of a typical recurrent neural network — Graph A2.1



X_t represents the input to the RNN, *A* the hidden layer and memory, and $h_t$the output. The right-side plot unfolds the RNN structure to illustrate left-to-right how each input only affects the current state of the network.

Source: Olah (2015).

However, early types of RNNs were unable to learn long-term relationships because the memory is constantly updated with new information. Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber (1997)) address this vanishing memory problem by adding a set of "gates" to the memories of the RNNs.[7] These gates are themselves neuronal layers whose output is a number between 0 and 1. This allows the gates to automatically control, based on data, the addition of new relevant information to the memory, the removal of information that is no longer relevant, and the retention of long-term concepts that are useful to predict the output. LSTMs also include a gate to decide which memories are active for each output.[8]

Exploring the structure of an LSTM network provides insight about how and why they function well. The key insight is that the data observed at period t and the result from the previous period both help to jointly determine how much of the previous state persists, as well as if anything from current data should be stored in the state-space. For example, a date with a particularly stark TAP deviation can make the model forget some of its state carried over from past periods and absorb a new value for the state,

[7] LSTMs, combined with connectionist temporal classification (CTC) of Graves et al (2006), solved the speech-to-text and OCR problems in the early 2010s. A comprehensive introduction to RNN and LSTMs can be found in Olah (2015).

[8] While recent advances in AI are underpinned by another neural network, the Transformers (Vaswani et al (2017)), LSTMs remain a competitive technique for data of the scale typically used in economics and financial settings. RNNs (including LSTM networks) process data sequentially and are not well-suited for parallel processing and thus do not take full advantage of parallel computing power. Transformers resolve this issue by eliminating recurrency through positional encoding. This allows these networks to train on much larger amounts of data, underpinning their much wider application compared to LSTMs. In economic or financial time series, however, LSTMs continue to perform well, if not better due to generally lower data requirements compared to the transformer.

similar to how Markov regime switching models work. In turn, this new value for the state influences how the model will forecast given the same input data (as if the parameters in a structural model changed depending on whether the period is in a "good" or in a "bad" state). Being a neural network, these interactions all occur nonlinearly, further adding flexibility to the model.

Formally, the LSTM process goes as follows. At each period, new observed inputs x_t enter from the bottom, while a vector of outputs (also known as "hidden states") h_t are produced at the top and transmitted to the subsequent stage too, where another version of this LSTM with the same parameters will repeat this process. The "long-term" information (also known as "carry state") flows from left to right, represented by the vector c_t. The top line in the Graph A2.2 scheme represents c_t. The values of this vector can be modified by a nonlinear combination of the new input x_t and the previous state h_{t-1}. In the first time step, both c_t and h_t are zero-valued vectors. The LSTM has three gate functions:

1. Forget Gate (Labelled 1): This is a single layer fully connected neural network with a sigmoid activation function σ that constrains the output to be between 0 and 1. If the sigmoid output is zero, the network will erase the memory c_t. Conversely, an output of one will allow the information in c_t to pass through fully. Formally, the forget gate is:

$$f_t = \sigma(b^f + W^f x_t + U^f h_{t-1}).$$

2. Input Gate (Labelled 2): This consists of two networks. The first is a single layer fully connected neural network with a sigmoid activation function that decides whether to add information to the memory (the opposing operation to the forget gate). The second is another single layer fully connected neural network with a hyperbolic tangential activation function ($tanh$) that limits the output to a range between -1 and +1. This second layer determines the specific information to be added, while the first network determines *how much* of this new information will effectively be added to the long-term carry state c_t. Formally, the inputs are:

$$i_t = \sigma(b^i + W^i x_t + U^i h_{t-1})$$

$$\tilde{c}_t = tanh(b^c + W^c x_t + U^c h_{t-1}).$$

After an LSTM cell receives the carry state from the previous time step, c_{t-1} and it goes through the forget gate and the input gate in sequence (the first and second multiplications in the equation below), it becomes the cell's current carry state, as such:

$$c_t = c_{t-1} \odot f_t + \tilde{c}_t \odot i_t,$$

with $\odot$ representing the Hadamard point-wise multiplication.

3. Output Gate (Labelled 3): This sigmoid network determines which part of the (now updated) memory is exposed at time t to be the LSTM cell's output. If the value is zero, no memory is exposed. If the value is one, the updated memory c_t is presented as the output h_t. Formally:

$$o_t = \sigma(b^o + W^o x_t + U^o h_{t-1})$$

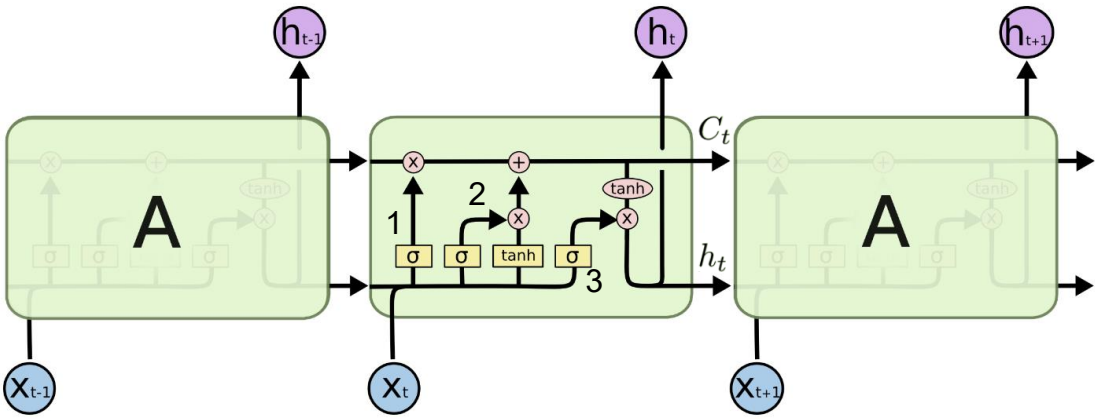
$$h_t = o_t \odot tanh\, c_t.$$

In all above regressions, $x_t \in \mathbb{R}^N$ is the column vector of N covariates, c_t, $h_t \in \mathbb{R}^{d_{LSTM}}$ are vectors of the same dimension $d_{LSTM} > 0$. Each element of $\{W^f, W^i, W^c, W^o\} \in \mathbb{R}^{d_{LSTM} \times N}$ are the parameter matrices related to contemporaneous inputs and each element of $\{U^f, U^i, U^c, U^o\} \in \mathbb{R}^{d_{LSTM} \times d_{LSTM}}$ are the parameter matrices related to the previous state. The elements of $\{b^f, b^i, b^c, b^o\} \in \mathbb{R}^{d_{LSTM}}$ are the column vectors of constants (known in the computer science literature as "biases"). And the operations described above by the Hadamard multiplication are independently applied to each vector component. In general, we represent an LSTM network as follows:

$$h_t = LSTM(x_t, d_{LSTM}).$$

Scheme of an LSTM network

Graph A2.2



X_t represents the input to the RNN, C_t the hidden layer and memory, and h_t the output.

Source: Olah (2015).

A.3 Prediction architecture

The purpose of our model is to map an extensive panel of daily financial market variables into h-day ahead predictions of the market anomalies for a window of w days. In what follows, we illustrate the technique with an example using $h = 60$ and $w = 20$, that is, a model that forecasts the average of the 20-day window that starts 60 days into the future. We use a novel architecture inspired by insights that financial market variables can interact nonlinearly, and that these interactions (and the importance of each variable) can change over time. The model also flexibly deals with seasonality. Importantly, in addition to the predictions, the network offers interesting by-products such as the time series of variable importance.

The architecture of the network is shown in Graph A3.1 below. Each rectangle is a different layer of the network. Starting from the top towards the bottom, the network takes in two types of data: a time series of two categorical variables (calendar quarter and a dummy for days around quarter-ends) and a time series of 125 continuous variables representing financial market series. The data is then sequentially processed by the network, as shown in the directed graph, eventually producing the TAP deviation forecast and the variable weights. It can be instructive to analyse the neural network as separate building blocks, next.

Temporal features. Seasonality can play a key role in financial market series. In our case, for example, many of the variables including TAP deviations themselves, display considerable changes at end-of-quarter periods. In addition, there might be other, less obvious ways in which seasonality can play a role. For this reason, we embed each quarter $Q_t \in \{1,2,3,4\}$ and a quarter-end dummy $QEnd_t \in \{0,1\}$, defined as 1 for the five days before and five days after quarter-ends. These embeddings map each possible value of Q_t and of $QEnd_t$ to a real-valued vector of dimension $d_{TempFeat}$, as such: $M_Q: Q_t \rightarrow \mathbb{R}^{d_{TempFeat}}$. In this case, we use $d_{TempFeat} = 16$. This results in two time series of embeddings, one of each

time variable. They are then concatenated to result in a single vector time series of dimension $2 \times d_{TempFeat}$. These temporal features inform the decision on the variable specific-weights, ω_t, as well as the final prediction.

Competitive variable selection. The vector time series of temporal features is concatenated with the time series of the N different financial market variables into a single, large time series vector with dimension $2 \times d_{TempFeat} + N$. This large panel time series is then fed to an LSTM layer that combines all these signals nonlinearly. At the end of this process, the network combines the information into an N-dimensional vector. Then at each point in time, these N values undergo a softmax transformation, which maps each result into a positive number that highlights the largest values with all summing to one. The "temperature" or smoothness of this transformation is a learned parameter. The temperature parameter determines just how competitive this process is (lower values make it more competitive). All variables are normalised; thus the weight is independent of each variable's scale. These time-varying, variable specific weights are then multiplied by each variable's series, resulting in competitively weighted variables. For some variables, the weight might be very close to zero.

Final transformation. Once weighted, the variables pass through a densely connected layer (with weights shared across time steps) to cast the signals into a uniform geometry and then another LSTM layer with nonlinear activation to extract increasingly refined features to predict the anomaly in question. Finally, another densely connected layer with output of dimension 1 results in the prediction for the TAP variable.

The proposed algorithm can be presented as follows:

$$u_t = \begin{bmatrix} W^1 Q_t \\ W^2 QEnd_t \end{bmatrix}$$

$$o_t = LSTM\big(LSTM\big(\big[\tilde{u}_t, x_t\big], 16\big), 125\big) \quad (4)$$

$$\tilde{o}_t = VARSEL(o_t)$$

$$\tilde{x}_t = \tilde{o}_t \odot x_t$$

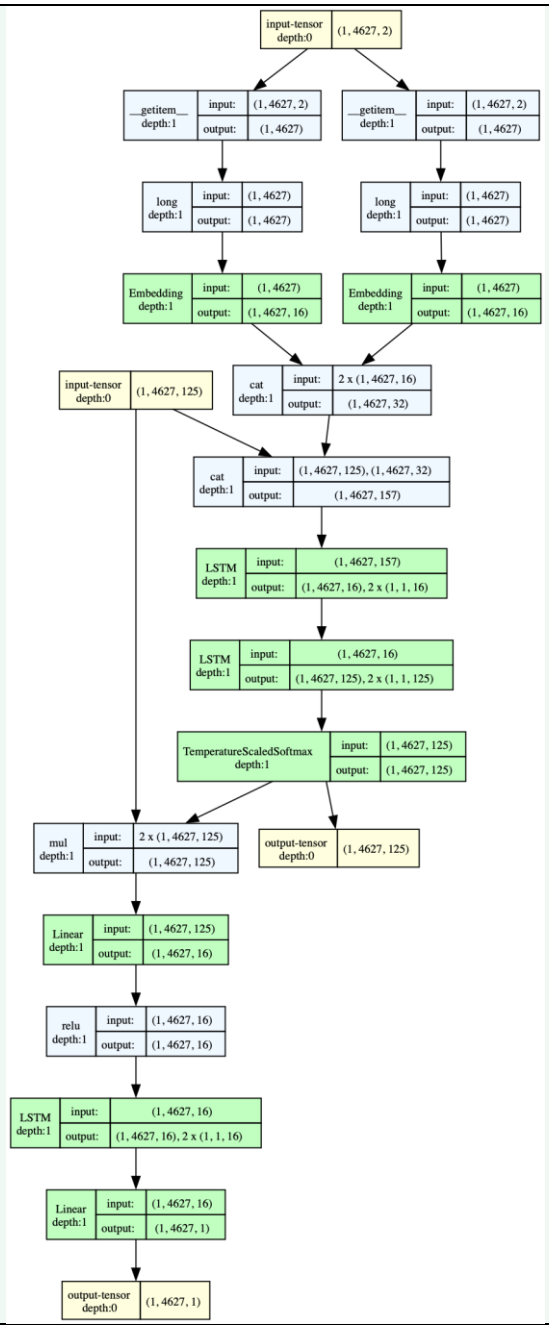
$$\hat{y}_{t+h,w} = LSTM\big(LSTM\big(\big[\tilde{u}_t, \tilde{x}_t\big], 16\big), 1\big),$$

where W^1 and W^2 are 4-dimensional real vectors and x_t is the period-t vector with the input variables, as detailed in the following section. The function VARSEL applies a softmax transformation across the 125 elements in each period, with a temperature parameter τ that is learned to optimise predictive ability:

$VARSEL(o_t) = \frac{e^{o_t^{(j)}/\tau}}{\sum_{j=1}^{125} e^{o_t^{(j)}/\tau}}$. This softmax transformation essentially highlights the differences between the weights of each variable. If the weights were all the same, the values would be equal to 1/125. But if, say, one variable tends to provide richer signal and therefore gets higher weight from the model, its share grows sharply as its weight becomes higher than other variables. Just how sharp is determined by the temperature τ>0: lower values tend to accentuate differences, whereas higher values (especially if above one) attenuate differences. The result is a time series of equal dimension to the financial market inputs, but with weights between 0 and 1 for each variable. Importantly, these variable-specific weights change over time, depending on the data up until that time. Further, the weights sum to one for each period, meaning they are "competitive" in the sense that a higher weight for one variable necessarily implies lower weights for the others.

Architecture of the prediction neural network

Graph A3.1



Neural network to predict 20-day TAP averages starting 60 days ahead used to illustrate the neural network architecture. The calculation graph from the input data to the TAP predictions flows from top to bottom. Numbers in parenthesis represent respectively a unit sample (always set to one in this example), the number of time steps in the full sample and the dimensionality of the data at each calculation node.

Source: Authors.

A.4 Input data

This Appendix provides descriptive statistics for the input data that are not described in Table 1 of the main text.

Input data: Foreign exchange bid-ask spreads

As a percentage of the bid-ask midpoint

Table A4.1

	Full sample							Training set	Test set
	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
USD-AUD	0.0393	0.0088	0.0247	0.0330	0.0376	0.0448	0.1306	0.0410	0.0313
USD-GBP	0.0233	0.0046	0.0157	0.0196	0.0225	0.0268	0.0518	0.0223	0.0276
USD-EUR	0.0225	0.0054	0.0134	0.0182	0.0220	0.0261	0.0819	0.0211	0.0290
USD-JPY	0.0263	0.0074	0.0152	0.0207	0.0244	0.0310	0.0537	0.0280	0.0188
USD-CHF	0.0374	0.0090	0.0206	0.0315	0.0365	0.0413	0.0719	0.0374	0.0373

Sources: LSEG Tick History; authors' calculations.

Input data: Equity market returns

In percent

Table A4.2

	Full sample							Training set	Test set
	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
S&P/ASX 200	0.02	1.06	-9.70	-0.48	0.03	0.56	7.00	0.01	0.02
Euro Stoxx 50	0.02	1.36	-12.40	-0.59	0.02	0.65	11.00	0.01	0.05
Nikkei 225	0.03	1.41	-11.41	-0.61	0.00	0.71	14.15	0.02	0.05
Swiss Market	0.01	1.06	-9.64	-0.46	0.03	0.52	11.39	0.01	0.01
FTSE 100	0.01	1.13	-10.87	-0.47	0.01	0.55	9.84	0.01	0.03
S&P 500	0.04	1.22	-11.98	-0.39	0.04	0.56	11.58	0.04	0.04

Sources: Bloomberg; authors' calculations.

Input data: Volatility indices

Table A4.3

	Full sample							Training set	Test set
	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
CBOE Volatility Index	19.62	8.97	9.14	13.57	17.21	22.79	82.69	19.51	20.11
S&P/ASX 200 Volatility Index	15.88	9.14	0.00	11.92	14.37	19.32	66.72	16.40	13.47
Euro Stoxx 50 Volatility Index	22.60	8.86	10.68	16.67	20.66	25.91	87.51	22.97	20.92

Nikkei Stock Average Volatility Index	23.72	8.87	12.19	18.39	21.57	26.31	92.03	24.39	20.67
Deutsche Borse VSMI Volatility Index	17.81	7.42	8.81	13.37	15.79	19.66	84.90	18.24	15.82
MOVE Index	84.56	32.99	36.62	60.00	74.51	103.69	264.60	80.86	101.49

Sources: Bloomberg; authors' calculations.

Input data: Spot exchange rates

In percentage changes

Table A4.4

	Full sample							Training set	Test set
	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
AUD-USD	0.00	0.78	-8.69	-0.39	0.02	0.41	7.18	0.00	-0.02
CHF-USD	0.01	0.63	-8.66	-0.31	-0.01	0.30	14.39	0.01	0.00
EUR-USD	0.00	0.55	-2.86	-0.30	0.01	0.29	4.01	0.00	-0.01
GBP-USD	-0.01	0.59	-7.10	-0.33	0.00	0.32	5.38	-0.01	-0.01
JPY-USD	0.00	0.61	-3.45	-0.33	-0.01	0.29	5.00	0.00	-0.04

Sources: Bloomberg; authors' calculations.

Input data: Forward exchange rates

In percentage changes

Table A4.5

	Full sample							Training set	Test set
	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
1-week									
AUD-USD	0.00	0.78	-8.36	-0.39	0.02	0.40	7.15	0.00	-0.02
EUR-USD	0.00	0.55	-2.61	-0.30	0.01	0.29	4.01	0.00	-0.01
GBP-USD	-0.01	0.60	-7.62	-0.33	0.00	0.32	5.37	-0.01	-0.01
JPY-USD	0.01	0.61	-4.43	-0.28	0.01	0.32	3.55	0.00	0.05
2-week									
AUD-USD	0.00	0.78	-8.74	-0.38	0.02	0.40	7.17	0.00	-0.02
EUR-USD	0.00	0.55	-2.59	-0.30	0.01	0.29	4.00	0.00	-0.01
GBP-USD	-0.01	0.60	-7.61	-0.33	0.00	0.32	5.36	-0.01	-0.01
JPY-USD	0.01	0.59	-10.58	-0.24	0.00	0.28	3.55	0.00	0.05
1-month									
AUD-USD	0.00	0.78	-8.38	-0.39	0.02	0.40	7.15	0.00	-0.02
EUR-USD	0.00	0.55	-2.61	-0.30	0.01	0.29	3.98	0.00	-0.01
GBP-USD	-0.01	0.60	-7.60	-0.33	0.00	0.32	5.33	-0.01	-0.01
JPY-USD	0.01	0.61	-4.41	-0.28	0.01	0.32	3.56	0.00	0.05
2-month									

AUD-USD	0.00	0.78	-8.37	-0.39	0.02	0.40	7.10	0.00	-0.02
EUR-USD	0.00	0.55	-2.61	-0.30	0.01	0.29	3.97	0.00	-0.01
GBP-USD	-0.01	0.60	-7.58	-0.33	0.00	0.32	5.29	-0.01	-0.01
JPY-USD	0.01	0.61	-4.45	-0.28	0.01	0.32	3.55	0.00	0.05
3-month									
AUD-USD	0.00	0.78	-8.37	-0.39	0.02	0.40	7.00	0.00	-0.02
EUR-USD	0.00	0.55	-2.61	-0.30	0.01	0.29	3.96	0.00	-0.01
GBP-USD	-0.01	0.60	-7.58	-0.33	0.00	0.32	5.25	-0.01	-0.01
JPY-USD	0.01	0.61	-4.40	-0.28	0.01	0.32	3.55	0.00	0.05
6-month									
AUD-USD	0.00	0.77	-8.38	-0.39	0.02	0.40	6.97	0.00	-0.02
EUR-USD	0.00	0.54	-2.61	-0.30	0.01	0.29	3.92	0.00	-0.01
GBP-USD	-0.01	0.59	-7.55	-0.33	0.00	0.32	5.22	-0.01	-0.01
JPY-USD	0.01	0.60	-4.37	-0.28	0.01	0.32	3.56	0.00	0.04

Sources: Bloomberg; authors' calculations.

Input data: 25 delta risk reversals[1]

Table A4.6

	Full sample							Training set	Test set
	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
1-month									
AUD-USD	-1.31	1.09	-8.25	-1.58	-1.02	-0.69	0.48	-1.36	-1.05
EUR-USD	-0.53	0.79	-4.18	-0.89	-0.40	0.02	2.87	-0.54	-0.49
GBP-USD	-0.81	0.84	-9.67	-1.12	-0.64	-0.31	0.58	-0.80	-0.87
USD-JPY	-1.10	1.35	-10.35	-1.44	-0.88	-0.36	1.04	-1.17	-0.79
USD-CHF	-0.12	0.66	-3.25	-0.53	-0.21	0.29	2.94	-0.15	0.01
2-month									
AUD-USD	-1.54	1.12	-8.05	-1.91	-1.26	-0.84	0.05	-1.61	-1.23
EUR-USD	-0.66	0.87	-4.23	-1.16	-0.52	-0.03	2.60	-0.68	-0.59
GBP-USD	-1.00	0.86	-7.63	-1.37	-0.83	-0.46	0.45	-0.99	-1.03
USD-JPY	-1.24	1.40	-10.20	-1.57	-1.03	-0.49	1.24	-1.32	-0.84
USD-CHF	-0.14	0.70	-3.14	-0.61	-0.26	0.30	2.60	-0.17	-0.02
3-month									
AUD-USD	-1.74	1.19	-8.10	-2.19	-1.46	-0.94	-0.08	-1.82	-1.36
EUR-USD	-0.77	0.93	-4.35	-1.32	-0.64	-0.05	2.23	-0.79	-0.67
GBP-USD	-1.15	0.88	-6.37	-1.63	-1.00	-0.58	0.54	-1.15	-1.15
USD-JPY	-1.35	1.47	-10.45	-1.70	-1.12	-0.60	1.46	-1.46	-0.86
USD-CHF	-0.16	0.74	-3.05	-0.68	-0.30	0.31	2.36	-0.19	-0.06
6-month									
AUD-USD	-2.03	1.26	-8.20	-2.58	-1.78	-1.16	-0.05	-2.13	-1.57
EUR-USD	-0.91	1.01	-4.40	-1.54	-0.82	-0.07	1.92	-0.94	-0.80
GBP-USD	-1.34	0.88	-4.62	-1.88	-1.26	-0.76	0.70	-1.34	-1.36

USD-JPY	-1.51	1.57	-10.30	-1.85	-1.25	-0.70	1.57	-1.65	-0.84
USD-CHF	-0.20	0.79	-2.90	-0.76	-0.35	0.30	2.30	-0.21	-0.15
1-year									
AUD-USD	-2.30	1.31	-8.30	-2.95	-2.05	-1.38	-0.23	-2.42	-1.79
EUR-USD	-1.02	1.06	-4.48	-1.70	-0.97	-0.15	1.47	-1.03	-0.95
GBP-USD	-1.50	0.87	-4.43	-2.05	-1.51	-0.91	0.45	-1.47	-1.61
USD-JPY	-1.67	1.70	-10.30	-2.08	-1.41	-0.77	1.41	-1.87	-0.77
USD-CHF	-0.25	0.86	-2.85	-0.88	-0.40	0.28	2.33	-0.25	-0.25
5-year									
AUD-USD	-2.47	1.06	-8.30	-3.08	-2.40	-1.82	1.34	-2.52	-2.25
EUR-USD	-0.68	0.74	-3.22	-1.20	-0.54	-0.10	0.60	-0.70	-0.57
GBP-USD	-1.38	0.69	-4.27	-1.72	-1.45	-1.17	0.84	-1.29	-1.81
USD-JPY	-2.45	1.94	-10.63	-3.21	-2.37	-1.14	1.11	-2.78	-0.93
USD-CHF	-0.66	0.60	-2.01	-1.15	-0.82	-0.20	1.08	-0.59	-0.97

[1] Defined as the implied volatility for call options minus the implied volatility for put options on the base currency with the same delta.

Sources: Bloomberg; authors' calculations.

Input data: Government bond/T-Bill yields

In percent

Table A4.7

	Full sample							Training set	Test set
	Mean	S.D.	Min	25%	50%	75%	Max	Mean	Mean
3-month									
Germany	0.60	1.63	-1.04	-0.62	-0.02	0.93	4.48	0.50	1.05
Japan	0.03	0.24	-0.48	-0.14	0.00	0.11	0.73	0.07	-0.14
US	1.38	1.83	-0.14	0.06	0.23	2.17	5.51	1.10	2.65
6-month									
US	1.49	1.85	-0.03	0.11	0.41	2.30	5.58	1.20	2.82
1-year									
Australia	2.96	1.84	-0.01	1.73	2.56	4.23	7.24	3.12	2.20
Germany	0.77	1.68	-0.93	-0.59	0.05	1.47	4.69	0.66	1.27
Switzerland	-0.33	0.58	-2.22	-0.86	-0.11	0.00	1.63	-0.41	0.05
UK	1.49	1.86	-0.15	0.35	0.56	1.98	5.89	1.28	2.46
US	1.56	1.78	0.04	0.17	0.56	2.42	5.51	1.28	2.85
2-year									
Australia	3.00	1.82	0.01	1.82	2.67	4.32	7.17	3.16	2.24
Germany	0.80	1.62	-1.00	-0.60	0.12	1.76	4.70	0.73	1.13
Japan	0.12	0.32	-0.36	-0.12	0.05	0.18	1.09	0.15	-0.04
Switzerland	0.18	1.08	-1.12	-0.76	-0.04	0.67	3.06	0.16	0.29
UK	1.57	1.78	-0.16	0.40	0.71	2.00	5.83	1.38	2.42
US	1.69	1.60	0.10	0.41	0.94	2.56	5.28	1.45	2.77

5-year									
Australia	3.25	1.74	0.25	2.09	3.04	4.84	6.87	3.40	2.59
Germany	1.11	1.59	-0.99	-0.31	0.61	2.38	4.76	1.12	1.06
Japan	0.29	0.46	-0.38	-0.09	0.17	0.48	1.58	0.33	0.07
Switzerland	0.46	1.12	-1.16	-0.54	0.17	1.14	3.19	0.47	0.38
UK	1.99	1.58	-0.14	0.76	1.43	2.95	5.76	1.91	2.36
US	2.19	1.27	0.19	1.30	1.82	2.87	5.23	2.06	2.77
10-year									
Australia	3.61	1.57	0.61	2.57	3.46	5.17	6.79	3.74	2.99
Germany	1.63	1.53	-0.86	0.31	1.52	3.02	4.68	1.72	1.22
Japan	0.66	0.62	-0.29	0.07	0.56	1.23	2.01	0.73	0.32
Switzerland	0.87	1.13	-1.12	-0.12	0.72	1.85	3.42	0.94	0.57
UK	2.51	1.44	0.08	1.30	2.17	3.75	5.55	2.50	2.51
US	2.76	1.08	0.51	1.94	2.61	3.56	5.29	2.73	2.90

Sources: Bloomberg; LSEG Datastream; authors' calculations.