Predicting Financial Market Stress with Machine Learning[1]

Iñaki Aldasoro
BIS

Peter Hördahl
BIS

Andreas Schrimpf
BIS & CEPR

Xingyu Sonya Zhu
BIS

December 11, 2025

Abstract

Using newly constructed market conditions indicators (MCIs) for three pivotal markets centered around the US dollar (Treasury, foreign exchange, and money markets), we show that tree-based machine learning (ML) models significantly outperform traditional time-series approaches in predicting the full distribution of future market stress. Through quantile regressions, we document that the random forest method achieves up to 27% lower quantile loss than autoregressive benchmarks, particularly at longer horizons (up to 12 months). Shapley value analysis reveals that variables related to macro expectations and uncertainty—especially about the monetary policy stance—are important predictors of future tail realizations of market conditions. For individual market segments, the state of the global financial cycle, as well as liquidity conditions, also play important roles. 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.

Keywords: machine learning, financial stress, quantile regressions, forecasting, Shapley value.
JEL Classification: G01, C53, G17, G12, G28.

[1]All authors are with the Bank for International Settlements (BIS), Monetary and Economic Department, Centralbahnplatz 2, CH-4002 Basel, Switzerland. Aldasoro: inaki.aldasoro@bis.org. Hördahl: peter.hoerdahl@bis.org. Schrimpf: andreas.schrimpf@bis.org. Zhu (corresponding author): sonya.zhu@bis.org. We would like to thank Jon Frost, Marco Lombardi, Fernando Pérez-Cruz, and workshop/seminar participants at the Barcelona School of Economics Summer Forum on Machine Learning in Economics, the Bank for International Settlements, the New Methods and Indicators in New Markets Workshop at the Bocconi University, the Hong Kong Monetary Authority, the Generative AI in Finance SI Paper Development Workshop, and the St. Gallen Financial Economics Workshop, for helpful comments and suggestions. We also thank Pietro Patelli for research assistance, and Doug Richardson for kindly sharing data from the Investment Company Institute. The views expressed in this paper are those of the authors and do not necessarily represent those of the BIS.

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 foreign exchange (FX) swaps – before propagating globally, as interconnected intermediaries and leveraged investors amplify shocks.[2]

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 ideally suited to machine learning (ML).

This paper makes three interrelated contributions. *First*, we construct novel market condition indicators (MCIs) for three pivotal US markets: Treasury, FX and money markets. Unlike traditional indices, the MCIs emphasize strains in market functioning and dislocations as reflected in episodes of heightened 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 such ML models consistently outperform classical time-series approaches (e.g., autoregressive and multivariate quantile regressions), particularly at longer horizons (up to 12 months).

Third, we suggest an approach that – despite the inherent black-box nature of ML compared to classical approaches – still retains a high degree of economic interpretability. 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 indicators related to investors' macro expectations and associated un-

[2] 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.

certainty surrounding such expectations are important drivers of distributional shifts. This is particularly so for variables related to monetary policy expectations and uncertainty, highlighting the outsized role of Fed policy expectations for market conditions. In addition, investor overextension (as captured by fund flows and the global financial cycle)[3] as well as liquidity conditions (Fed Treasury purchases and commercial paper (CP) issuance) also play important roles in some cases. Last but not least, the MCIs of other markets as well as past realizations of the same market's MCI are material for some market segments, indicating the importance of spillovers across integrated markets as well as the self-reinforcing nature of market stress.

Our MCIs are constructed through a two-step process.[4] 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- Overnight Index Swaps (OIS) spreads), and arbitrage breakdowns (e.g., covered interest parity (CIP) deviations, discrepancies in the triangular no-arbitrage between currency triplets).[5] 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 for example 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 1 to 12 months ahead. 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 limitations of linear models in capturing complex interactions – such as the feedback between dealer balance sheets and funding liquidity – that may drive market stress.

[3] See, among others, Ben-Rephael et al. (2012), Ben-Rephael et al. (2021), Rey (2013), Miranda-Agrippino and Rey (2020).

[4] We make the MCI series available to researchers together with this paper.

[5] 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) on violations of the law of one price in currency triplets, and Krishnamurthy (2002) on on/off-the-run spreads.

Tree-based models excel in this setting given their better ability to deal with rapidly shifting market conditions and non-linearities in 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 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 economic conditions and 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 state 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 and market regulators, flagging market-specific stress that aggregate indices may miss. Second, the ML architecture offers a template for stress-testing frameworks, where forecasting the distribution of outcomes (e.g., right-tail illiquidity 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 validate 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.[6] Our use of Shapley values, which builds on Lundberg et al. (2018), helps bridge the gap between ML and interpretability.[7] 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).[8] 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 the measurement of financial market conditions.[9] 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 of concern for policymakers (Markets Committee, 2022). Moreover, we contribute by concentrating on key markets interconnected through systemically important intermediaries, whose operations span multiple markets and provide access to leverage to other market participants. 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 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.[10] We bridge these by introducing high-frequency, market-specific indicators and leveraging ML to model their interactions. The

[6] 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.

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

[8] 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.

[9] 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). See Chavleishvili and Kremer (2025) for a recent application of real-time financial crisis measurement.

[10] 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).

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 6 concludes.

2 Measuring market conditions

The importance of monitoring financial market conditions cannot be overstated. Episodes of market stress can severely impair price formation and the efficient allocation of capital, 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 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 extreme periods of market dysfunction as well as times when markets remain functional but witness significant strains and low liquidity.

The construction of the MCIs involves a two-step process, focusing on one market at a time.[11] In the first step, we collect variables capturing volatility, market and funding (il-)liquidity as well as impaired intermediation 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

[11] The procedure builds on (Aldasoro et al., 2022).

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.[12] 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 (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 strained market conditions over the past two decades. For example, FX market volatility and dislocation indicators spiked amid 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 behaviors can also be observed 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

[12]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.

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.

Variables	Definition	Source
	Panel A: FX market	
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	
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 repo rate (2-week moving average)	Bloomberg
	Panel C: Money market	
Commercial paper-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

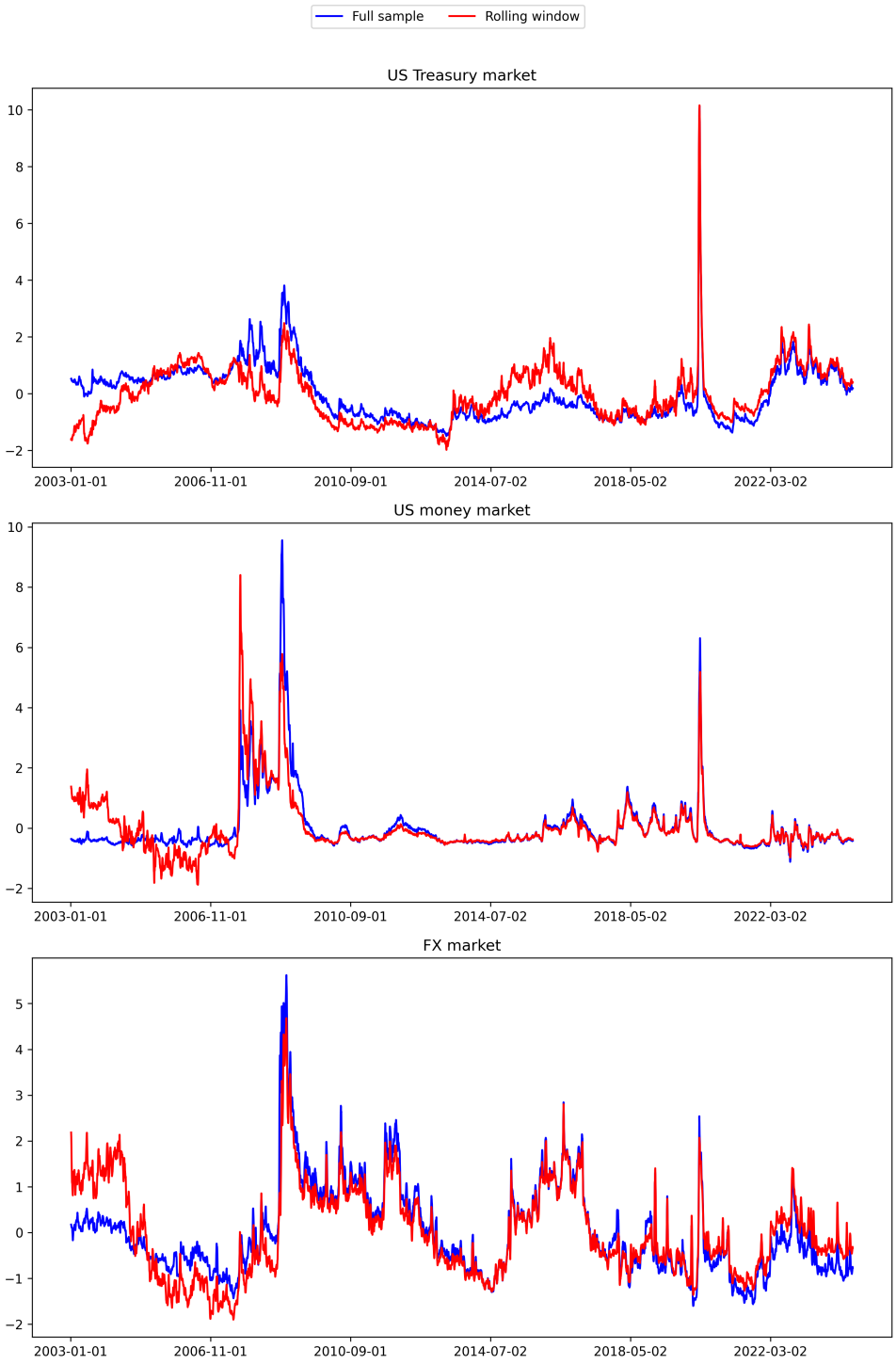
market-specific episodes. Figure 1 presents the estimate of the MCIs for the three markets, based on the PCA analysis and showing both the full-sample and the rolling-window estimates.[13] The MCIs for the three markets spiked (to differing degrees) during the GFC and the Covid-19 pandemic. But they 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 reflect. 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 the Chicago Fed's National Financial Conditions Index both show very high correlations with the VIX (despite the VIX being only one of many inputs for these indices). 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.

[13] 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.

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.



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 – that is, 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,[14] allowing us to answer questions like: What factors drive the worst 10% of outcomes in money markets?

Quantile forecasting with high-dimensional data, such as the 44 predictors in our case, presents two significant challenges. The first challenge is dealing with non-linearities. In such cases, the relationships between predictors may not be straightforward and can involve interactions, such as multiplicative effects or dependencies that only emerge beyond certain thresholds. The second challenge is the presence of sparse signals. While many predictors may seem uninformative on average, they can become critically important when focusing on specific quantiles.

Tree-based ML models such as random forest are uniquely suited to this task. By recursively partitioning input data into different subsets, 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: a simple autoregressive model and a multivariate model with multiple predictors (which requires a discussion of the predictors used). As a third step, we then present the tree-based models and assess their performance against the autoregressive benchmark, which outperforms the multivariate one out of sample.

3.1 Defining the problem.

Across all candidate models, we use the quantile loss to evaluate the distance between forecasted quantiles for market conditions and the quantiles observed out-of-sample.

The objective of the τ-th quantile regression is to find the best projection $f(\cdot)$ that minimizes the quantile loss

$$\min_{f(\cdot)} \frac{1}{N} \sum_{t=1}^{T} \rho_\tau(y_t - f(\boldsymbol{x_t})), \tag{1}$$

where $\rho_\tau(\cdot)$ is the so-called check function (a loss function that helps estimate the τ-th condi-

[14] 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). Chavleishvili and Kremer (2025) also use quantile regressions to show the predictive power of their Composite Index of Systemic Risk.

tional quantile of a dependent variable), defined as

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

In equation (1), $\boldsymbol{x_t} = \left[x_t^1, \cdots, x_t^k, \cdots, x_t^p\right]^T$ is a vector of dimension $p \times 1$, where p is the number of independent variables. The projection function $f(\boldsymbol{x_t})$ maps the values of independent variables observed at time t to the space of the dependent variable, in our case producing the desired forecast. The loss function (2) is asymmetric for $\tau \neq 0.5$: for high values of τ, underpredictions are penalized more than overpredictions, and vice versa for low values of τ.

In a linear model, the projection can be written as a linear function of dependent variables, $f(\boldsymbol{x_t}) = \boldsymbol{x_t^T \beta}$, with $\boldsymbol{\beta} = \left[\beta^1, \cdots, \beta^k, \cdots, \beta^p\right]^T$. β^k 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 $\boldsymbol{\beta}$. In a random forest model, $f(\cdot)$ is a tree-structured predictor that projects $\boldsymbol{x_t}$ to the space of the dependent variable.

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 candidate models and test their out-of-sample performance with the same estimation and testing windows. For a given month t, we run our estimates using monthly observations from $t - 84$ to $t - 1$. To evaluate the out-of-sample performance, we then forecast the h-period ahead ($h = 1, 2, ...12$) quantile distribution of MCIs with the explanatory variables observed at time t. For a given forecast horizon, we first calculate the quantile loss from predicted quantiles for each testing window. Then we compute the average value across different testing windows.

It is worth noting that we need to observe the realized value to calculate the quantile loss of out-of-sample predictions. Consequently, as the forecast horizon increases, the available number of data points that can be used for out-of-sample test decreases. Specifically, for a one-month forecast horizon, we have 153 non-overlapping estimation windows, each paired with a corresponding out-of-sample observation. In contrast, for a 12-month forecast horizon, the number of non-overlapping estimation windows and out-of-sample observations reduces to 142.

3.2 Benchmark models

Autoregressive model. We start with the simplest time-series model to predict future market conditions, where we only consider their recent observed values. Specifically, we use a one-period lag, and run the following regression to predict the τ-th quantile of the dependent variable y_t:

$$y_t^\tau = \boldsymbol{\beta} y_{t-1} + \varepsilon_t. \tag{3}$$

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 MCIs in the three key financial markets we study to influence each other to capture feedback effects. Accordingly, we modify equation (3) as follows:

$$y_{i,t}^{\tau} = \beta_1 y_{tr,t-1} + \beta_2 y_{mm,t-1} + \beta_3 y_{fx,t-1} + \varepsilon_t, \tag{4}$$

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

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

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

where x_k with $k = 1, \cdots, K$ 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 of 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 intermediation 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 through reserves. Additionally, we collect the level and change in US Treasury's General Account (the reserve account the Treasury holds at the Federal Reserve, which can influence money market conditions), 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 dealers' Treasury coupon holdings, the excess bond premium (Gilchrist and Zakrajšek (2012)), the global financial cycle index (Rey (2013)), and margin requirement for 10-year Treasury futures contracts (to capture the ease of synthetically using leverage, e.g. to arbitrage the cash futures basis (see, e.g. Schrimpf et al., 2020)). 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 in the distribution of S&P 500 returns.

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 (TGA (bln) or chTGA (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 (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
TGA (bln)	239.00	259.87	366.18	3.86	26.48	107.24	345.19	1804.94
chTGA (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 and financial indicators								
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

3.2.1 Performance of baseline models

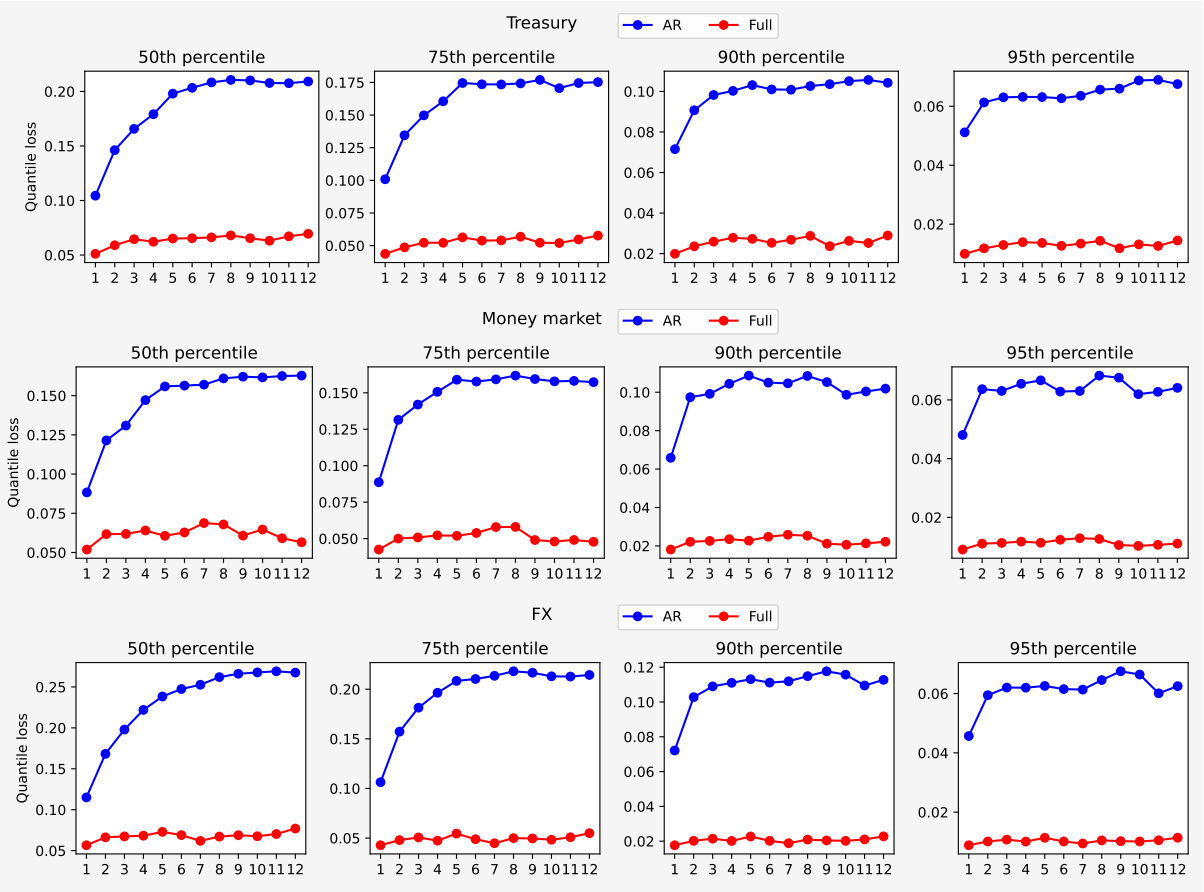
Figure 2 and Figure 3 report the in-sample and out-of-sample quantile loss for different forecast horizons spanning from one to twelve months. Rows indicate markets (FX, money market, and Treasury), whereas columns indicate the median, 75th percentile, 90th percentile and 95th percentile. The blue lines reflect the performance of the auto-regressive model (AR), whereas the red 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, as is common 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.

Figure 2: In-sample quantile loss

This graph shows the average loss of *in-sample* predictions by the 50th, 75th, 90th, and 95th quantile regressions. Blue lines (labeled as *AR*) display the average quantile loss for different forecast horizons when the the quantile regression model only includes current values of market conditions (as in Equation (4)), and red lines (labeled as *Full*) indicate the average loss when the regression model includes 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.

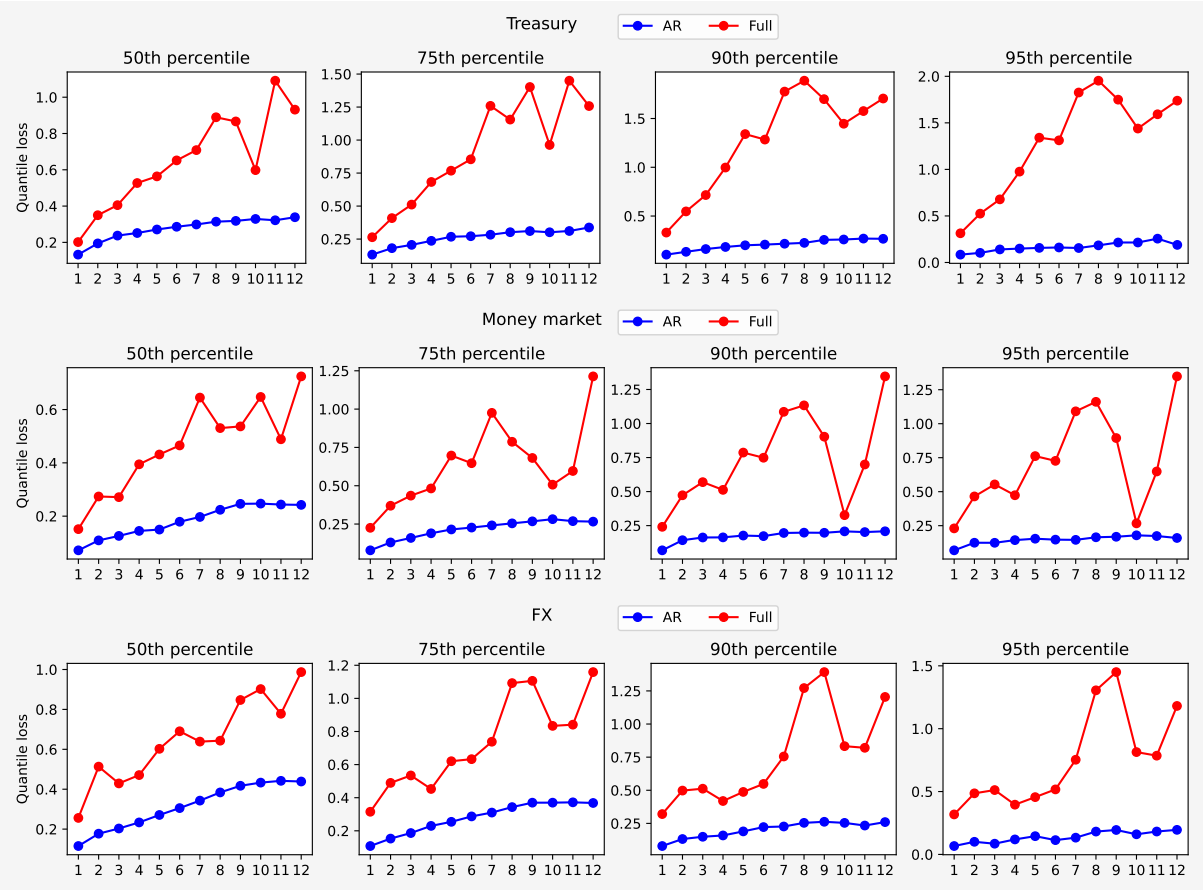


3.3 Tree-based models

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 datasets when the loss function is not smooth and the dataset contains many

Figure 3: Out-of-sample quantile loss

This graph shows the average loss of *out-of-sample* predictions by the 50th, 75th, 90th, and 95th quantile regressions. Blue lines (labeled as *AR*) display the average quantile loss for different forecast horizons when the quantile regression model only includes current values of market conditions (as in equation (4)), and red lines (labeled as *Full*) display the average loss when the regression model includes 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.



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 the random forest method given the relative computational ease of its implementation. Moreover, compared with neural network models, tree-based models like random forest have the advantage of allowing for easier interpretability of the 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 bounded from above by a function determined by the correlation between prediction errors from different trees and their average value. By averaging across multiple trees, the random forest method reduces the risk of overfitting, 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 proposes a generalization of random forests ('quantile regression forests') as a method for inferring conditional quantiles, and shows 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.[15]

Performance of random forests. Figure 4 compares the out-of-sample quantile loss of the random forest model with that of our benchmark autoregressive model. Columns represent 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 a better performance of the RF model.

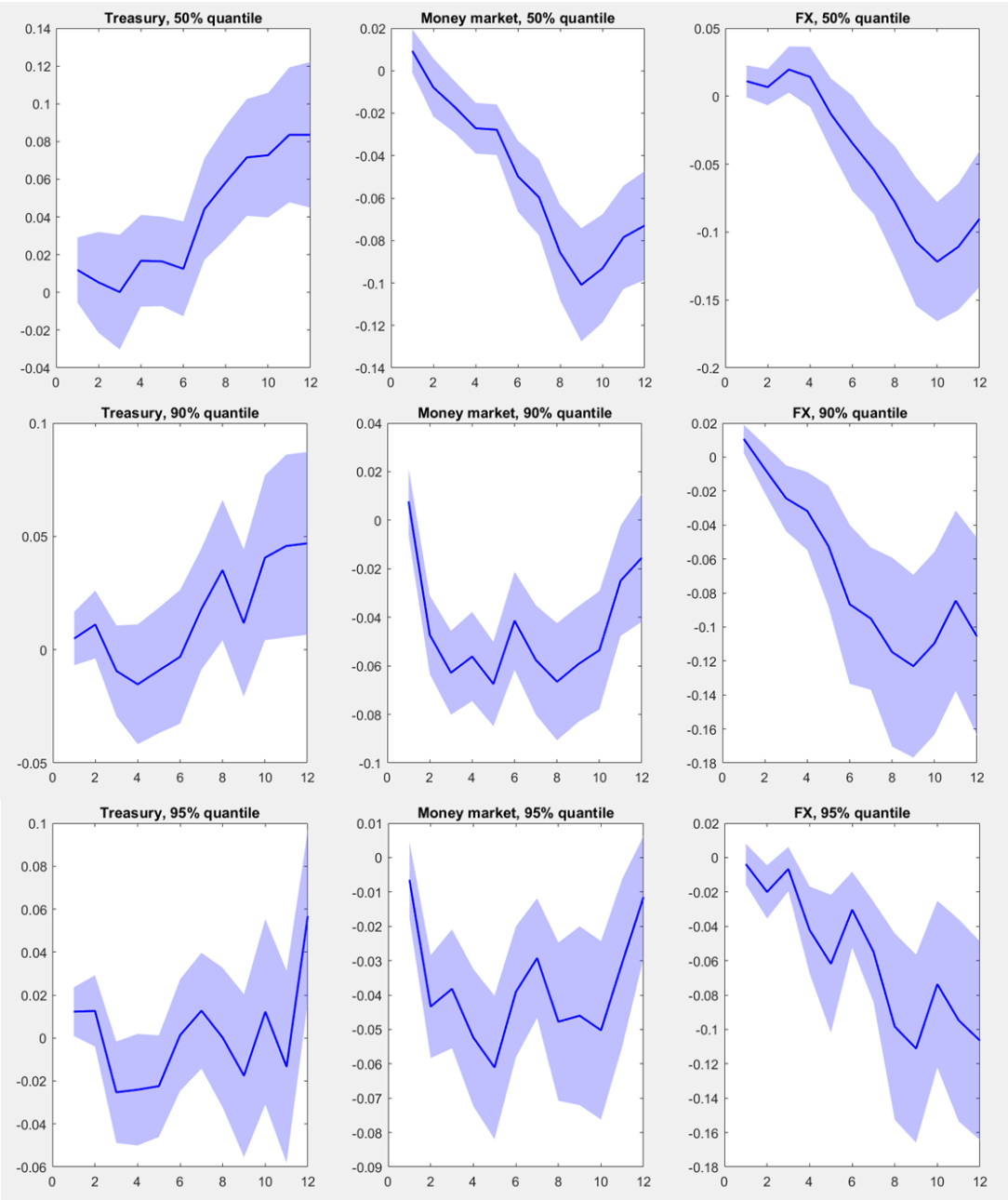
Our results suggest that the random forest method can significantly outperform the benchmark. In particular, RF has a significantly lower quantile loss than the benchmark model in the money and FX markets. This holds both across horizons and quantiles. For the money market, the RF model performance is 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.

In the FX market, the RF model's ability to capture relevant variables extends over a longer

[15]For each MCI, forecast horizon, and forecast window, we estimate the random forest model using 500 trees. Other hyperparameters are set to standard values, e.g.: minimum leaf size = 5; fraction of predictors sampled at each split = 1/3; split criterion = MSE; no pruning. Around 63% of observations (with replacement) are used for bootstrap sampling, with the remaining 37% used for out-of-bag validation (no cross-validation).

Figure 4: Out-of-sample performance of random forest

This graph compares quantile losses between the random forest and autoregressive models based on out-of-sample predictions across forecast horizons. Negative values indicate better performance of the random forest model. Shaded areas indicate 90% confidence intervals. The sample period is from January 2003 to May 2024.



horizon. Its performance relative to the autoregressive model continues to improve up to the 9-month or 10-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. The RF model slightly underperforms in terms of predicting the median, but its performance improves for predictions of the upper tail of the distribution, especially at shorter and intermediate horizons (3-5 months). While it is not clear why there appears to be less predictability for the Treasury market, one possible contributing factor could be decisive central bank interventions in this market at times of stress.

In Appendix A2, we also evaluate the performance of the neural network model. Consistent with Grinsztajn et al. (2022), our findings indicate that the neural network model does not necessarily outperform the tree-based model when applied to small-sized tabular datasets, which are commonly used in economic research.

4 Which variables help predict market conditions?

Given that machine learning models are often 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 underlying drivers. Shapley value analysis, a concept rooted in cooperative game theory (Shapley, 1953), offers a rigorous framework for attributing model predictions to individual features, ensuring both thoroughness (by accounting for all possible feature interactions) and transparency. This method is uniquely suited for 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 financial intermediary distress, a nuance linear models cannot capture. 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. In its traditional formulation, Shapley values explore the contribution of a player to the outcome of a game by sequentially picking one player at a time and evaluating *all* possible coalitions of the remaining players. In our context, we sequentially evaluate forecast accuracy for each explanatory variable under all possible 'coalitions' (i.e. combinations) of all other explanatory

variables.

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, in our context this is the full set of variables that can explain the predicted tail realizations of the MCIs.
- S is a subset of features not containing i, namely the set of explanatory variables excluding any given variable i.
- $f(S)$ is the model prediction using the features in subset S, i.e. the forecast of the tail realization of the MCI over a given horizon when using the explanatory variables collected in the set S. Accordingly, $f(S \cup \{i\} - f(S)$ assesses the marginal contribution of explanatory variable i.
- $|S|$ is the number of features (i.e. predictors) in subset S.
- $|N|$ is the total number of predictors.

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.

We illustrate the results from the Shapley analysis by calculating SHAP values for six-month ahead RF predictions of the 90th quantile for each of our MCIs. We choose the six-month horizon as 'representative' given that it is in the middle of our forecast horizon range. We focus on the 90th quantile since we are mainly interested in explaining upper tail predictions, i.e. for stressed conditions.

Figure 5 shows a summary plot of Shapley values in predicting the 90th quantile of money market (MM) conditions six months ahead. Rows display the input variables, sorted by importance as measured by their respective mean absolute Shapley values. The x-axis indicates the Shapley value of the specified input variable for each out-of-sample prediction (dots); i.e. the impact of the input variable for each forecasted 90th quantile. The colors of dots range from blue (low values of the specific input variable) to red (high values).

Figure 5: Important features for predicting the money market MCI

This graph displays the Shapley values derived from out-of-sample random forest predictions for the 90th quantile of the money market MCI, six months ahead. 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 of importance based on the mean absolute Shapley values of the explanatory variables.

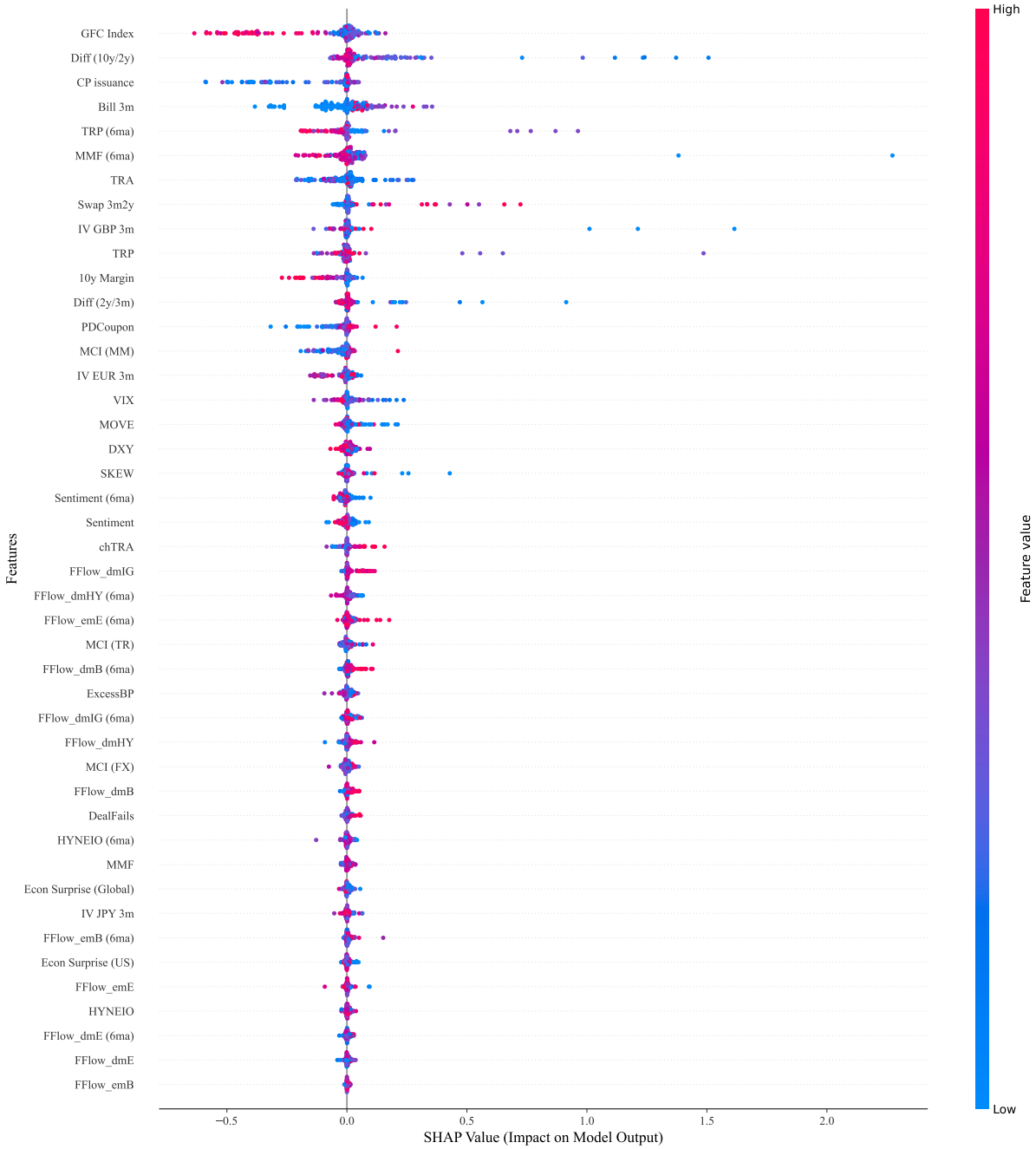
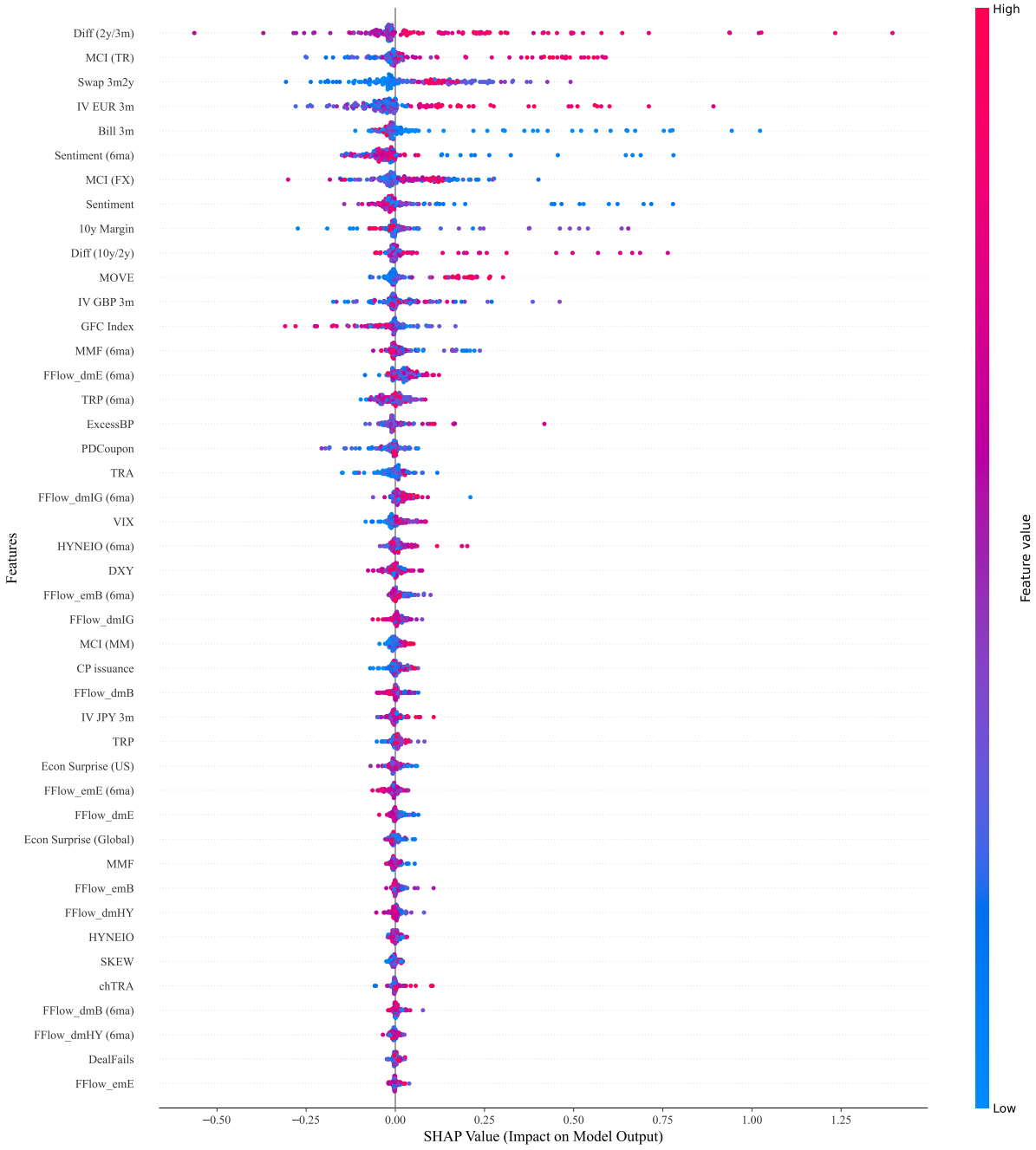


Figure 6: Important features for predicting the FX MCI

This graph displays the Shapley values derived from out-of-sample random forest predictions for the 90th quantile of the FX MCI, six months ahead. 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 of importance based on the mean absolute Shapley values of the explanatory variables.



A handful of variables stand out as predictors of future tail realizations of money market conditions, reflecting financial and macro economic conditions as well as market liquidity. The top feature is the global financial cycle (Miranda-Agrippino and Rey, 2020) indicator. High GFC values (red dots), which indicate booming financial conditions, tend to reduce the predicted 90th quantile of the MM MCI, i.e. signal a lower risk of MM stress six months ahead. The

second-most important feature is the 10-year / 2-year yield slope, a popular and usually reliable recession predictor. Low (blue) slope values, corresponding to an inverted yield curve and thus to greater recession risk, are typically associated with predictions of significantly higher values of the 90th quantile, indicating greater market stress risk. Another important variable is the 3-month Treasury bill rate, with low interest rate levels – i.e. easy Fed policy – lowering the risk of future money market stress. Two liquidity variables round out the top five important features. Low CP issuance tends to predict lower values of the 90th quantile, as the liquidity pressure in money markets is reduced from low competition for funds from the CP segment. Finally, higher Fed purchases of Treasuries signal a lower risk of money market pressures as liquidity is ample, whereas lower (possibly negative) Fed purchases tend to have the opposite effect.

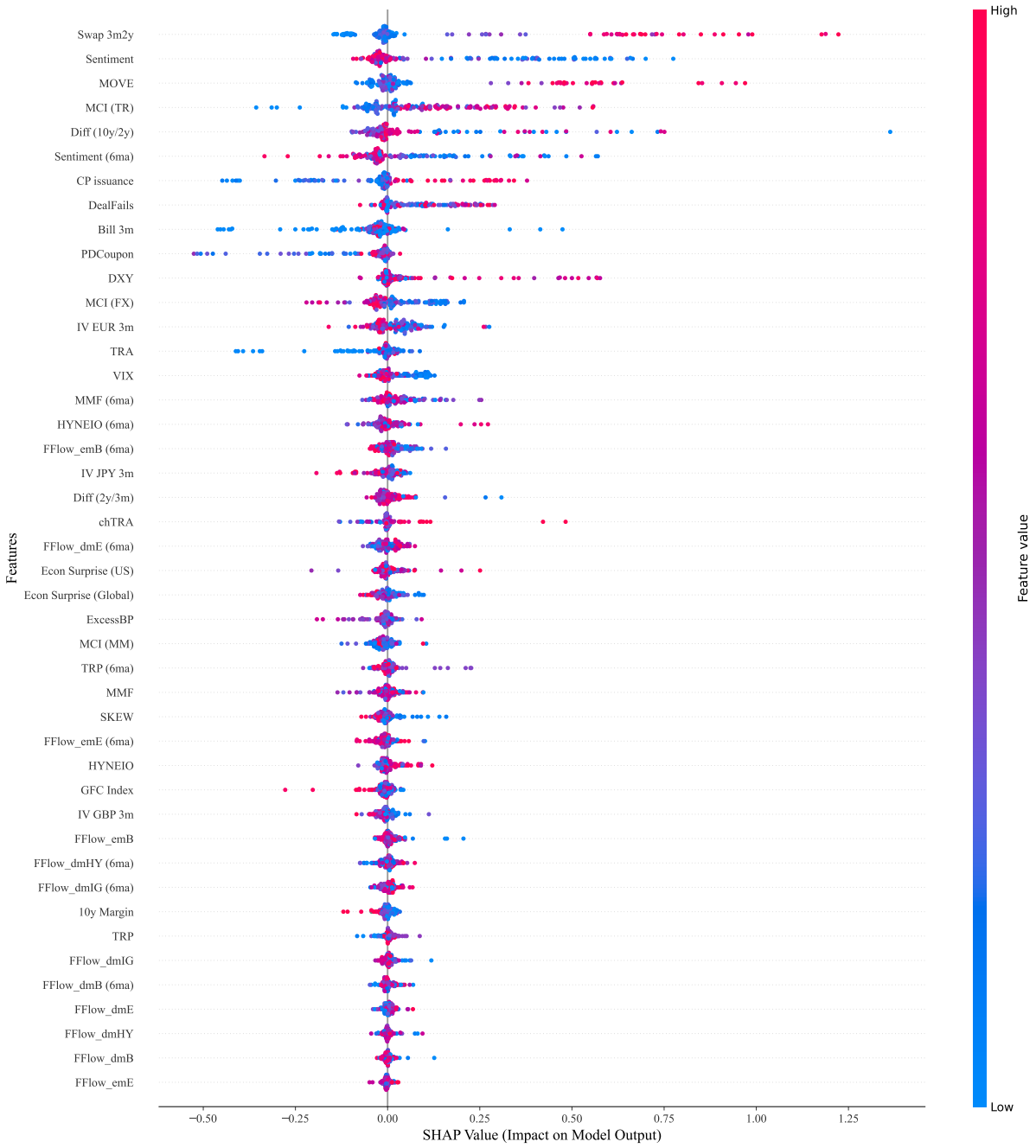
Figure 6 provides the same analysis for the 90th quantile of the FX MCI, six months ahead. As was the case for the MM segment, macro-related variables feature prominently, especially those related to monetary policy. The top ranking variable is a short yield slope, the difference between the 2-year yield and the 3-month rate, which is closely related to near-to-medium term Fed policy rate expectations. High (red) values of this variable tend to predominantly predict higher upper quantile values, i.e. tighter FX market conditions. In other words, when the Fed is expected to tighten policy, the model predicts a higher risk of upcoming FX market stress. In addition, low levels of the 3-month interest rate generally point to tighter FX market conditions too (this is the fifth most important variable). Uncertainty about future interest rates and thus about future policy rates also matters for FX market conditions: high (low) swaption volatility is predictive of high (low) FX MCI tail outcomes. The same holds for FX implied volatility, as captured by the 3-month EUR/USD implied volatility variable. These are the third and fourth most important variables for the FX market segment. The second-most important feature is the Treasury MCI, for which high values also tend to increase the 90th percentile of the future distribution of the FX MCI. This shows that spillovers of stress conditions across markets can also play an important role.

Finally, Figure 7 presents the results for the Treasury market. Similar to the other two market segments, macro and monetary policy variables are prominent. The highest ranking feature is the aforementioned swaption volatility, for which, again, high values point to tighter future market conditions. Long-term yield uncertainty, as captured by the MOVE volatility index has similar predictive characteristics for Treasury MCI tail outcomes. Similar to the money market, the long-term yield spread has important information for future Treasury market conditions. Interestingly, for this variable, both very high and very low values can, at times, point to a higher risk of stressed market conditions. This illustrates the non-linear features of RF quantile forecasting. A different type of variable enters as the second most important one: investor sentiment, as captured by intra-family flows into equity funds (Ben-Rephael et al., 2012). Low (blue colored) values indicate low risk-taking appetite among investors, which often results in

flows to safe assets such as Treasuries and thus less tight Treasury market conditions. Finally, lagged Treasury MCI values also matter (fourth most important variable), illustrating a type of self-reinforcing dynamic within this market segment.

Figure 7: Important features for predicting the Treasury market MCI

This graph displays the Shapley values derived from out-of-sample random forest predictions for the 90th quantile of the Treasury MCI, six months ahead. 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 of importance based on the mean absolute Shapley values of the explanatory variables.



5 Forecasting the full distribution of MCIs

Equipped with random forest estimates of future MCI quantiles, we can fit the entire distribution of future MCI values for a particular forecast horizon. This enables us to study if and how MCI distributions vary over time, as well as their sensitivity to predictors. Moreover, having the full distribution provides another way to evaluate the out-of-sample forecasting performance of the RF method.

In order to extract the predicted MCI distributions we follow Adrian et al. (2019). Specifically, for a given forecast horizon, we use the RF method to forecast the quantiles {0.05, 0.10, 0.15, ..., 0.95} for the specific MCI being examined. We then fit a skewed t-distribution to these data by minimizing the sum of squared deviations of the t-distribution's implied quantiles from the forecasted quantiles. The skewed t-distribution is sufficiently flexible to capture a broad range of distribution shapes; compared to a normal distribution, it allows for fat tails as well as asymmetry (Azzalini and Capitanio, 2003).

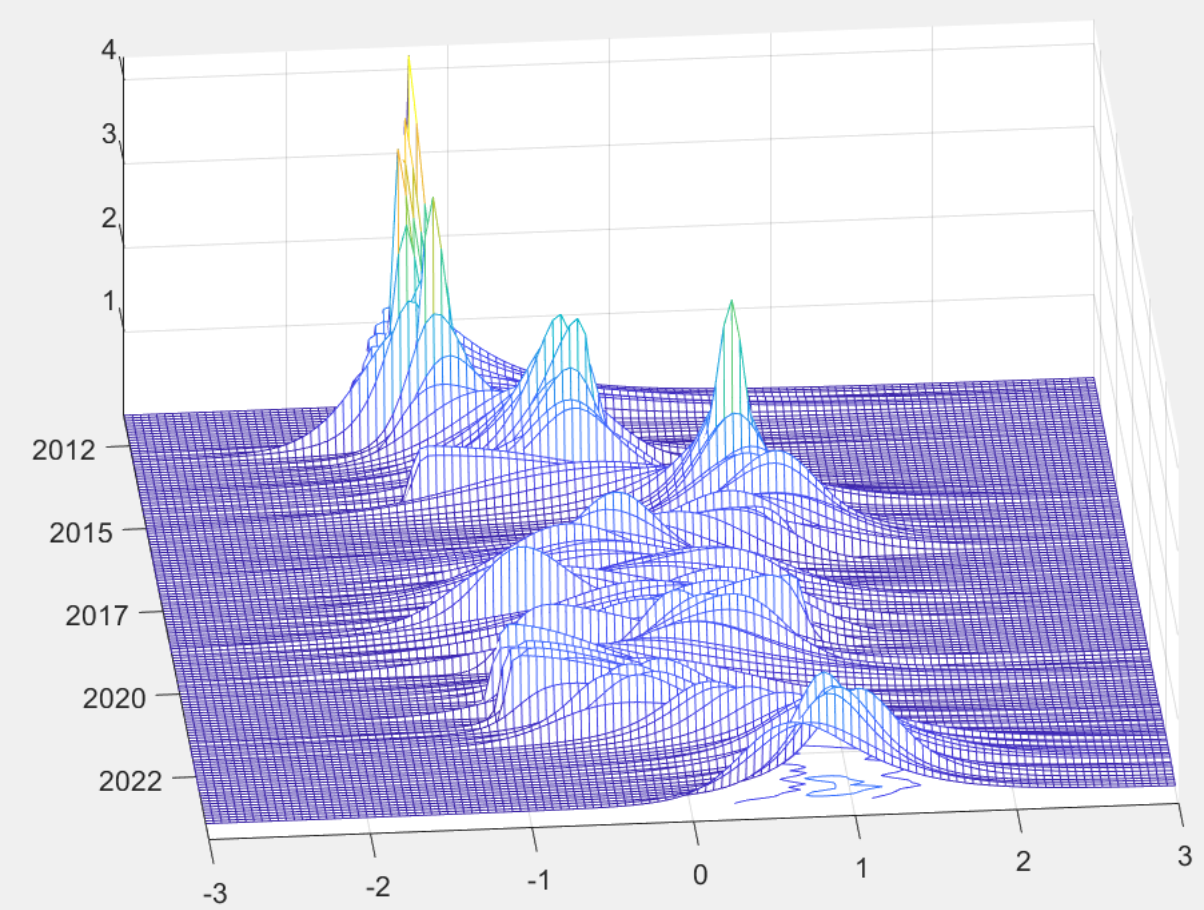
The estimated predicted distributions vary considerably over time. Figure 8 illustrates this using the 6-month ahead MCI for the Treasury market as an example. Not only the mean and the variance of the fitted distributions vary over time; the fatness of the tails also changes and the distributional asymmetry shifts considerably, with both positive and negative skewness evident at different times. This underscores the importance of focusing on the tail of the distribution rather than the mean or the median when constructing early-warning signals of market stress.

The substantial time-variation in the distributions illustrated in Figure 8 raises the question whether the fitted distributions accurately describe movements in MCI distributions. To assess the accuracy of our RF-based density forecasts we evaluate the predicted cumulative distribution at future realized values using the probability integral transform (PIT). In practice, this amounts to measuring the percentage of realized future MCI observations that fall below a given model-implied predicted quantile. If the model is accurate, the cumulative distribution of the PITs will lie on a 45-degree line. By contrast, if the model fails in capturing some part of the distribution, we would see a pronounced deviation from the 45-degree line corresponding to that part of the distribution.

As Figure 9 shows, the random forest model produces forecasts that capture well the realized distribution of MCIs across a range of forecast horizons and for all three markets considered. In all cases, the predicted cumulative distribution of the PITs are close to the 45-degree line, never straying outside the theoretical 95% confidence bands of the diagonal line.[16] This shows that the RF method does a good job in capturing all parts of the MCI distribution, including the upper tail that is relevant for capturing stressed conditions.

[16] The confidence bands are estimated according to Rossi and Sekhposyan (2019); see also the discussion in Adrian et al. (2019).

Figure 8: Treasury MCI distributions six months ahead
This graph displays estimated distributions for predicted Treasury MCI, six months ahead. The distributions are obtained by fitting a skewed t-distribution to a range of forecasted quantiles obtained using the random forest method.



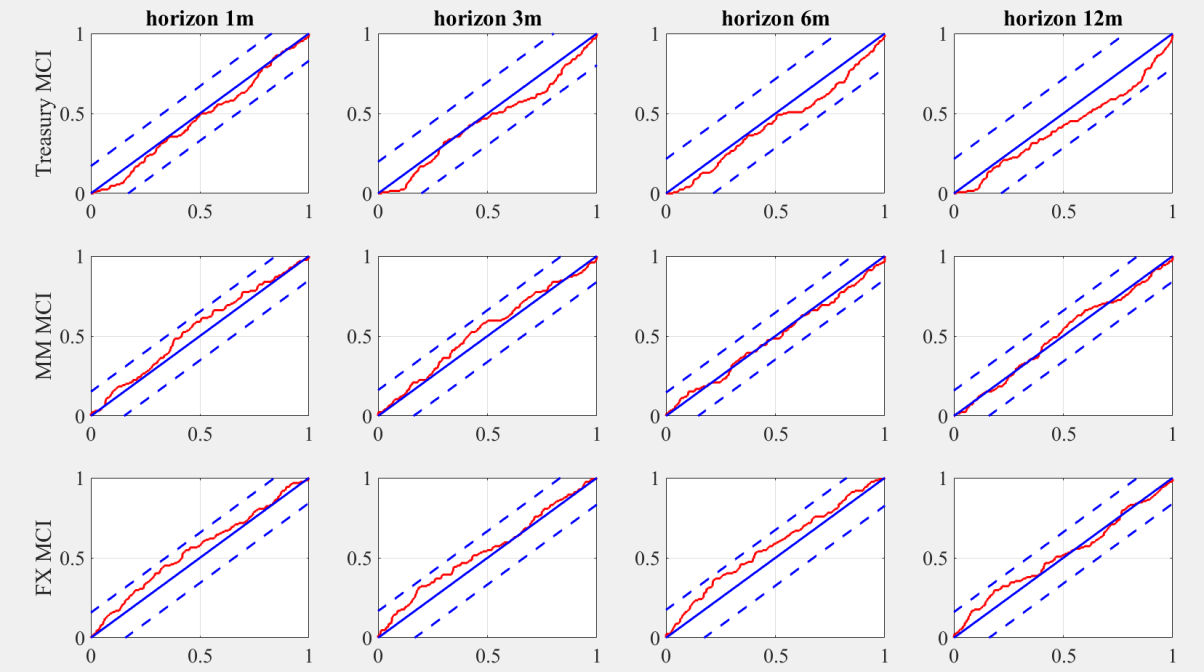
Having validated the forecasting performance of our MCI distributions, we next turn to some examples that illustrate how different types of predictors may affect the distribution forecasts. We focus on the six-month forecast horizon, allowing us to inform the choice of variables we study using the Shapley analysis in Section 4.

We proceed as follows. For each market segment, we examine how the exclusion of the top predictive feature affects the 6-month ahead MCI distribution at some interesting point in time for that market. In an environment characterized by a large number of predictors, individual predictors are often highly correlated. As a result, the flexibility of the RF approach enables it to typically "make up" for any individual dropped predictor, and produce a forecast that is little changed. In what follows, we therefore exclude not only the top predictor, but also any other predictor that has an absolute correlation with it that is higher than 0.6 (a "high", but admittedly arbitrary cut-off).[17]

[17] In practice, we do not fully drop these variables, but set all their historical values equal to their respective median values. This way, we can preserve the structure of the random forests while eliminating any informational value

Figure 9: Cumulative PITs of MCI density forecasts

This graph displays the empirical cumulative probability integral transforms (PITs) of density forecasts from the RF model, evaluated at future realizations (red lines). The diagonal 45-degree lines correspond to the ideal outcome of a perfect predictive model (blue solid lines). The dashed lines are 95% confidence bands around this theoretical ideal. The x-axis displays the quantiles and the y-axis the empirical cumulative distribution.



The first panel in Figure 10 considers the six-month ahead Treasury MCI at end-January 2022. This is an interesting period because it is in the run-up to the first Fed rate hike in an environment of elevated inflation (the first increase in a series of rate hikes subsequently took place in March 2022). At this forecast date, the unrestricted RF model's MCI density was centered at a positive value, pointing to tighter than typical Treasury market conditions ahead. But if the model is restricted by excluding the top predictor (swaption implied volatility) and highly correlated features[18], the resulting distribution paints a more sanguine picture, with the mode now firmly in negative territory.

Moving on to the second panel in Figure 10, this shows density forecasts for the money market at end-September 2016. This was a month before the US SEC's money market reform was implemented, the lead-up to which was associated with some tension in money markets. As in the previous example, if we restrict the most important predictor (the GFC index) and associated correlated predictors[19], the resulting MCI distribution is centered deeply in negative territory, suggesting easy conditions ahead. The unrestricted distribution, by contrast, is centered half a standard deviation higher, pointing to more tightness. Moreover, it has higher variance than the

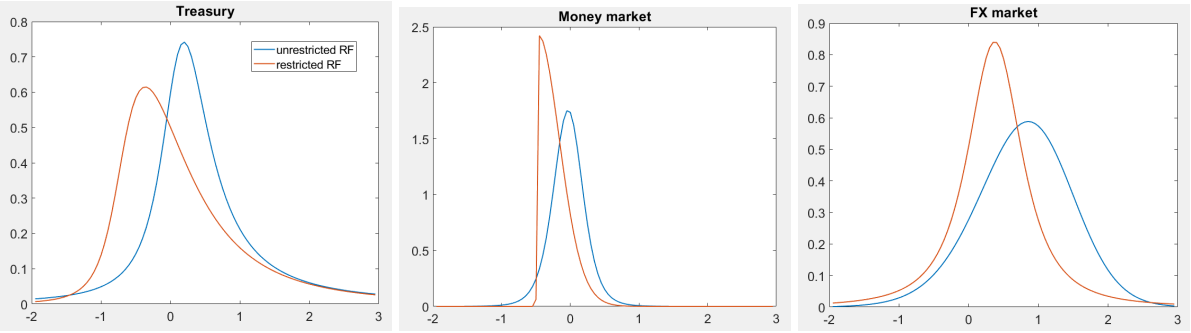
from the affected variables.

[18]This includes three variables: the MOVE index, the lagged Treasury MCI, and CP issuance.

[19]The Treasury General Account balance, the Treasury holdings of primary dealers, and Treasury futures margin requirements.

Figure 10: Sample MCI distributions six months ahead

This graph displays estimated distributions for predicted MCIs, six months ahead at different points in time. The distributions are obtained by fitting a skewed t-distribution to a range of forecasted quantiles obtained using the random forest method. "Unrestricted RF" is the estimated distribution from a random forest model that includes all predictive features; "restricted RF" is the estimated distribution from a random forest model that drops the top features (as well as highly correlated ones). The forecast date for the Treasury densities is January 2022; for the money market September 2016; and for the FX market December 2014.



restricted one, indicating a greater probability of high-tail outcomes.

Finally, the third panel of Figure 10 considers FX market densities in December 2014, just before the Swiss franc de-peg from the euro. This event tightened FX market conditions materially. Although the de-peg event itself was essentially fully unpredictable, our RF model shows that conditions in markets in the run-up to the event suggested already an elevated likelihood of market stress. Indeed, our unrestricted RF model points to outcomes associated with high levels of stress having significant probability mass. The unrestricted predictive density is centered at a high value (around one standard deviation above the mean), and it has a high variance, further boosting the probability of very high outcomes. The restricted density, instead, is centered much to the left of the unrestricted one, has lower variance, and therefore substantially lower probability mass at high MCI values.[20]

These examples all show that variables with important leading indicator properties for tail outcomes can play a key role in shaping the entire distribution of future MCIs. Of course, given the non-linear nature of the random forest method and the MCI distributions implied by its estimates, the impact of restricting these variables also varies over time. In any case, these examples illustrate that interpretability—a key criterion for successful use for early-warning exercises in a policy context—is achievable for results based on random forest analysis.

6 Conclusion

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

[20]The restricted model drops the 2-year/3-month US yield slope, and the correlated variables 10-year/2-year yield slope and 3-month interest rate.

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 macro-related investor expectations and associated uncertainty indicators, especially about the monetary policy stance, in explaining higher forecasted tail realizations of market conditions. Global financial cycle dynamics and liquidity indicators also play an important role. In addition, in some market segments 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).

We also show that the random forest model allows us to very accurately fit the entire distribution of future MCI values for any given forecast horizon, across market segments. The resulting MCI distributions vary considerably over time, with respect to mean, variance, skewness and kurtosis. By estimating the full distribution of future MCIs, we are able to examine which predictors are key in shaping these distributions.

These findings have important implications. For policymakers, the MCIs offer a real-time diagnostic tool to identify market dysfunction, complementing existing FSIs and FCIs. In addition to the important role played by monetary policy, 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. Moreover, our use of Shapley values illustrates the feasibility of achieving some degree of explainability in applications of ML.

Our work points to at least four avenues for future exploration. First, extending the MCIs to corporate bond, equity 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 further. Third, incorporating alternative data—such as text-based sentiment—may 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/sfwpfi/201810.html.

Azzalini, Adelchi, and Antonella Capitanio. 2003. "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 367–389.

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.

Carlson, Mark, Kurt Lewis, and William Nelson. 2014. "Using policy intervention to identify financial stress." *International Journal of Finance & Economics* 19 59–72.

Chavleishvili, Sulkhan, and Manfred Kremer. 2025. "CISS of death: measuring financial crises in real time." *Review of Finance* 29 (3): 685–710.

Chavleishvili, Sulkhan, 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 Shachar, 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 Somogyi. 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*.

Markets Committee. 2022. "Market dysfunction and central bank tools." May.

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.

Rossi, Barbara, and Tatevik Sekhposyan. 2019. "Alternative tests for correct specification of conditional predictive densities." *Journal of Econometrics* 208 638–657.

Scheicher, Martin, and Andreas Schrimpf. 2022. "Liquidity in bond markets – navigating in troubled waters." *SUERF Policy Brief* (395): .

Schrimpf, Andreas, Vladyslav Sushko, and Hyun Song Shin. 2020. "Leverage and margin spirals in fixed income markets during the Covid-19 crisis." *BIS Bulletin* (2): .

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.

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.

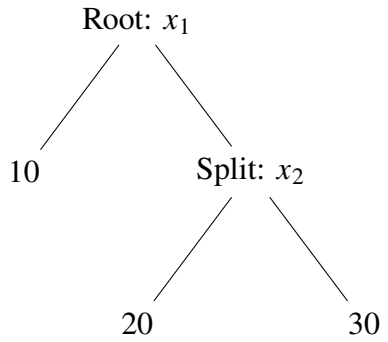
Tree Traversal. For each feature, the algorithm traverses the tree to determine the contribution of that feature at each node. This involves calculating the change in the model's prediction when a feature is included versus when it is excluded.

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:

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

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.

A2 Performance of the neural network model

To evaluate the performance of the neural network model, we use the same estimation and testing windows as outlined in the main text. Within each estimation window, we train the neural network model to tune four key hyperparameters: the number of hidden layers (selected from [1, 2, 3]), the number of neurons in each layer (selected from [16, 32, 48, 64, 96]), the activation function (selected from ['relu', 'tanh', 'sigmoid']), and the learning rate (selected from [0.001, 0.01, 0.1]). After identifying the optimal hyperparameters, we assess the model's performance by calculating the quantile loss for the out-of-sample observations.

Figure 1 illustrates the out-of-sample quantile loss of the neural network model relative to the autoregressive model specified in equation (4). Overall, the neural network model does not outperform the autoregressive model in the short-term (up to 3 months) for predicting stress across all three markets. For the US Treasury market MCI, the neural network shows better performance only for predicting the 95% percentile of its distribution. For the money market MCI, the neural network performs better than the autoregressive model starting from the three-month forecast horizon for predicting the 90% and 95% percentiles, and from the five-month forecast horizon for predicting the median market condition. In the main text, we show that the random forest model consistently outperforms the autoregressive model in forecasting liquidity condition stress in the money market starting from one or two months ahead. Similarly, the performance of the neural network model is not better than the one of the autoregressive model for predicting the FX market MCI from one- to five-month forecast horizons. In the longer term, however, the performance of the neural network model improves and becomes relatively better, but still comparable with the one of random forest. Therefore, our analysis here confirms that the neural network model does not strictly outperform the tree-based model for our exercise, particularly in predicting money market and FX market conditions.

Figure 1: Out-of-sample performance of neural network

This graph compares quantile losses between the neural network and autoregressive models based on out-of-sample predictions across forecast horizons. Negative values indicate better performance of the neural network model. Shaded areas indicate 90% confidence intervals. The sample period is from January 2003 to May 2024.