

BIS Working Papers No 1250

Predicting financial market stress with machine learning

by Iñaki Aldasoro, Peter Hördahl, Andreas Schrimpf and
Xingyu Sonya Zhu

Monetary and Economic Department

March 2025

JEL classification: G01, C53, G17, G12, G28.

Keywords: Machine learning, financial stress, quantile
regressions, forecasting, Shapley value.

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2025. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Predicting financial market stress with machine learning*

Iñaki Aldasoro
BIS

Peter Hördahl
BIS

Andreas Schrimpf
BIS & CEPR

Xingyu Sonya Zhu
BIS

February, 2025

Abstract

Using newly constructed market condition indicators (MCIs) for three pivotal US markets (Treasury, foreign exchange, and money markets), we demonstrate that tree-based machine learning (ML) models significantly outperform traditional time-series approaches in predicting the full distribution of future market stress. Through quantile regression, we show that random forests achieve up to 27% lower quantile loss than autoregressive benchmarks, particularly at longer horizons (3–12 months). Shapley value analysis reveals that funding liquidity, investor overextension and the global financial cycle are important predictors of future tail realizations of market conditions. The MCIs themselves play a prominent role as well, both in the same market (self-reinforcing dynamics *within* markets) and across markets (spillovers *across* markets). These results highlight the value of ML in forecasting tail risks and identifying systemic vulnerabilities in real time, bridging the gap between high-frequency data and macroeconomic stability frameworks.

JEL Codes: G01, C53, G17, G12, G28.

Keywords: machine learning, financial stress, quantile regressions, forecasting, Shapley value.

*All authors are with the Bank for International Settlements (BIS), Monetary and Economic Department, Centralbahnplatz 2, CH-4002 Basel, Switzerland. *Corresponding author:* Sonya Zhu (sonya.zhu@bis.org). We would like to thank Jon Frost, Marco Lombardi, Fernando Pérez-Cruz, and seminar participants at the BIS for helpful comments and suggestions. The views expressed here are those of the authors only, and not necessarily those of the BIS. Declaration of interest: none.

1 Introduction

Financial market stress is a persistent threat to macroeconomic stability, with cascading effects on credit provision, asset prices, and economic growth. The Great Financial Crisis (GFC), the Covid “dash for cash” and recurring episodes of market illiquidity underscore the systemic risks posed by unstable and malfunctioning financial markets. Such episodes often originate in seemingly isolated segments – such as money markets or FX swaps – before propagating globally, as interconnected intermediaries and leveraged investors amplify shocks.¹

Policymakers and academics alike have long sought tools to measure and forecast these stress dynamics in real time. Traditional approaches, including financial stress indices (FSIs) and financial conditions indices (FCIs), provide aggregate snapshots of market health but often conflate broad sentiment shifts (e.g., equity volatility via the VIX) with structural vulnerabilities like liquidity shortages or arbitrage breakdowns. This conflation limits their usefulness in identifying market-specific stress, which is critical for targeted interventions. Addressing these gaps requires a framework that prioritizes real-time data and accommodates non-linear dynamics – a task uniquely suited to machine learning (ML).

This paper makes two interrelated contributions. First, we construct novel market condition indicators (MCIs) for three pivotal US markets: Treasury, foreign exchange (FX) and money markets. Unlike traditional indices, the MCIs emphasize market microstructure dislocations as reflected in episodes of illiquidity and deviations from no-arbitrage conditions that reflect the balance sheet constraints of intermediaries and the impairment of arbitrage. Second, we employ tree-based artificial intelligence (AI) models (random forest (RF)) to forecast the full distribution of future market conditions via quantile regressions (Koenker and Bassett, 1978; Adrian et al., 2019). Our results show that ML models consistently outperform classical time-series approaches (e.g., autoregressive and multivariate quantile regressions), particularly at longer horizons (3–12 months). A well-recognized drawback of AI/ML models is lack of explainability, i.e. the challenge of understanding how complex models arrive at their output, and which input variables play a meaningful role in producing that output. We rely on Shapley values to address this issue (Shapley, 1953). This is critical from a policy perspective as it can inform which variables help explain shifts in the forecast distribution of MCIs. We find evidence that investor overextension (e.g., fund flows, the global financial cycle) and

¹ See Brunnermeier and Pedersen (2009), Brunnermeier and Oehmke (2009), Adrian and Shin (2010), Duffie (2020), Duffie (2021) and Ranaldo and Rossi (2017), among many others.

intermediary liquidity constraints (e.g., primary dealer security holdings) are important drivers of such distributional shifts.² Similarly, the MCI of other markets as well as past realizations of the same market’s MCI also play a relevant role, indicating the importance of spillovers in these integrated markets as well as the self-reinforcing nature of market stress.

Our MCIs are constructed through a two-step process.³ First, we curate market-specific variables that proxy for volatility (e.g., implied volatility in FX markets, the MOVE index for Treasuries), illiquidity (e.g., bid-ask spreads, on/off-the-run spreads, repo-OIS spreads), and arbitrage breakdowns (e.g., covered interest parity (CIP) deviations, discrepancies in the triangular no-arbitrage between currency triplets).⁴ These variables, standardized and aggregated via rolling-window principal component analysis (PCA), yield daily real-time indices normalized to zero mean and unit variance – i.e. positive (negative) values signal tighter(looser)-than-average market conditions. The MCIs capture well-known stress episodes across markets, e.g. the FX MCI spikes during the 2011–2012 European debt crisis and the 2016 Brexit referendum, while the Treasury MCI reflects deteriorating liquidity during the post-2013 taper tantrum and the 2020 pandemic. Critically, the MCIs provide complementary information to the VIX, as seen in 2015–2016 when FX stress surged without a corresponding equity volatility spike. This granularity addresses a key limitation of FSIs/FCIs, which often over-rely on equity-derived signals (Carlson et al., 2014; Kliesen et al., 2012).

Armed with the MCIs, we next set out to assess how best to forecast their realizations over horizons ranging from 3 to 12 months. We benchmark against two classical approaches: an autoregressive (AR) model and a multivariate quantile regression incorporating 44 predictors that may signal fragilities. We consider predictors in two broad categories, one reflecting investors’ risk perceptions and overextension, and the other capturing more structural features of markets such as the market-making capacity of intermediaries and funding liquidity conditions (Brunnermeier and Pedersen, 2009; Duffie et al., 2023; Gorton and Metrick, 2012). While the multivariate model outperforms the AR model in-sample, its out-of-sample accuracy deteriorates at longer horizons, suggesting overfitting. For example, at the 12-month horizon, the multivariate model’s quantile loss for the Treasury MCI exceeds the AR model’s by 18%. This underscores the limita-

² See, among others, Ben-Rephael et al. (2012), Ben-Rephael et al. (2021), Rey (2013), Miranda-Agrippino and Rey (2020), Du et al. (2023).

³ We make the MCI series available to researchers together with this paper.

⁴ See, among others, Amihud and Mendelson (1986) for an early analysis of bid-ask spreads as a measure of illiquidity, Du et al. (2018) on CIP deviations, Huang et al. (2025) violations of the law of one price in currency triplets, and Krishnamurthy (2002) on on/off-the-run spreads.

tions of linear models in capturing complex interactions – such as the feedback between dealer balance sheets and funding liquidity – that may drive market stress.

Tree-based models excel in this setting given their better ability to deal with rapidly shifting market conditions and non-linearities in the relationships. Random forest, in particular, with its ensemble of decorrelated decision trees, allows for non-linearities and interactions between predictors while at the same time being robust against overfitting (Breiman, 2001; Grinsztajn et al., 2022). For instance, for the 90th quantile of FX market stress (3-month horizon), RF achieves a 27% lower quantile loss than the AR model. In unreported robustness checks, we find that extreme gradient boosting (XGBoost) shows comparable performance, although it comes with higher computational costs.

Having established the superiority of ML models for forecasting the upper quantiles of the distribution of MCIs, we next assess which variables may drive this performance. The interpretability of ML models is often criticized, but Shapley values (Lundberg et al., 2018) mitigate this by quantifying each predictor’s marginal contribution—an important criterion for rendering the methods we develop useful for vulnerability monitoring and policymaking. For FX markets, the implied volatility of EUR and GBP derivatives, market sentiment as captured by fund flows, as well as past realizations of the FX and Treasury MCIs stand out. Similarly, intra-family fund flows into riskier segments (Ben-Rephael et al., 2021) and a measure of the global financial cycle (Rey, 2013) emerge as important predictors of future tail realization of money market stress. Investor overextension in seemingly calm and low-volatility periods thus often sows the seeds for subsequent market stress. These insights can help policymakers focus on key areas in the monitoring of non-bank intermediaries and leverage cycles, as flagged by recent Financial Stability Reports (e.g. Federal Reserve Board (2022)). The relevance of previous MCI realizations as well other markets’ MCIs highlight the self-reinforcing nature of illiquidity spirals (Brunnermeier and Pedersen, 2009; Plantin and Shin, 2018; Aymanns et al., 2017) and the interconnectedness of these markets, which while distinct, are very much intertwined in the modern market-based financial system.

Our framework has three policy applications. First, the MCIs serve as real-time diagnostic tools for central banks, flagging market-specific stress that aggregate indices miss. Second, the ML architecture offers a template for stress-testing frameworks, where forecasting the distribution of outcomes (e.g., left-tail liquidity shocks) is critical. Relatedly, the approach also offers a tool for policymakers to forecast emerging signs of market stress in real-time. Third, Shapley values identify systemic vulnerabilities – such as dealer balance-sheet constraints or crowded trades (Duffie et al., 2023; Brown et al., 2022)– that warrant macroprudential oversight. For academics, the results val-

idate ML’s role in financial economics, particularly in settings with sparse signals and non-linear dynamics.

Related literature. Our paper contributes to various strands of the literature. First and foremost, we contribute to the literature on machine learning applications in banking and finance (Fouliard et al., 2021). A growing literature studies ML in the context of asset pricing and return predictability (see for example Jensen et al. (2024) and Kelly et al. (2024) and references therein). Gu et al. (2020) demonstrate large economic gains to US equity investors using ML forecasts relative to leading regression-based strategies. In line with their paper and the findings of Grinsztajn et al. (2022), we demonstrate tree-based models’ superiority in forecasting financial variables. Our focus on market dysfunction and its implications for financial stability also connects our work with Aquilina et al. (2025), who show that recurrent neural networks can help predict violations of triangular arbitrage relations for the euro-yen-dollar triplet in foreign exchange markets over a 60-day horizon.⁵ Our use of Shapley values, which builds on Lundberg et al. (2018), helps bridge the gap between ML and interpretability.⁶ Recent work has also explored the ability of ML and related methods to predict banking and financial crises (Ward, 2017; Beutel et al., 2019; Bluwstein et al., 2020).⁷ We contribute by focusing on financial market stress prediction, zooming in on higher frequency dimensions as opposed to the slow-moving imbalances that characterizes the build-up towards banking crises.

Second, we contribute to the literature on financial stress measurement.⁸ While FSIs (Monin, 2017) and FCIs (Hatzius et al., 2010; Adrian et al., 2022) aggregate broad signals, our approach emphasizes disruptions to market liquidity, mispricing, and the breakdown of standard arbitrage relations that are the tell-tale sign of impaired market functioning. Moreover, we contribute by focusing on key markets that are interconnected by market-making and collateral to support leveraged strategies. This focus on identifying periods of market malfunctioning – which on several occasions over the past two decades required costly “dealer of last resort” interventions by the central bank (Aldasoro et al., 2025) – is what sets our work apart from that of others measuring broader

⁵ We use this currency triplet, plus others involving the British pound and the Swiss franc, as part of the set of input variables for the construction of our FX MCI.

⁶ Tarashev et al. (2016) provide an application of Shapley values to identify systemic banks.

⁷ Early work on banking crises prediction (without ML techniques) includes Borio and Lowe (2002) and Borio and Drehmann (2009). For more recent work see Schularick and Taylor (2012) and Greenwood et al. (2022), among others.

⁸ This literature grew significantly after the GFC, as did related but distinct work on early warning indicators of banking crises (see previous footnote) and systemic risk (Bisias et al., 2012; Acharya et al., 2016; Adrian and Brunnermeier, 2016; Hollo et al., 2012).

funding market conditions.

Finally, our work intersects two foundational strands of financial economics: market microstructure and its role in liquidity provision, and intermediary-based asset pricing with its focus on intermediary constraints as drivers of systemic risk.⁹ We bridge these by introducing high-frequency, market-specific indicators and leveraging ML to model their interactions. The market microstructure literature emphasizes how trading frictions, information asymmetries, and institutional design shape liquidity and price formation. Seminal work by Kyle (1985) and Glosten and Milgrom (1985) established bid-ask spreads as proxies for adverse selection and inventory costs, while more recent studies (e.g., Brunnermeier and Pedersen (2009)) link liquidity to funding conditions – a theme central to our MCIs. Recent empirical work highlights the role of arbitrage breakdowns as signals of market stress (Du et al., 2018; Rime et al., 2022; Huang et al., 2025). Our MCIs operationalize the lessons from this literature by synthesizing microstructure variables (e.g., bid-ask spreads, cross-currency basis deviations) into daily stress indicators. This aligns with Bai et al. (2018) who highlight that market-specific liquidity metrics are critical for diagnosing systemic risk.

Roadmap. The rest of the paper is structured as follows. Section 2 presents the construction of the MCIs, reviews their properties, and discusses how they relate to alternative indicators of financial stress or financial conditions. Section 3 presents benchmark quantile forecasting models against which our random forest model is then assessed. Section 4 uses Shapley values to identify which predictors carry the most information for forecasting the tails of the MCI distribution across markets. Finally, Section 5 concludes.

2 Measuring market conditions

The importance of monitoring financial market conditions cannot be overstated. Episodes of market stress can severely impair the efficient allocation of capital and price formation, leading to broader economic disruptions. The development of robust measures to gauge market conditions is thus crucial for policymakers and market participants alike. This section introduces newly developed market conditions indicators for three key market segments: the US Treasury market, US money markets, and the FX market centered

⁹ Financial intermediaries – banks, dealers, hedge funds, and money market funds – play a pivotal role in liquidity provision, but their capacity to do so is procyclical (Adrian and Shin, 2010) and constrained by balance-sheet costs (He and Krishnamurthy, 2013; Du et al., 2023) and regulation (Scheicher and Schrimpf, 2022).

around the US dollar. The MCIs aim to capture market volatility, illiquidity, and deviations from standard no-arbitrage conditions. They provide a comprehensive view of market health and potential stress, covering periods of market dysfunction as well as times when markets remain functional but operate with historically low liquidity.

The construction of the MCIs involves a two-step process, focusing on one market at a time.¹⁰ In the first step, we collect variables capturing volatility, market and funding (il-)liquidity as well as impaired market-making more broadly, for each of the three markets. We aim to strike a balance between the coverage of different aspects of market conditions and availability of a reasonably lengthy daily time series. Guided by this trade-off, we start our analysis in 2003.

In the second step, we build market-specific composite indicators. We express all variables so that higher values indicate worse market conditions. Then we standardize them through a z-score transformation to put them on an equal footing, ensuring they have a mean of zero and a standard deviation of one. Afterwards, we build out-of-sample MCIs through a rolling-window principal component analysis (PCA). We do so in order to consider only past information and avoid look-ahead bias in our forecasting exercise. We start with a three-year initial estimation window, whereby the MCI for each market in the first three years is defined as the first principal component, i.e. the linear combination of the input series that captures most of their variability. Afterwards, we expand the window at a monthly frequency and redo the PCA each month with the expanded window. For ease of interpretation, we normalize the MCIs to have zero mean and unit standard deviation. Positive values of the MCIs therefore indicate tighter-than-average market conditions, signaling potential stress.

Table 1 lists the input variables used in the first step. For the FX market, we include the cross-currency basis, violations of the law of one price (VLOOP) in currency triplets, bid-ask spreads in various currency pairs and JPMorgan’s FX volatility index.¹¹ For the US Treasury market, we include various indicators. Market liquidity is assessed with time-to-quote (a measure of how long bond dealers take to respond to a request for quote in the TradeWeb system), quoted spreads for securities of various maturities or deviations of observed bond yields from a fitted smooth yield curve (Hu et al., 2013), among others. Price dislocation in turn is captured through interest rate swap spreads

¹⁰ The procedure builds on, and expands upon, a related policy article (Aldasoro et al., 2022).

¹¹ The cross-currency basis is the difference between the interest paid to borrow one currency by swapping it against another and the cost of directly borrowing this currency in the cash market. A non-zero value indicates a violation of covered interest parity (CIP), signaling market dislocation (Du et al., 2018; Rime et al., 2022). Similarly, VLOOP indicates impairments to arbitrage (Huang et al., 2025). Quoted spreads are a standard measure of market liquidity, whereas FX volatility indices in turn capture market uncertainty.

(measured as the spread between overnight index swaps (OIS) and Treasury securities at various maturities and expressed in absolute terms), the Treasury futures basis (Barth and Kahn, 2021) or the on-the-run liquidity premium (Krishnamurthy, 2002; Duffie et al., 2023), among others. Volatility is captured with the MOVE index. Finally, for US money markets we include the spreads between repo and OIS rates, between various high-quality financial commercial paper and OIS rates (Rime et al., 2022), the TED spread and the LIBOR-OIS spread.

Taken together, the input variables we use capture known episodes of poor market conditions over the past two decades. For example, FX market volatility and dislocation indicators signaled the European sovereign debt crisis in 2011–12, the 2015 de-pegging of the Swiss franc, Brexit and the US money market fund (MMF) reform in 2016. In the Treasury market, measures of market liquidity and volatility rose around the GFC, the taper tantrum, the Covid-19 pandemic, and the start of the war in Ukraine. Similar measures also play a role in money markets: repo-OIS spreads as well as CP-OIS spreads surged during the GFC, the US MMF reform, the VIX spike in early 2018, the September 2019 repo turmoil, and the pandemic.

These patterns are also well-reflected in our MCIs, which capture both broad-based and market-specific episodes. Figure 1 presents the estimate of the MCIs for the three markets, based on the PCA analysis and showing both the full-sample and the rolling-window estimates.¹² The MCIs for the three markets spiked (to differing degrees) during the GFC and the Covid-19 pandemic. But they also signaled deterioration in market conditions in other less extreme episodes. For example, the FX MCI inched up during the European sovereign debt crisis and again in 2015/2016 (Swiss franc de-peg, Brexit and US MMF reform). The Treasury MCI, in turn, moved slightly into positive territory after the taper tantrum in 2013, as well as during idiosyncratic flash events in 2014 and 2021. It has been on the rise in more recent times, in line with market commentary on deteriorating Treasury market liquidity (Scheicher and Schrimpf, 2022; Federal Reserve Board, 2022). Finally, the money market MCI exhibited moderate increases around the MMF reform and the 2018 VIX spike, although it admittedly remained largely subdued outside the GFC and the Covid-19 crisis.

Importantly, the MCIs provide complementary information to the VIX, thus offering distinct insights into market conditions (Aldasoro et al., 2022). While MCIs correlate positively with the VIX, they often capture nuances that the VIX alone does not re-

¹² The full-sample PCA utilizes the entire dataset to compute the principal components, which can introduce future information and potentially bias estimation. Our rolling estimation addresses this issue by using only past information.

flect. For instance, during the 2015-16 period, the FX MCI rose significantly without a commensurate increase in the VIX, highlighting market-specific stress not evident from the VIX. In contrast, most FSIs and FCIs are tightly linked to the VIX, as it is a key input in their construction. The strong correlation between these indices and the VIX means that they often reflect similar information. For example, the Office of Financial Research’s FSI and Goldman Sachs’ US FCI both show very high correlations with the VIX (despite the VIX being only one of many inputs). This indicates that they similarly capture broad market sentiment and volatility. However, this tight linkage can limit their ability to identify market-specific stress episodes. This is a gap that the MCIs effectively fill by focusing on market-specific variables and conditions, in addition to their focus on price dislocations, illiquidity and impairments to arbitrage.

3 Predicting quantiles of market conditions indices

Financial market stress is inherently asymmetric and heavy-tailed. Policymakers care not just about the average forecast of market conditions but about tail risks – the upper percentile stress scenarios that could trigger contagion. Classical approaches such as linear regression estimate conditional means, flattening these critical extremes. Quantile regression, by contrast, models the entire distribution,¹³ allowing us to answer questions like: What factors drive the worst 10% of outcomes in money markets?

Quantile forecasting with high-dimensional data (44 predictors in our case) poses two challenges. First, non-linearities: relationships between predictors may be multiplicative or threshold-dependent. Second, sparse signals: many features or predictors may be uninformative on average but critical in specific quantiles.

Tree-based ML models such as random forest and XGBoost are uniquely suited to this task. By recursively partitioning the data into regions with similar stress levels, they capture local interactions and variable importance without imposing linear assumptions.

In this section we evaluate the out-of-sample accuracy of two widely used time series models in predicting the quantiles of MCIs and contrast that with tree-based ML models. We proceed in three steps. First, we define the forecasting problem. Second, we present the two benchmarking classical models we consider: autoregressive and multivariate (which requires a discussion of the predictors used). We then present the tree-based models and assess their performance against the benchmark.

¹³ For the seminal work on quantile regressions, see Koenker and Bassett (1978). For a recent multivariate application to forecasting and stress-testing, see Chavleishvili and Manganelli (2024).

3.1 Defining the problem.

For both baseline models considered, we use the quantile loss to evaluate the distance between forecasted quantiles for market conditions and the quantiles observed out-of-sample.

In a classic linear model, the objective of the τ -th quantile regression is to find the vector β that minimizes

$$\frac{1}{N} \sum_{t=1}^T \rho_{\tau}(y_t - \beta \mathbf{x}_t^T), \quad (1)$$

where $\rho_{\tau}(\cdot)$ is the so-called check function, defined as

$$\rho_{\tau}(u) = \begin{cases} \tau u & \text{if } u > 0 \\ (\tau - 1)u & \text{if } u \leq 0 \end{cases} \quad \text{for any } u \in \mathfrak{R} \quad (2)$$

In equation (1), $\mathbf{x}_t^T = [x_t^1, \dots, x_t^k, \dots, x_t^p]$ is a vector of dimension $1 \times p$, where p is the number of predictors. $\beta = [\beta^1, \dots, \beta^k, \dots, \beta^p]^T$ is a $p \times 1$ vector that captures the marginal change in the τ -th quantile due to the marginal change in x_t^k . The solution that minimizes the objective function delivers the optimal values of β .

Although our MCIs are available at daily frequencies, we take their monthly average to allow for the introduction of low-frequency explanatory variables into the model later on. We estimate the various models and test the out-of-sample performance with an estimation-test window of 84 months. In other words, for a given month t , we run our estimates using monthly observations from $t - 83$ to t . We then forecast the h -period ahead quantile distribution of MCIs with the explanatory variables observed at t . For each forecast horizon $h = 1, 2, \dots, 12$, we calculate the quantile loss from predicted quantiles and compute the average loss across different testing windows. For the one-month forecast horizon, we have 153 estimation and testing windows, and for the 12-month forecast horizon, we have 142 estimation and testing windows.

3.2 Benchmark models

Autoregressive model. We start with the simplest time-series model, where we only consider the past values of the market conditions indices. Specifically, we use a one-period lag, and run the following regression to predict the τ_{th} quantile of the dependent variable y_t :

$$y_t^{\tau} \sim \beta y_{t-1} + \varepsilon_t. \quad (3)$$

Equation (3) is similar to the quantile version of an AR(1) regression. To improve the performance of the model, we also allow for the market conditions in the three key financial markets we study to be correlated with each other. Accordingly, we modify equation (3) as follows:

$$y_{i,t}^\tau \sim \beta_1 y_{tr,t-1} + \beta_2 y_{mm,t-1} + \beta_3 y_{fx,t-1} + \varepsilon_t, \quad (4)$$

where $y_{i,t}$ and $i \in \{tr, mm, fx\}$ refers to the Treasury, Money, or FX market conditions at time t .

Multivariate time series model. An alternative approach is to consider various additional predictors to inform the forecasting of market conditions:

$$y_{i,t}^\tau \sim \beta_1 y_{tr,t-1} + \beta_2 y_{mm,t-1} + \beta_3 y_{fx,t-1} + \sum_{k=1}^K \gamma_k x_{k,t-1} + \varepsilon_t, \quad (5)$$

where x_k with $k = 1, \dots, K$ are K additional predictors that could be useful for forecasting future market conditions.

To this end, we consider a wide range of financial and economic indicators, providing a comprehensive view of market dynamics and macroeconomic states that are potential predictors for market conditions. When selecting indicators that can help explain a rising risk of future stress, we focus on variables in the two categories discussed above: (i) those that point to investor overextension or that signal perceptions of higher risk in the market; and (ii) variables which are informative about liquidity conditions and market-making capacity. Our dataset includes up to 44 explanatory variables. Table 2 presents their summary statistics.

The first group of variables capture the direct impact from the US government and central bank on market liquidity. Predictors include Federal Reserve’s Treasury securities purchases, both in total amounts and as a 6-month moving average, which helps to capture the impact from central banks’ supply of liquidity. Additionally, we collect the level and change in US Treasury’s General Account, as well as the volume of failed Treasury deliveries by primary dealers, offering insights into market liquidity and stability.

The second group of predictors involve fund flow data. We consider both the previous month, and the 6-month moving average to allow for longer fund flow trends to have an impact on market conditions. Specifically, we use EPFR fund flow data for developed and emerging markets, covering bonds, equities, high-yield, and investment-grade securities.

Besides cross-country fund flows, we also utilize intra-family fund flows into money market funds, high-yield funds, and equity funds, compiled by the Investment Company Institute following Ben-Rephael et al. (2012) and Ben-Rephael et al. (2021). Fund flow data shed light on investor sentiment and risk preferences across asset classes and regions.

We also consider various macroeconomic and macro-financial indicators. These include the Citi Economic Surprise indices for the US and the global economy, which measure how actual economic data compares to expectations, as well as coupon holdings, the excess bond premium (Gilchrist and Zakrajšek (2012)), the global financial cycle index (Rey (2013)), and the margin requirement for 10-year Treasury futures contracts (to capture the balance-sheet cost and margins for intermediaries to make markets). Under this category, we also include borrowing costs, as captured by short-term interest rates, as well as medium- and long-run term spreads.

Lastly, we consider indicators that proxy for risk in financial markets. This group includes the broad dollar index, implied volatility for EUR, GBP and JPY, the Merrill Lynch Option Volatility Estimate (MOVE) index for interest rate risk, VIX for near-term risk in the US stock market, and the SKEW index, capturing perceived tail risk of the distribution of S&P 500 returns.

3.2.1 Performance of baseline models

Figure 2 and Figure 3 report the in-sample and out-of-sample quantile loss for different forecast horizons spanning from one to twelve months. Rows indicate markets (FX, money market, and Treasury), whereas columns indicate the median, 75th percentile and 90th percentile. The blue lines reflect the performance of the auto-regressive model (AR), whereas the orange lines capture the performance of the multivariate models with the full (Full) list of 44 input variables.

In-sample, the multivariate model delivers a consistently better performance across horizons. By including more explanatory variables, the multivariate model utilizes more information, and Figure 2 indicates that it has a lower quantile loss than the autoregressive model. The in-sample performance of the multivariate model is unambiguously better (lower loss) regardless of the quantiles the model is fitting or the forecast horizons considered.

Out of sample, however, the order reverses. Across horizons, markets and quantiles, Figure 3 shows that the multivariate model delivers higher average quantile losses than the AR model. Given our interest in out of sample forecasting of MCIs, below we focus on the better-performing AR model as a benchmark when assessing the performance of

tree-based models.

3.3 Tree-based models

Within ML models, tree-based variants remain state-of-the-art for tabular data (Grinsztajn et al., 2022). Tree-based models beat neural network models across a wide range of hyperparameter choices for tabular dataset when the loss function is not smooth and the dataset contains many uninformative features. The quantile loss function is indeed non-smooth and the predictor variables we consider may also contain uninformative features. Tree-based models therefore fit our purpose well. Since we are not interested in optimizing for performance across ML models, we focus on random forest for its time efficiency. Moreover, compared with neural network models, tree-based models like random forest have the advantage of allowing for interpretability of results which, as discussed above, is key for our application.

The random forest method, formally introduced by Breiman (2001), is a procedure that creates ensembles of independent decision trees for a classification or regression task. They operate by constructing multiple decision trees during training and generating average predictions of individual trees. The prediction error of a random forest is upper bounded by a function determined by the correlation between prediction errors from different trees, and their average value. By averaging multiple trees, the random forest method reduces the risk of overfitting, which is a common issue with individual decision trees.

Meinshausen (2006) argues that random forests are able to provide information not only about the conditional mean, but also about the full conditional distribution of a variable of interest. He proposed a generalization of random forests (‘quantile regression forests’) as a method for inferring conditional quantiles, and showed that it provides accurate estimates even in settings with high-dimensional predictor variables. We rely on this method when estimating and predicting quantiles of our MCIs.

Performance of RF. As discussed above, among the two baseline models, the autoregressive model has a better out-of-sample performance than the full-feature model. In Figure 4, we therefore compare the out-of-sample quantile loss of the random forest model with the one of the autoregressive model. Columns denote markets (Treasury, money market, FX), whereas rows indicate quantiles (median, 90th and 95th). Concretely, we present the difference between the quantile loss of the RF relative to the autoregressive model, along with 90% confidence intervals. Negative values signal better performance of the

RF model.

Our results suggest that the random forest method can significantly outperform the benchmark. In particular, RF has a significantly lower quantile loss than the baseline model in the money and FX markets. It does so both across horizons and quantiles. For the money market, the RF model performance is comparable to the autoregressive model at the one-month forecast horizon. However, starting from the two-month forecast horizon, the RF model significantly outperforms the autoregressive model. This suggests that the RF model effectively captures the most informative variables for predicting the money market conditions two months ahead.

In the FX market, the RF model’s ability to capture relevant variables extends over a longer horizon. Its performance relative to the autoregressive model continues to improve up to the 11-month forecast horizon. This indicates that the RF model successfully picks up variables that provide robust and sustained predictive capability for future FX market conditions.

For the US Treasury market, however, the random forest does not significantly outperform the autoregressive model for predicting market conditions. While the RF model slightly underperforms in terms of median predictions, its performance improves when considering the upper tail of the distribution, especially at shorter and intermediate horizons (3-5 months).

4 Which variables help predict market conditions?

In an era where machine learning (ML) models are increasingly criticized as “black boxes”, the interpretability of their predictions is paramount – particularly in economics, where policymakers and academics require not only accurate forecasts but also actionable insights into the underlying drivers. Shapley values, a concept rooted in cooperative game theory (Shapley, 1953), offer a rigorous framework for attributing model predictions to individual features, ensuring both fairness (by accounting for all possible feature interactions) and transparency. This method is uniquely suited to disentangling the complex, non-linear relationships that underpin financial market stress, such as the interplay between dealer balance sheets and investor leverage cycles.

Traditional approaches to interpretability, such as coefficient magnitudes in linear regressions or permutation importance scores, fall short in two key ways. First, they fail to capture context-dependent effects: a variable like the excess bond premium may matter only during periods of intermediary distress, a nuance linear models cannot encode. Second, they ignore feature interactions – for example, the compounding effect

of Fed balance sheet expansion and fund flows on Treasury market liquidity. Shapley values address these gaps by quantifying the marginal contribution of each predictor to a specific forecast, conditional on all possible combinations of other variables. In the context of machine learning models, Shapley values are used to explain the contribution of each feature to the prediction made by the model.

Given a feature set \mathbf{x} and a model f , the Shapley value for a feature i represents the average contribution of that feature to the prediction across all possible subsets of features. Specifically, the Shapley value ϕ_i for feature i is given by:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

where:

- N is the set of all features.
- S is a subset of features not containing i .
- $f(S)$ is the model prediction using the features in subset S .
- $|S|$ is the number of features in subset S .
- $|N|$ is the total number of features.

For tree-based models, such as decision trees, random forests, and gradient boosting machines, the calculation of Shapley values can be optimized using the structure of the trees. Lundberg et al. (2018) develop fast exact tree solutions for SHAP (SHapley Additive exPlanation) values, which are the unique consistent and locally accurate attribution values for tree models. In the appendix, we illustrate the key steps in computing SHAP values using a two-feature example.

Figure 5 shows a summary plot of Shapley values in predicting the 90th quantile of money market condition three months ahead. Rows display the input variables, sorted by their respective mean absolute Shapley values. The x-axis indicates the Shapley value of the specified input variable in an out-of-sample prediction instance (dots). The colors of dots range from blue (low values of the specific input variable) to red (high values).

A handful of variables stand out as predictors of future tail realizations of money market conditions, reflecting both funding liquidity and investor overextension. Perhaps not surprisingly, current money market condition are the most important feature. Lower current values of the money market MCI (blue dots) tend to predict a lower 90th percentile of money market condition in three months. The opposite occurs with higher

values of the money market MCI (red dots). This points to the strong self-reinforcing nature of stressed money market conditions. In turn, lower (higher) purchases of Treasury securities by the Fed, as well as low (high) values of the global financial cycle (Miranda-Agrippino and Rey, 2020) indicator, tend to be predictive of tighter (easier) money market conditions. In turn, lower values of 3-month Treasury bill issuance as well as US Treasury deposits with the Federal Reserve (i.e. the Treasury General Account) are linked with lower future tail realizations of the money market MCI. Finally, lower values of the slope of the yield curve (as captured by the difference between 10-year and 2-year rates) help predict tighter money market conditions in the future.

Figure 6 provides the same analysis for FX market conditions. The top ranking variable is the implied volatility of the 3-month EUR-USD exchange rate, capturing the degree of uncertainty attached by options markets to movements in the exchange rate.¹⁴ Higher (lower) values are predictive of higher (lower) FX MCI in the future. Similar to the money market MCI, higher values of the FX MCI also carry important information about future tail realizations of the the FX MCI. In addition, high values of the Treasury MCI also tend to increase the 90th percentile of the future distribution of FX MCI. Moreover, when market sentiment is low, as captured by low values in the net intra-family flows into equity funds (Ben-Rephael et al., 2012), the forecasted 90th percentile of FX MCIs moves up. Most of the top variables explaining the forecasted tail of the FX MCI reflect investor overextension and risk perceptions, rather than funding liquidity or the market-making capacity of intermediaries (Du et al., 2023). To be sure, some of the latter variables also matter, but they tend to rank lower.

Finally, Figure 7 presents the results for the Treasury market. Similar to the other two MCIs, high current values of the MCI predict future tail realizations of the indicator. Equity market sentiment also ranks highly, with lower current values predicting future Treasury market stress. Higher values of the broad dollar index point to future shifts in the 90th percentile of the Treasury market MCI, consistent with the financial channel of Bruno and Shin (2015). The results for the Treasury MCI should be interpreted with caution, however, since (as shown before) the out-of-sample performance is not significantly better than that of the autoregressive model.

5 Conclusion

This paper demonstrates that tree-based machine learning models, coupled with newly constructed market condition indicators, provide a powerful toolkit for predicting finan-

¹⁴ Similarly, the implied volatility of GBP-USD is important, ranking fifth.

cial market stress. By focusing on three pivotal US markets – Treasury, FX, and money markets – we show that MCIs capture dislocations in liquidity, volatility, and arbitrage conditions that traditional financial condition and stress indices often miss. Our quantile regression framework reveals that ML models, particularly random forests, outperform classical time-series approaches across forecast horizons. Shapley value analysis highlights the outsized role of funding liquidity, investor overextension, and global financial cycle dynamics in explaining higher forecasted tail realizations of market conditions. In addition, the market conditions indices themselves affect future realizations of the MCI in the same market (self-reinforcing dynamics *within* markets) and across markets (spillovers *across* markets).

These findings have important implications. For policymakers, the MCIs offer a real-time diagnostic tool to identify market dysfunction, complementing existing FSIs and FCIs. The prominence of variables like intra-family fund flows and the global financial cycle underscores the need to monitor non-bank intermediaries and investor leverage cycles. For academics, our results validate the efficacy of tree-based ML in financial forecasting, particularly in environments riddled with non-linearities and sparse signals – a conclusion echoing Grinsztajn et al. (2022).

Our work points to four avenues for future exploration. First, extending the MCIs to corporate bond and derivatives markets could enhance systemic risk monitoring. Second, hybrid models combining ML with structural frameworks (e.g., DSGE models with financial frictions) may improve interpretability. Third, incorporating alternative data – dealer inventories, text-based sentiment, or blockchain-derived liquidity metrics – could refine predictions. Finally, real-time implementation requires addressing computational bottlenecks, such as the latency of Shapley value calculations for high-dimensional models.

References

- Acharya, Viral V., Lasse H. Pedersen, Thomas Philippon, and Matthew Richardson.** 2016. “Measuring Systemic Risk.” *The Review of Financial Studies* 30 (1): 2–47.
- Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone.** 2019. “Vulnerable Growth.” *American Economic Review* 109 (4): 1263–1289.
- Adrian, Tobias, and Markus K. Brunnermeier.** 2016. “CoVaR.” *American Economic Review* 106 (7): 1705–41.
- Adrian, Tobias, Federico Grinberg, Nellie Liang, Manmohan Sheheryar, and Jason Yu.** 2022. “The term structure of growth-at-risk.” *American Economic Journal: Macroeconomics* 14 (283-323): .
- Adrian, Tobias, and Hyun Song Shin.** 2010. “Dealer leverage and liquidity.” *Journal of Financial Intermediation* 19 (3): 418–437.
- Aldasoro, Iñaki, Peter Hördahl, and Sonya Zhu.** 2022. “Under pressure: market conditions and stress.” *BIS Quarterly Review* (19): 31–45.
- Aldasoro, Iñaki, Wenqian Huang, and Nikola Tarashev.** 2025. “Central bank liquidity backstops, bank regulation and risk-taking by asset managers.” *Management Science* (forthcoming).
- Amihud, Yakov, and Haim Mendelson.** 1986. “Asset Pricing and the Bid-Ask Spread.” *Journal of Financial Economics* 17 (2): 223–249.
- Aquilina, Matteo, Douglas Araujo, Gaston Gelos, Taejin Park, and Fernando Perez-Cruz.** 2025. “Harnessing artificial intelligence for monitoring financial markets.” *BIS Working Paper* (forthcoming).
- Aymanns, Christoph, Co-Pierre Georg, and Benjamin Golub.** 2017. “Illiquidity spirals in Coupled Over-The-Counter Markets.” Working Papers on Finance 1810, University of St. Gallen, School of Finance, <https://ideas.repec.org/p/usg/sfwphi/201810.html>.
- Bai, Jennie, Arvind Krishnamurthy, and Charles-Henri Weymuller.** 2018. “Measuring Liquidity Mismatch in the Banking Sector.” *The Journal of Finance* 73 (1): 51–93.

- Barth, Daniel, and R. Jay Kahn.** 2021. “Hedge Funds and the Treasury Cash-Futures Disconnect.” OFR Working Paper 21-01, Office of Financial Research, Available at SSRN: <https://ssrn.com/abstract=3817544>.
- Ben-Rephael, Azi, Jaewon Choi, and Itay Goldstein.** 2021. “Mutual fund flows and fluctuations in credit and business cycles.” *Journal of Financial Economics* 139 84–108.
- Ben-Rephael, Azi, Shmuel Kandel, and Avi Wohl.** 2012. “Measuring investor sentiment with mutual fund flows.” *Journal of Financial Economics* 104 363–382.
- Beutel, Johannes, Sophia List, and Gregor von Schweinitz.** 2019. “Does machine learning help us predict banking crises?” *Journal of Financial Stability* 45 100693.
- Bisias, Dimitrios, Mark Flood, Andrew Lo, and Stavros Valavanis.** 2012. “A survey of systemic risk analytics.” *Annual Review of Financial Economics* 4 255–296.
- Bluwstein, Kristina, Marcus Buckmann, Andreas Joseph, Miao Kang, Sujit Kapadia, and Ozgur Simsek.** 2020. “Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach.” Working Paper 848, Bank of England.
- Borio, Claudio, and Mathias Drehmann.** 2009. “Assessing the risk of banking crises - revisited.” *BIS Quarterly Review* 29–46.
- Borio, Claudio, and Philip Lowe.** 2002. “Assessing the risk of banking crises.” BIS Working Papers 114, Bank for International Settlements.
- Breiman, Leo.** 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32.
- Brown, Gregory W, Philip Howard, and Christian T Lundblad.** 2022. “Crowded Trades and Tail Risk.” *The Review of Financial Studies* 35 (7): 3231–3271.
- Brunnermeier, Markus K, and Martin Oehmke.** 2009. *Bubbles, Liquidity, and the Macroeconomy*. National Bureau of Economic Research.
- Brunnermeier, Markus K, and Lasse Heje Pedersen.** 2009. “Market Liquidity and Funding Liquidity.” *Review of Financial Studies* 22 (6): 2201–2238.
- Bruno, Valentina, and Hyun Song Shin.** 2015. “Cross-Border Banking and Global Liquidity.” *The Review of Economic Studies* 82 (2): 535–564.

- Carlson, Mark, Kurt Lewis, and William Nelson.** 2014. “Using policy intervention to identify financial stress.” *International Journal of Finance & Economics* 19 59–72.
- Chavleishvili, Sul Khan, and Simone Manganelli.** 2024. “Forecasting and stress testing with quantile vector autoregression.” *Journal of Applied Econometrics* 39 (1): 66–85.
- Du, Wenxin, Benjamin Hébert, and Wenhao Li.** 2023. “Intermediary balance sheets and the treasury yield curve.” *Journal of Financial Economics* 150 (3): 103722.
- Du, Wenxin, Alexander Tepper, and Adrien Verdelhan.** 2018. “Deviations from covered interest parity.” *The Journal of Finance* 73 (3): 915–957.
- Duffie, Darrell.** 2020. “Still the World’s Safe Haven? Redesigning the U.S. Treasury Market After the COVID-19 Crisis.” *Hutchins Center Working Paper 62*.
- Duffie, Darrell.** 2021. “Resilience Redux in the U.S. Treasury Market.” *Brookings Institution*.
- Duffie, Darrell, Michael Fleming, Frank Keane, Claire Nelson, Or Shchar, and Peter Van Tassel.** 2023. “Dealer Capacity and U.S. Treasury Market Functionality.” Staff Reports 1070, Federal Reserve Bank of New York.
- Federal Reserve Board.** 2022. “Financial Stability Report.” May.
- Fouliard, Jeremy, Michael Howell, Hélène Rey, and Vania Stavrakeva.** 2021. “Answering the Queen: Machine Learning and Financial Crises.” NBER Working Paper 28302, National Bureau of Economic Research.
- Gilchrist, Simon, and Egon Zakrajšek.** 2012. “Credit Spreads and Business Cycle Fluctuations.” *American Economic Review* 102 (4): 1692–1720.
- Glosten, Lawrence R., and Paul R. Milgrom.** 1985. “Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders.” *Journal of Financial Economics* 14 (1): 71–100.
- Gorton, Gary, and Andrew Metrick.** 2012. “Securitized Banking and the Run on Repo.” *Journal of Financial Economics* 104 (3): 425–451.
- Greenwood, Robin, Samuel Hanson, Andrei Shleifer, and Jakob Sørensen.** 2022. “Predictable financial crises.” *The Journal of Finance* 77 863–921.

- Grinsztajn, Leo, Edouard Oyallon, and Gael Varoquaux.** 2022. “Why do tree-based models still outperform deep learning on typical tabular data?” In *Advances in Neural Information Processing Systems*, Volume 35. 507–520, Curran Associates, Inc..
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu.** 2020. “Empirical asset pricing via machine learning.” *The Review of Financial Studies* 33 (5): 2223–2273.
- Hatzius, Jan, Peter Hooper, Frederic Mishkin, Kermit Schoenholtz, and Mark Watson.** 2010. “Financial conditions indexes: A fresh look after the financial crisis.” *NBER Working Paper* 16150.
- He, Zhiguo, and Arvind Krishnamurthy.** 2013. “Intermediary Asset Pricing.” *American Economic Review* 103 (2): 732–770.
- Hollo, Daniel, Manfred Kremer, and Marco Lo Duca.** 2012. “CISS—A composite indicator of systemic stress in the financial system.” ECB Working Paper 1426, European Central Bank.
- Hu, Grace Xing, Jun Pan, and Jiang Wang.** 2013. “Noise as Information for Illiquidity.” *The Journal of Finance* 68 (6): 2341–2382. <https://doi.org/10.1111/jofi.12083>.
- Huang, Wenqian, Angelo Ranaldo, Andreas Schrimpf, and Fabricius Somyi.** 2025. “Constrained liquidity provision in currency market.” *Journal of Financial Economics*, forthcoming.
- Jensen, Theis Ingerslev, Bryan T. Kelly, Semyon Malamud, and Lasse Heje Pedersen.** 2024. “Machine Learning and the Implementable Efficient Frontier.” Research Paper 22-63, Swiss Finance Institute, Available at SSRN: <https://ssrn.com/abstract=4187217> or <http://dx.doi.org/10.2139/ssrn.4187217>.
- Kelly, Bryan, Semyon Malamud, and Kangying Zhou.** 2024. “The Virtue of Complexity in Return Prediction.” *The Journal of Finance* 79 (1): 459–503. <https://doi.org/10.1111/jofi.13298>.
- Kliesen, Kevin, Michael Owyang, and E Vermann.** 2012. “Disentangling diverse measures: a survey of financial stress.” *Federal Reserve Bank of St Louis Review* 94 (5): 369–397.
- Koenker, Roger, and Gilbert Bassett.** 1978. “Regression Quantiles.” *Econometrica* 46 (1): 33–50.

- Krishnamurthy, Arvind.** 2002. “The bond/old-bond spread.” *Journal of Financial Economics* 66 (2): 463–506, Limits on Arbitrage.
- Kyle, Albert S.** 1985. “Continuous Auctions and Insider Trading.” *Econometrica* 53 (6): 1315–1335.
- Lundberg, Scott M, Gabriel G Erion, and Su-In Lee.** 2018. “Consistent individualized feature attribution for tree ensembles.” *arXiv preprint arXiv:1802.03888*.
- Meinshausen, Nicolai.** 2006. “Quantile regression forests.” *Journal of Machine Learning Research* 7 (983-999): .
- Miranda-Agrippino, Silvia, and Hélène Rey.** 2020. “US monetary policy and the global financial cycle.” *The Review of Economic Studies* 87 (6): 2754–2776.
- Monin, Philip.** 2017. “The OFR Financial Stress Index.” *Office of Financial Research Working Papers* (17-04): .
- Plantin, Guillaume, and Hyun Song Shin.** 2018. “Exchange rates and monetary spillovers.” *Theoretical Economics* 13 (2): 637–666.
- Ranaldo, Angelo, and Enzo Rossi.** 2017. “Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums.” *Journal of Finance* 72 (3): 1119–1151.
- Rey, Hélène.** 2013. “Dilemma not trilemma: the global cycle and monetary policy independence.” *Proceedings - Economic Policy Symposium - Jackson Hole*.
- Rime, Dagfinn, Andreas Schrimpf, and Olav Syrstad.** 2022. “Covered interest parity arbitrage.” *The Review of Financial Studies*, forthcoming.
- Scheicher, Martin, and Andreas Schrimpf.** 2022. “Liquidity in bond markets – navigating in troubled waters.” *SUERF Policy Brief* (395): .
- Schularick, Moritz, and Alan M. Taylor.** 2012. “Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008.” *American Economic Review* 102 (2): 1029–1061.
- Shapley, Lloyd S.** 1953. “A Value for n-Person Games.” *Contributions to the Theory of Games* 2 307–317.

Tarashev, Nikola, Kostas Tsatsaronis, and Claudio Borio. 2016. “Risk attribution using the Shapley value: methodology and policy applications.” *Review of Finance* 20 (3): 1189–1213.

Ward, Felix. 2017. “Spotting the danger zone: Forecasting financial crises with classification tree ensembles and many predictors.” *Journal of Applied Econometrics* 32 359–378.

Figure 1: **Market conditions indices.** This graph shows the five-day moving average of market condition indices for the US Treasury, money and foreign exchange (FX) markets (upper, middle, and lower panels respectively). Blue lines are the original series in Aldasoro et al. (2022) using a full-sample principal component analysis (PCA). Red lines are the series constructed using PCA with a rolling estimation that starts with a three-year window and expands each time by one month. The sample period is from 01/01/2003 to 31/05/2024.

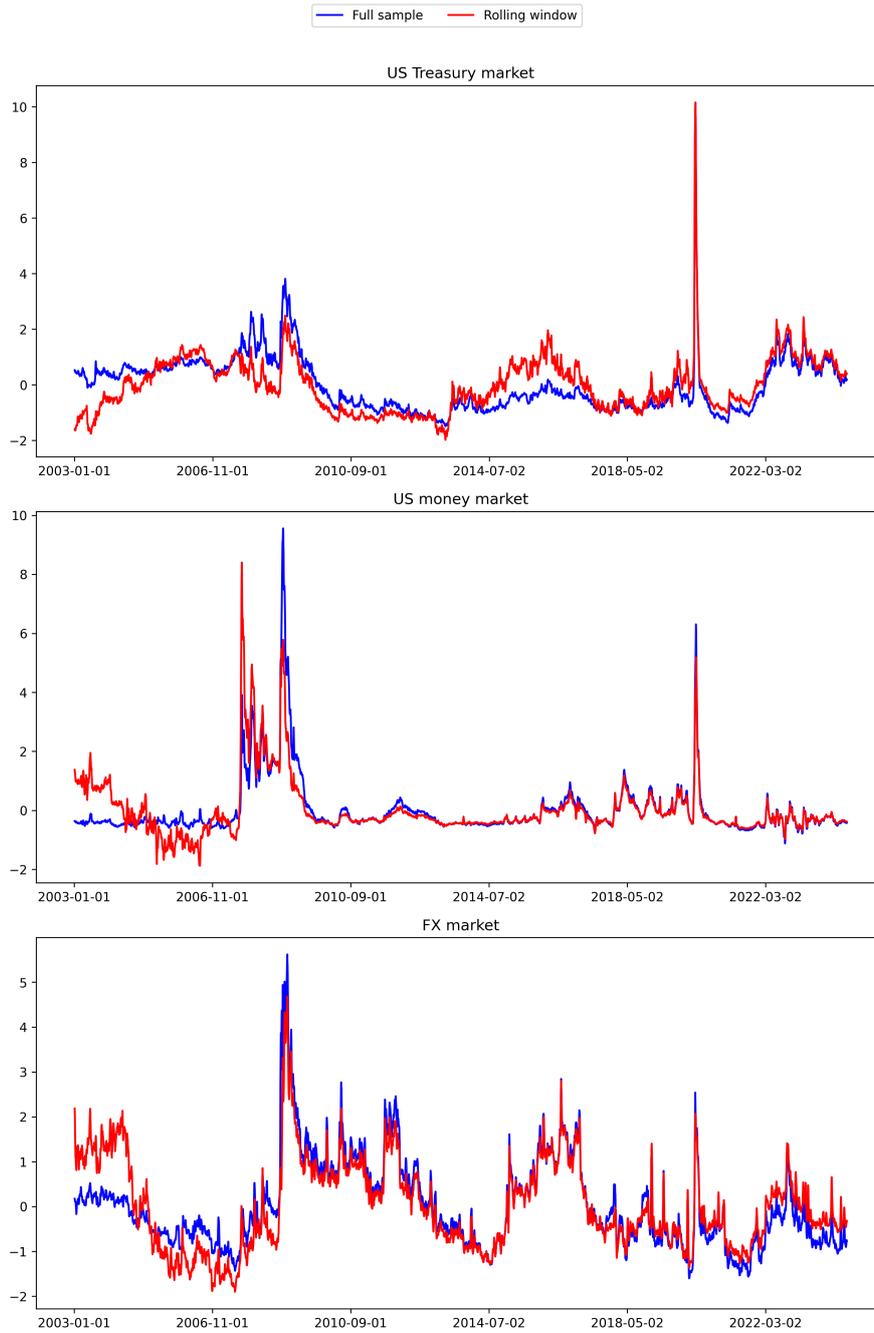


Figure 2: **In-sample quantile loss.** This graph shows the average loss of in-sample predictions by the 50-th (left panels), 75-th (central panels), and 90-th (right panels) quantile regressions. Blue lines (labeled as *AR*) display the average **in-sample** loss for different forecast horizons when the quantile regression models only include current values of market conditions (as in Equation (4)), and red lines (labeled as *Full*) indicate the value when the quantile regression models include a list of 44 explanatory variables (as in Equation (5)). The horizontal axis is the forecast horizon, ranging from one to twelve months. The sample period is from January 2003 to May 2024.

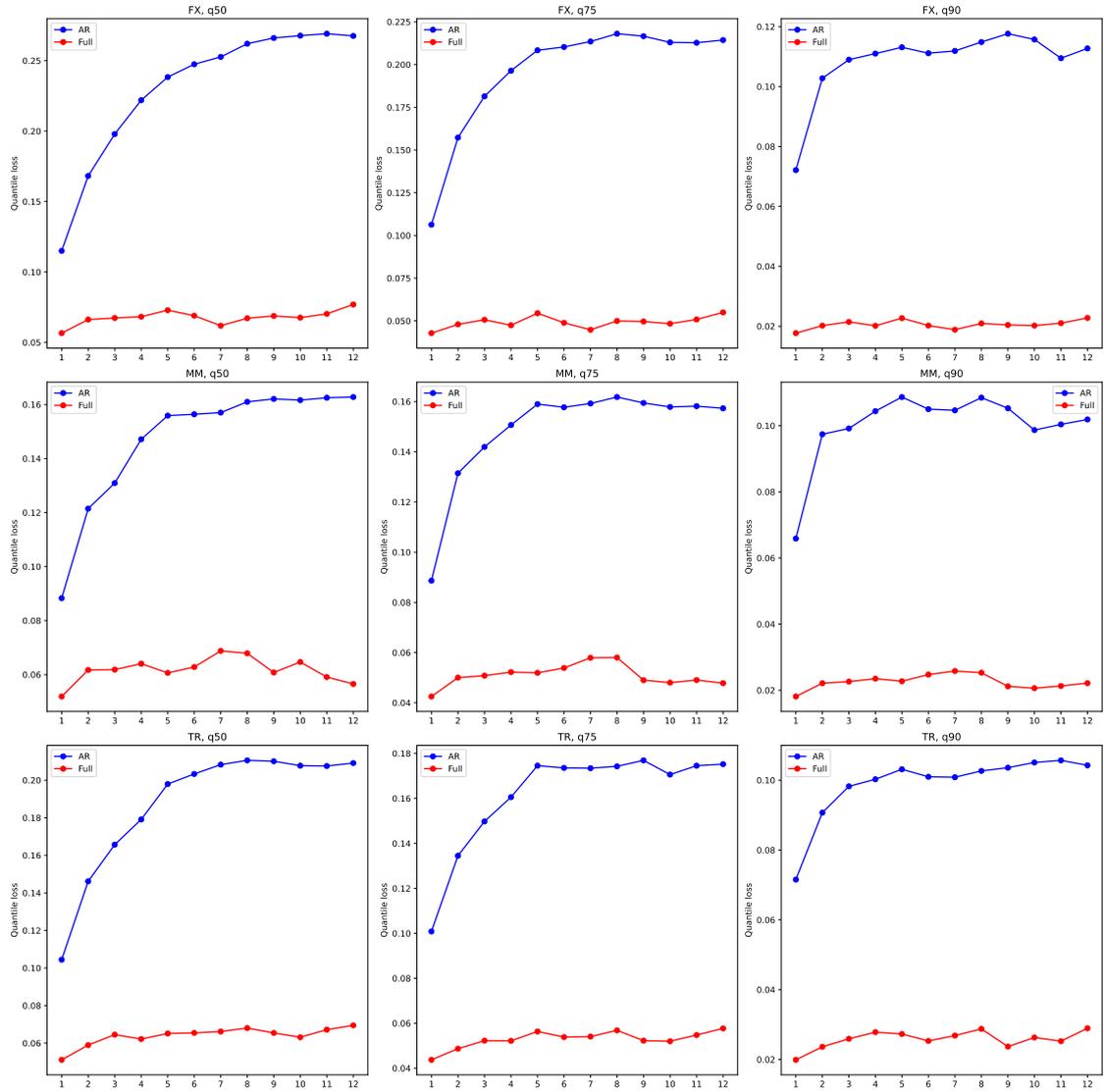


Figure 3: **Out-of-sample quantile loss.** This graph shows the average loss of in-sample predictions by the 50-th (left panels), 75-th (central panels), and 90-th (right panels) quantile regressions. Blue lines (labeled as *AR*) display the average **out-of-sample** loss for different forecast horizons when quantile regression models only include current values of market conditions (as in Equation (4)), and red lines (labeled as *Full*) display the average loss when quantile regression models include a list of 44 explanatory variables (as in Equation (5)). The horizontal axis is the forecast horizon, ranging from one to twelve months. The sample period is from January 2003 to May 2024.

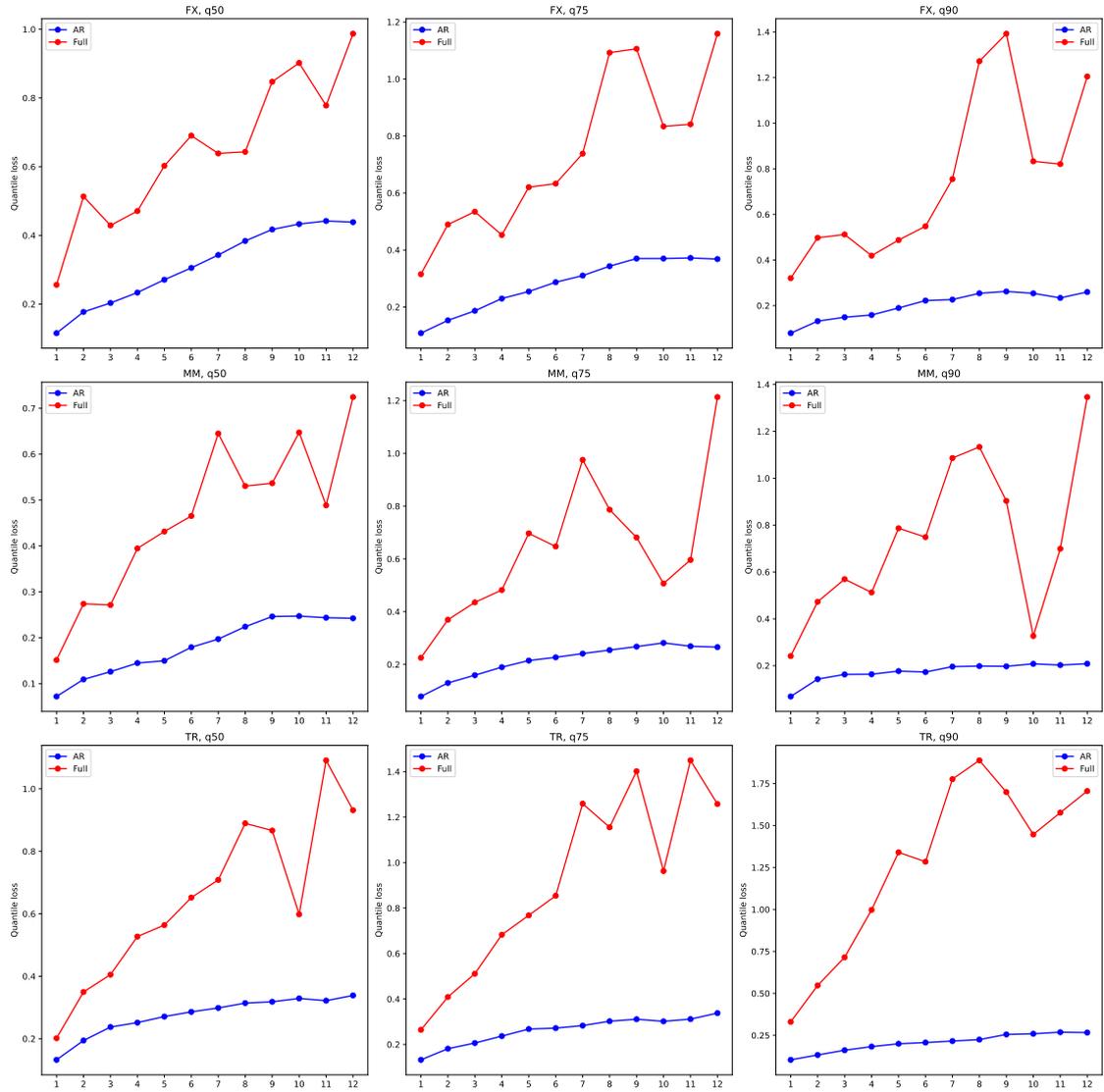


Figure 4: **Out-of-sample performance of random forest.** This graph shows differences in the quantile loss from the random forest model and the one from the autoregressive model. Quantile loss is calculated from our out-of-sample predictions for different forecast horizons. Negative values indicate that the random forest model has a better performance than the autoregressive model. Shaded areas are the 90% confidence intervals. The sample period is from January 2003 to May 2024.

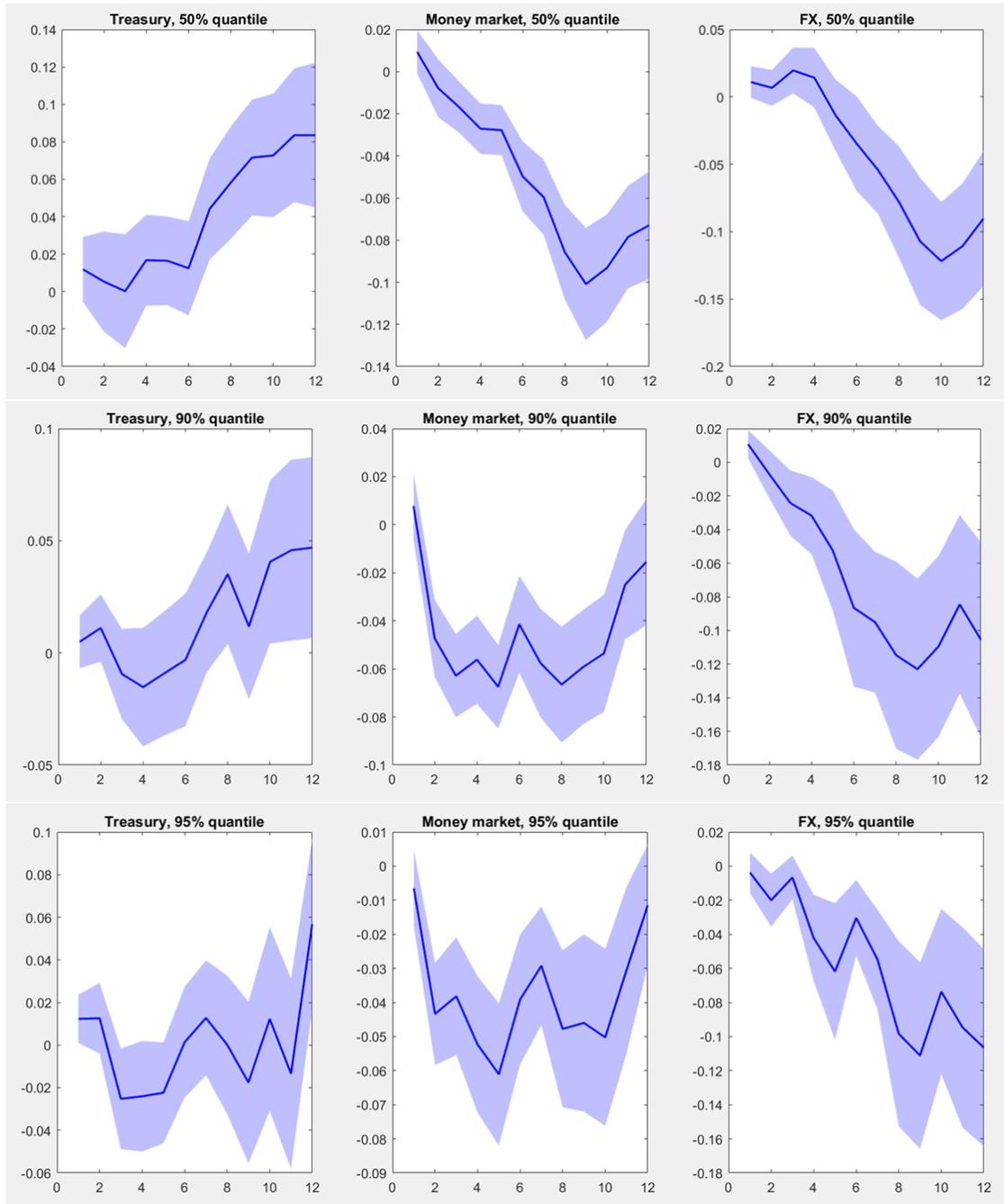


Figure 5: **Important features for predicting the money market MCI.** This graph displays the Shapley values derived from the random forest's out-of-sample predictions for the 90th quantile of the money market MCI, three months in advance. Each row shows Shapley values (indicated by the x-axis) of the specified explanatory variables in out-of-sample predictions. Rows are displayed in descending order based on the absolute Shapley values of the specified explanatory variable.

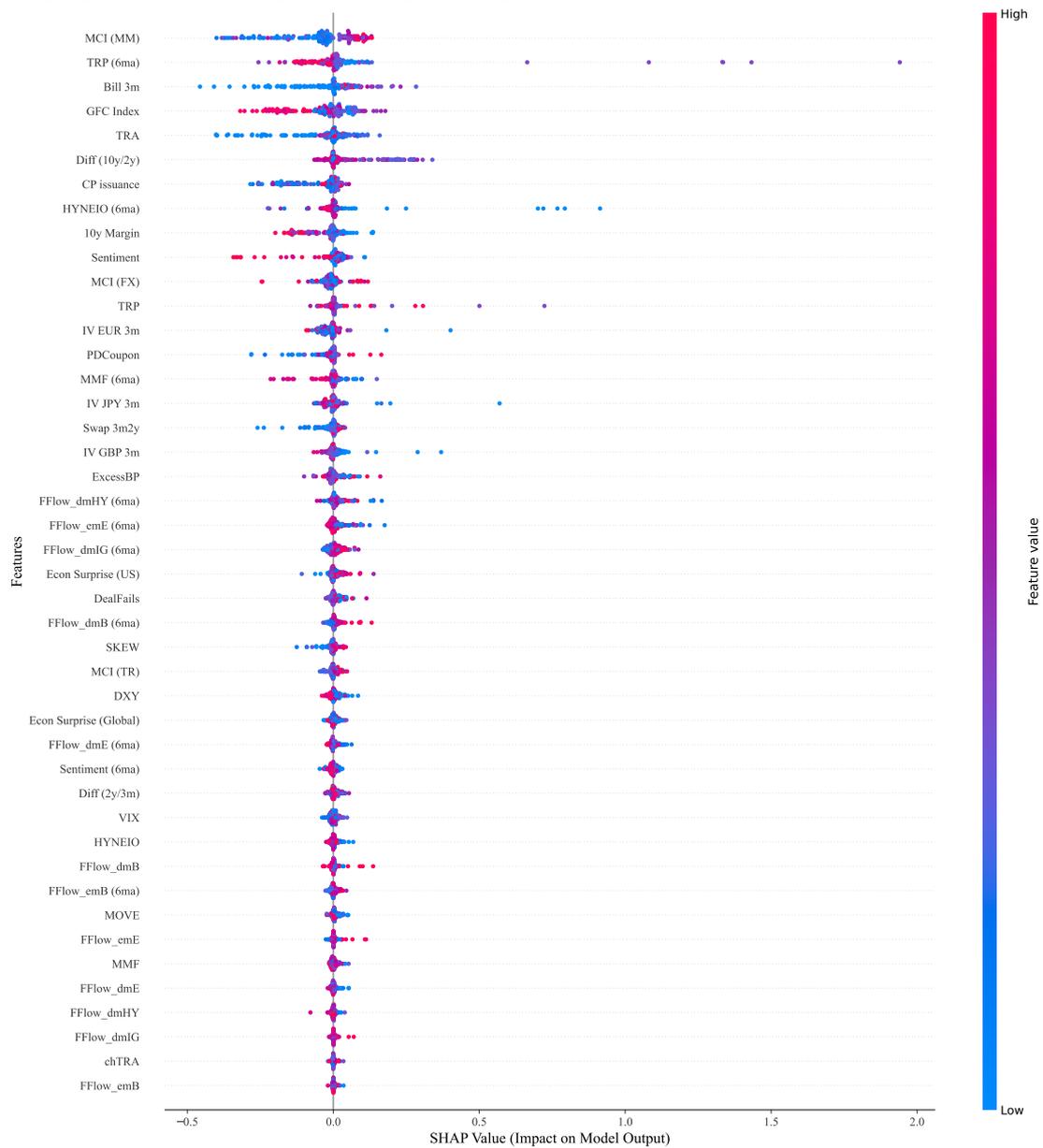


Figure 6: **Important features for predicting the FX MCI.** This graph displays the Shapley values derived from the random forest's out-of-sample predictions for the 90th quantile of the FX MCI, three months in advance. Each row shows the Shapley values (indicated by the x-axis) of the specified explanatory variables in out-of-sample predictions. Rows are displayed in descending order based on the average absolute Shapley values of the specified explanatory variable.

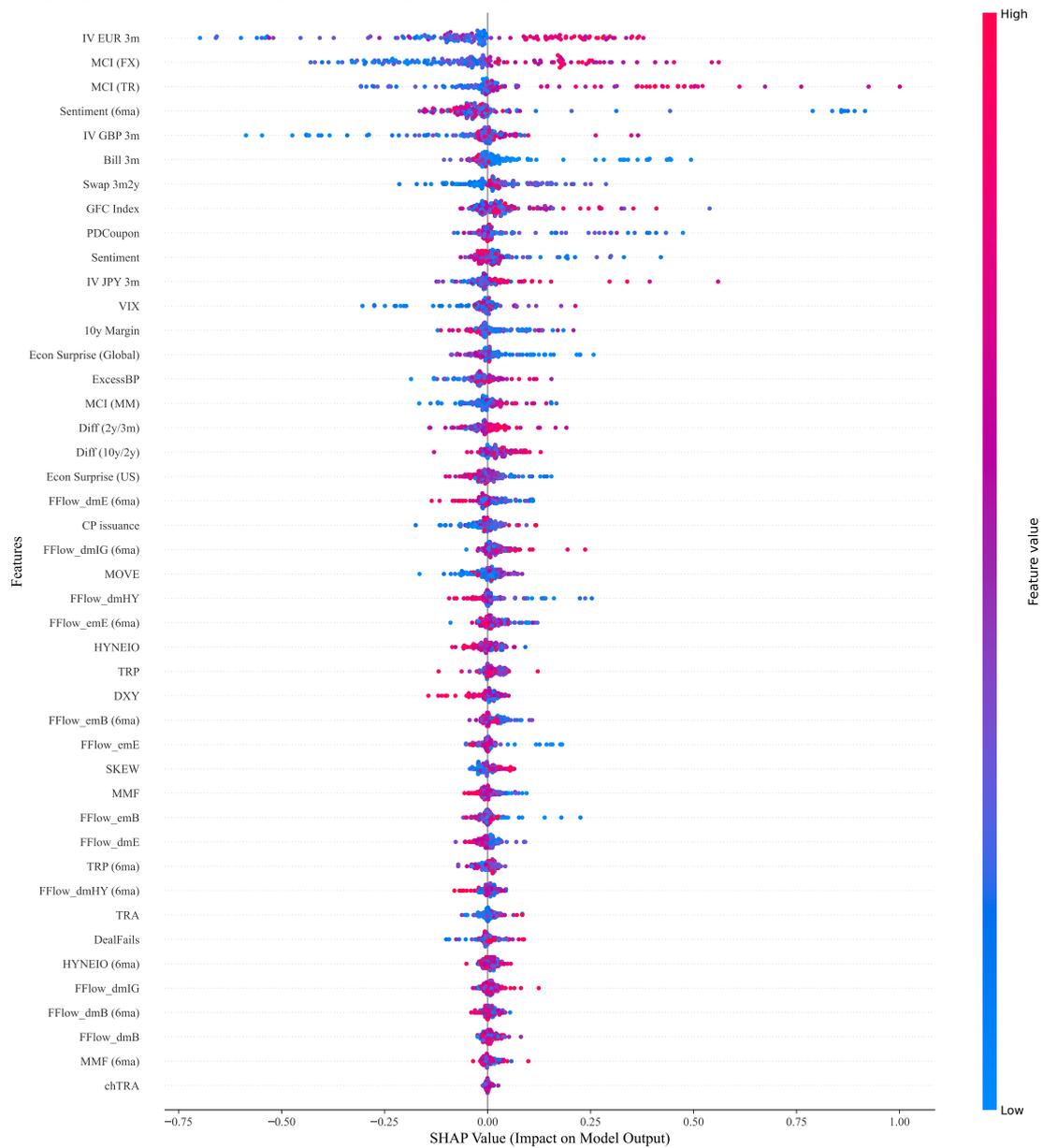


Figure 7: **Important features for predicting the Treasury market MCI.** This graph displays the Shapley values derived from the random forest's out-of-sample predictions for the 90th quantile of the Treasury MCI, three months in advance. Each row shows the Shapley values (indicated by the x-axis) of the specified explanatory variables in out-of-sample predictions. Rows are displayed in descending order based on the average absolute Shapley values of the specified explanatory variable.

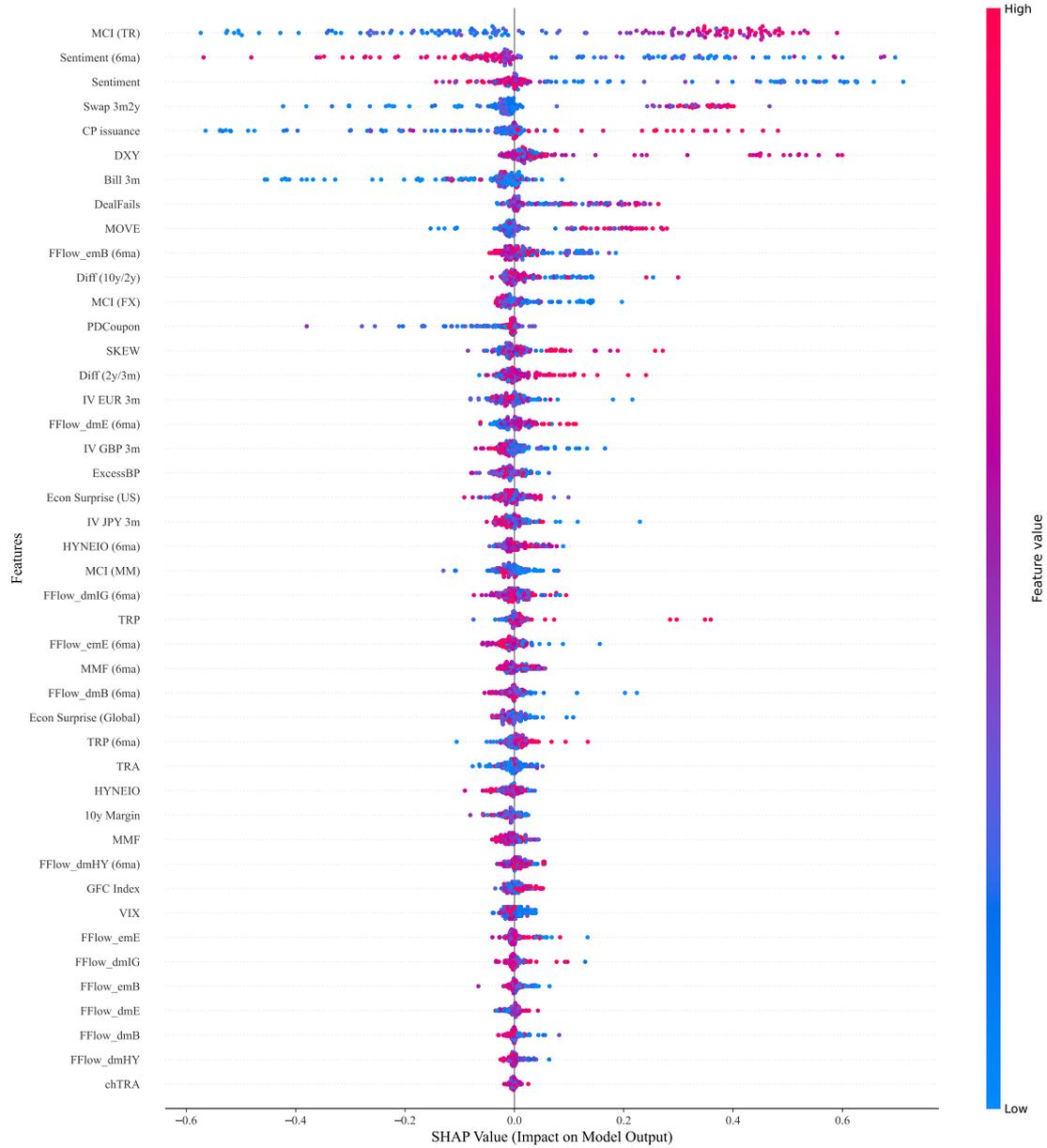


Table 1: **Input variables for MCI construction, by market.** Panels A to C list out the input variables for estimating the FX, Treasury and Money market MCIs. Our estimation covers the period from 01/01/2003 to 31/05/2024.

Panel A: FX market		
Variables	Definition	Source
JP Morgan FX volatility index	3-month option-implied volatility for G10 and emerging market currencies.	Bloomberg
Quoted spread	Quoted bid-ask spread of spot exchange rates (EUR, JPY, GBP, CHF vs USD).	Datascope
Cross-currency basis	EUR-JPY-GBP-CHF/USD cross-currency basis.	Bloomberg
Triangular VLOOP	Five-day moving average of triangular no-arbitrage deviations between USD-EUR-FX, USD-GBP-FX, with FX = EUR/GBP, CHF, JPY, as in Huang et al (2021).	Datascope
Panel B: Treasury market		
Variables	Definition	Source
MOVE Index	US Treasury yield volatility implied by 1-month options on 2, 5, 10 and 30-year Treasuries.	Bloomberg
Quoted spread (pre 2008)	Average spread of the 1st to 4th off-the-run US Treasuries (2, 5, 10 years).	Bloomberg
Quoted spread (post 2008)	Average spread of off-the-run US Treasuries; (maturities: 0–1.5, 1.5–3.5, 3.5–5.5, 5.5–10 years).	Tradeweb
Time-to-quote	Time taken by dealers to return their first quote for inquiries which ended up as a trade, for 5.5-to-10-year off-the-run US Treasuries.	Tradeweb
Liquidity Index (GVLQ-USD)	For US Government Securities. Average yield deviation relative to a fitted yield curve across US Treasuries with maturity beyond 1 year.	Bloomberg, Fed
On-the-run premium	Difference between par-coupon yields of seasoned 10-year Treasuries and yields on newly issued 10-year Treasuries, as in Christensen et al (2017).	Bloomberg
Absolute OIS spread	Absolute value of OIS-US Treasury spread (maturities: 6 months, 2 years, 5 years).	Bloomberg
Absolute Treasury futures basis	Absolute value of the implied minus actual 5-year trade repo rate (2-week moving average).	Bloomberg
Panel C: Money market		
Variables	Definition	Source
Commercial paper (CP) - OIS spread	3-month (A1/P1, Financial (Fin) AA) or 1-month (Non-Fin AA, or Non-Fin A2/P2) CP rates minus OIS.	Bloomberg
Repo - OIS spread	Absolute value of the 1-month or 1-week US GCF repo rates minus OIS.	Bloomberg
TED spread	3-month LIBOR minus Treasury-bill rate.	Bloomberg
LIBOR - OIS spread	3-month LIBOR minus 3-month OIS.	Bloomberg

Table 2: Summary statistics of predictors. **Panel A:** net monthly purchase of US government bonds by the Federal Reserve, in trillion dollars (TRP (tln)), as well as its six-month moving average (TRP (tln, 6ma)), monthly average and change in US Treasury’s deposits with Federal Reserve Banks (TRA (bln) or chTRA (bln)), primary dealer fails in US Treasury settlements (DealFails (bln)), and primary dealer net coupon treasury holdings (PDCoupon), as in Du et al. (2023). **Panel B:** monthly net fund flows (FFlow) from the EPFR to the developed/emerging bond (with suffixes dmB or emB) and equity (dmE or emE) markets, high-yield bond (dmHY), investment grade (dmIG), intra-family fund flows into money market funds (MMF), high-yield bond funds (HYNEIO), or equity funds (Sentiment), as in Ben-Rephael et al. (2012) and Ben-Rephael et al. (2021). **Panel C:** excess bond premium (ExcessBP) from Gilchrist and Zakrajšek (2012), the broad dollar index (DXY), the global financial cycle index (GFC Index) based on Rey (2013), the Citi economic surprise indices for the US or the global economy (Econ Surprise), the maintenance margin of 10-year treasury futures reported by CBOE in thousands of dollars (10y Margin), the dollar volume of total commercial paper issuance in billions (CP issuance), 3 month US treasury bills issuance (Bill 3m), and differences between 10-year and 2-year Treasury yields (10y/2y), and between 2-year and 3-month Treasury yields (2y/3m). **Panels D and E:** MOVE, VIX, SKEW indices, the three-month EUR/GBP/JPY implied volatility (IV EUR/GBP/JPY 3m), and current values of market liquidity conditions. All variables are aggregated into monthly frequency if the original data are available at a higher frequency. The sample period is from January 2003 to May 2024.

	count	mean	std	min	25%	50%	75%	max
Panel A: Fed/government related								
TRP (tln)	239.00	15.41	81.42	-101.23	-3.01	0.20	25.02	986.28
TRP (tln, 6ma)	239.00	16.05	57.89	-63.12	-2.31	1.78	29.47	312.78
TRA (bln)	239.00	259.87	366.18	3.86	26.48	107.24	345.19	1804.94
chTRA (bln)	239.00	3.01	99.66	-468.84	-35.59	-0.09	37.40	529.92
DealFails (bln)	239.00	218.46	342.60	28.86	96.39	163.02	245.43	4702.74
PDCoupon (tln)	239.00	28.55	95.90	-168.73	-24.65	48.00	92.69	208.50
Panel B: Fund Flows								
FFlow_dmB	239.00	0.37	0.74	-4.47	0.03	0.44	0.76	2.67
FFlow_dmE	239.00	-0.00	0.30	-1.36	-0.15	0.02	0.17	0.72
FFlow_dmHY	239.00	0.08	1.37	-6.09	-0.65	0.19	0.80	3.52
FFlow_dmIG	239.00	-0.03	2.18	-20.82	-0.14	0.49	0.84	2.14
FFlow_emB	239.00	0.46	2.11	-11.13	-0.70	0.40	1.65	6.71
FFlow_emE	239.00	0.30	1.09	-3.68	-0.32	0.31	0.83	4.43
FFlow_dmB (6ma)	239.00	0.36	0.50	-0.87	0.03	0.31	0.67	1.91
FFlow_dmE (6ma)	239.00	0.00	0.21	-0.71	-0.14	0.02	0.15	0.44
FFlow_dmHY (6ma)	239.00	0.05	0.78	-2.20	-0.49	0.07	0.53	2.01
FFlow_dmIG (6ma)	239.00	-0.10	1.64	-8.10	-0.14	0.35	0.78	1.57
FFlow_emB (6ma)	239.00	0.46	1.60	-4.57	-0.53	0.33	1.53	4.42
FFlow_emE (6ma)	239.00	0.29	0.68	-1.46	-0.17	0.29	0.62	2.48
MMF	239.00	-0.01	0.22	-0.89	-0.08	0.01	0.08	0.79
HYNEIO	239.00	-0.03	0.05	-0.35	-0.04	-0.02	-0.00	0.09
Sentiment	239.00	-0.01	0.10	-0.27	-0.09	-0.02	0.05	0.26
MMF (6ma)	239.00	-0.01	0.13	-0.34	-0.10	-0.01	0.05	0.47
HYNEIO (6ma)	239.00	-0.03	0.03	-0.20	-0.04	-0.02	-0.01	0.04
Sentiment (6ma)	239.00	-0.02	0.10	-0.24	-0.09	-0.02	0.05	0.22
Panel C: Macroeconomic conditions								
ExcessBP	239.00	-0.00	0.67	-0.96	-0.32	-0.14	0.05	3.50
DXY	239.00	89.21	9.34	71.80	81.13	89.52	96.62	112.12
GFC Index	239.00	0.04	1.23	-2.69	-0.94	-0.10	0.80	2.85
Econ Surprise (US)	239.00	5.76	47.60	-125.78	-21.55	5.61	33.22	236.48
Econ Surprise (Global)	239.00	6.10	28.28	-90.89	-10.22	5.16	21.50	104.27
10y Margin (k)	238.00	1.33	0.41	0.60	1.10	1.35	1.54	2.25
CP issuance (bln)	239.00	101.59	30.01	58.89	79.72	89.98	116.90	194.56
Bill 3m	239.00	1.53	1.84	0.00	0.08	0.32	2.42	5.48
Diff (10y/2y)	239.00	1.04	0.96	-0.96	0.22	1.05	1.73	2.84
Diff (2y/3m)	239.00	0.33	0.45	-1.07	0.08	0.37	0.62	1.77
Panel D: Volatility								
MOVE	239.00	84.52	30.96	41.87	61.41	77.58	98.83	221.94
VIX	239.00	19.10	8.39	10.13	13.54	16.79	22.18	62.67
SKEW	239.00	126.10	9.28	110.94	119.26	123.41	132.10	156.37
IV EUR 3m	239.00	8.97	2.98	4.56	6.83	8.47	10.47	21.64
IV GBP 3m	239.00	9.24	2.81	5.16	7.52	8.38	10.45	22.23
IV JPY 3m	239.00	9.64	2.72	5.32	7.72	9.18	11.35	20.45
Swap 3m2y	239.00	4.66	2.52	1.00	2.84	4.02	5.88	12.71
Panel E: MCIs								
MCI (TR)	239.00	0.05	0.96	-1.80	-0.71	0.02	0.72	6.34
MCI (MM)	239.00	-0.05	0.96	-1.53	-0.44	-0.30	-0.01	5.02
MCI (FX)	239.00	-0.09	0.93	-1.77	-0.77	-0.30	0.42	4.15

Appendix

A1 Example: TreeSHAP with Two Features

In this section, we review the key steps for computing features' Shapley (or SHAP) values using a tree with two features as an example.

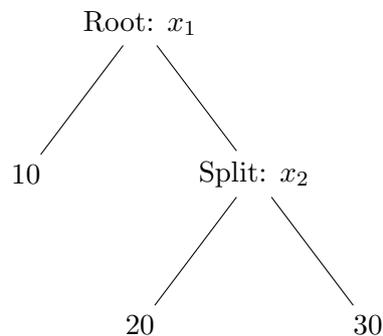
Tree Traversal. For each feature, the algorithm traverses the tree to determine the contribution of that feature at each node. This involves calculating the change in the model's prediction when a feature is included versus when it is excluded.

Path Contribution. For each path in the tree, the algorithm calculates the contribution of each feature along that path. This is done by considering the marginal contribution of the feature at each split in the tree.

Average Contribution. The Shapley value for each feature is then the average of its contributions across all possible paths and all trees in the ensemble (if using a forest or boosting model).

Consider a decision tree with the following structure:

- The root node splits on x_1 .
- The left child of the root node is a leaf node with a prediction of 10.
- The right child of the root node splits on x_2 .
- The left child of the node that splits on x_2 is a leaf node with a prediction of 20.
- The right child of the node that splits on x_2 is a leaf node with a prediction of 30.



To calculate the Shapley value for x_1 , we need to consider the contribution of x_1 to the prediction over all possible subsets of features.

Step 1: Calculate $f(S)$ and $f(S \cup \{x_1\})$

- $S = \emptyset$ (no features):

- Average prediction without any features:

$$\frac{10 + 20 + 30}{3} = 20$$

- $S = \{x_2\}$:

- Prediction using x_2 :

- * If x_2 is in the left child: 20

- * If x_2 is in the right child: 30

- Average prediction:

$$\frac{20 + 30}{2} = 25$$

- $S \cup \{x_1\} = \{x_1\}$:

- Prediction using x_1 :

- * If x_1 is in the left child: 10

- * If x_1 is in the right child:

- Further split on x_2 : average of 20 and 30:

$$\frac{20 + 30}{2} = 25$$

- Average prediction:

$$\frac{10 + 25}{2} = 17.5$$

Step 2: Calculate Contributions

- Contribution of x_1 when $S = \emptyset$:

$$f(S \cup \{x_1\}) - f(S) = 17.5 - 20 = -2.5$$

- Contribution of x_1 when $S = \{x_2\}$:

$$f(S \cup \{x_1\}) - f(S) = 17.5 - 25 = -7.5$$

Step 3: Average Contributions

- The Shapley value for x_1 is the average of its contributions across all subsets:

$$\phi_{x_1} = \frac{1}{2} ((-2.5) + (-7.5)) = \frac{1}{2} (-10) = -5$$

In this example, the Shapley value for x_1 is calculated by considering its marginal contributions across all possible subsets of features. The contributions are averaged to provide a fair representation of x_1 's impact on the model's predictions. This method ensures that each feature's contribution is fairly assessed, taking into account all possible interactions and dependencies in the model. The TreeSHAP algorithm efficiently computes these values by leveraging the tree structure, making it feasible to apply even for large and complex tree-based models.

Previous volumes in this series

1249 March 2025	The role of geopolitics in international trade	Han Qiu, Dora Xia and James Yetman
1248 March 2025	Climate Minsky Moments and Endogenous Financial Crises	Matthias Kaldorf and Matthias Rottner
1247 March 2025	Geopolitics meets monetary policy: decoding their impact on cross-border bank lending	Swapan-Kumar Pradhan, Viktors Stebunovs, Előd Takáts and Judit Temesvary
1246 March 2025	Monetary policy along the yield curve: can central banks affect long-term real rates?	Paul Beaudry, Paolo Cavallino and Tim Willems
1245 February 2025	Putting AI agents through their paces on general tasks	Fernando Perez-Cruz and Hyun Song Shin
1244 February 2025	Artificial intelligence and relationship lending	Leonardo Gambacorta, Fabiana Sabatini and Stefano Schiaffi
1243 February 2025	Mutual funds and climate news	Giulio Cornelli, Leonardo Gambacorta, Tommaso Oliviero and Koji Takahashi
1242 January 2025	Privacy-enhancing technologies for digital payments: mapping the landscape	Raphael Auer, Rainer Böhme, Jeremy Clark and Didem Demirag
1241 January 2025	Estimating non-linear heterogeneous agent models with neural networks	Hanno Kase, Leonardo Melosi and Matthias Rottner
1240 January 2025	The granular origins of inflation	Santiago Alvarez-Blaser, Raphael Auer, Sarah M. Lein and Andrei A. Levchenko
1239 January 2025	The use and disuse of FinTech credit: When buy-now-pay-later meets credit reporting	Yanfei Dong, Jiayin Hu, Yiping Huang, Han Qiu and Yingguang Zhang
1238 December 2024	Fiscal stimulus plans and households' expectations	Fiorella De Fiore, Marco Jacopo Lombardi and Albert Pierres Tejada
1237 December 2024	The macroeconomics of green transitions	Gregor Boehl, Flora Budianto and Előd Takáts
1236 December 2024	Savings-and-credit contracts	Bernardus van Doornik, Armando Gomes, David Schoenherr and Janis Skrastins

All volumes are available on our website www.bis.org.