Pricing Macro-Panic: Forecasting the Tails of Bond Risk Premia*

Jiaxun Liu†

First Draft: March 2026 *This Draft: October 2026*
Click here for latest version

Abstract

Moving beyond conventional mean-targeted forecasts, I predict the entire distribution of bond risk premia to capture extreme, crisis-driven tail movements. Augmenting standard predictors with an extracted upper-tail macroeconomic factor—representing a latent "macro-panic"—achieves full predictive dominance over baseline models across the quantile spectrum. Methodologically, a Quantile Random Forest (QRF) efficiently captures non-linear crisis thresholds using mean-selected predictors, offering a computationally tractable, high-performing alternative to Quantile LASSO. The resulting forecasts exhibit pronounced regime-dependent cyclicality, becoming heavily countercyclical during recessions. Crucially, this approach reveals a term-structure duality: short-term tail risks are driven by acute macro-liquidity panics ("dash for cash"), whereas long-term risks are anchored by structural safe-haven demand ("flight to safety"). Finally, I document a decay in macroeconomic predictability along the yield curve: while the QRF dominates at shorter maturities, naive empirical quantiles become highly competitive at the long end.

Keywords: Bond Risk Premium, Tail Risk, Term Structure, Forecasting, Quantile Random Forest, Macro-Finance, Systemic Risk
JEL: C53, E44, G17

*I would like to thank Christian Brownlees for the support throughout the project, Daniele Bianchi and Daniele Massacci for the valuable feedback and all of the insightful conversations at the FinEML Conference 2025, the 50th symposium of the Spanish Economic Association, the RCEA ICEEF Conference 2026 and the IAAE Annual conference 2026. All errors are my own.

†Faculty of Economics and Business, Universitat Pompeu Fabra. Email: jiaxun.liu@upf.edu

1 Introduction

During severe economic shocks, the compensation investors demand for holding duration risk experiences extreme, bidirectional swing—either spiking due to binding intermediary constraints or collapsing amid massive flights to safety. Yet, for over two decades, the standard empirical asset pricing literature has predominantly modeled bond risk premia at the conditional mean. Within this traditional framework, researchers actively debate the optimal predictor set. Proponents of the "spanning hypothesis" argue that all necessary forecasting information is already embedded within the yield curve itself (Cochrane and Piazzesi, 2005; Adrian et al., 2013). This view is reinforced by studies warning that macroeconomic variables might just look predictive due to small-sample illusions (Bauer and Hamilton, 2018). Conversely, alternative approaches attempt to augment yield factors with broad macroeconomic aggregates (Ludvigson and Ng, 2009) or forward-looking survey expectations (Eriksen, 2017). Regardless of the predictors chosen, these classical models implicitly assume that excess returns respond linearly and proportionally to standard business cycle fluctuations.

While this assumption holds during quiet economic expansions, it potentially masks the true nature of tail risk in the bond market. Systemic market distress is characterized by threshold effects and sudden structural breaks, rather than smooth macroeconomic transitions (Brunnermeier and Pedersen, 2009; He and Krishnamurthy, 2013). Traditional linear frameworks fail to capture how sudden financial frictions govern both the extreme highs and extreme lows of the yield curve. Capturing these crisis dynamics requires a shift from conditional means to the state-dependent extremes of "Term-Premia-at-Risk," adapting the distributional framework of Adrian et al. (2019). By moving beyond conventional mean-targeted forecasts, this paper predicts the entire distribution of bond risk premia, aiming to capture the extreme, crisis-driven tail movements that define severe market risk.

To understand why these extremes are so critical, it is necessary to examine the bidirectional feedback loop between the macroeconomy and the bond market. Classical economic theories dictate that rational investors demand higher compensation when macroeconomic uncertainty rises (Campbell and Cochrane, 1999; Bansal and Yaron, 2004). Conversely, the

way bond premia affect the real economy is fiercely debated, largely because a decrease in the term premium can reflect two different economic states. On one hand, market-driven term premium compression often signals an impending recession, as fearful investors execute a "flight to safety" into long-term bonds (Hamilton and Kim, 2002; Wright, 2006). On the other hand, central banks operate on the premise that an artificial, policy-driven decrease in the term premium—such as through Quantitative Easing (QE)—injects vital stimulus into the economy (Bernanke, 2013; Waller, 2024). As criticized by Rudebusch et al. (2006) and Ireland (2015), this complex macro-finance interaction is largely neglected by traditional New Keynesian models, which typically collapse the yield curve into a single short-term policy rate and abstract away from the real economic effects of long-term duration risk. Operating outside these traditional theoretical constraints, the modern practitioner's reliance on QE creates its own systemic vulnerabilities—prolonged accommodative monetary policy can encourage risk-taking, setting the stage for systemic financial crises when adverse shocks eventually hit (Adrian and Liang, 2016). When these financial conditions rapidly tighten, economic growth becomes highly vulnerable (Adrian et al., 2019).

Because policy responses and market panics are most extreme during crises, forecasting the tails of the bond risk premium is crucial. Historically, sharp downward movements in long-term yields have been strong leading indicators of impending recessions (Estrella and Mishkin, 1998; Rudebusch and Williams, 2009; Chauvet and Senyuz, 2016). This lower-tail compression is largely driven by a "flight-to-safety", where panicked investors willingly accept much lower yields to hold assets they can instantly convert to cash (Longstaff, 2002; Nagel, 2016). However, when true panic spreads and intermediary constraints bind, liquidity can evaporate entirely. During a "dash for cash", investors liquidate even safe Treasury holdings, breaking the normal flight-to-safety dynamic and causing risk premia to violently spike. Illustrating this structural break during the March 2020 crash, former Federal Reserve Governor Lael Brainard (2021) noted:

> "The abrupt repositioning and repricing of portfolios led to a dash for cash, as even relatively safe Treasury holdings were liquidated... Trading conditions deteriorated rapidly in the second week of March as a wide range of investors

sought to raise cash by liquidating the Treasury securities they held."

Consequently, by forecasting the entire distribution, this paper suggests that these extremes do not merely coexist but operate sequentially. During sudden market breaks, the acute phase of a crisis is led by a scramble for liquidity that forces upper-tail premia to violently spike (He et al., 2022). This is followed by a delayed safe-haven compression at the lower bound, drastically widening the overall distribution of expected returns (Swanson and Williams, 2014; Vissing-Jorgensen, 2021; Duffie, 2020; Haddad et al., 2021).

Business cycles are inherently asymmetric: economic expansions are typically prolonged and stable, whereas recessions are brief and violent. Because traditional "level" macroeconomic factors are designed to capture the average state of the economy, they are heavily anchored to these quiet expansionary periods and naturally smooth over sudden crisis events. To address this limitation and explicitly capture non-linear tail dynamics, I extract quantile factors (Chen et al., 2021) from a large macroeconomic panel using the FRED-MD database (McCracken and Ng, 2016), forecasting the non-parametrically smoothed zero-coupon Treasury Bond yield curve of Liu and Wu (2021). By estimating factors at the extreme ends of the macroeconomic distribution, I am able to disentangle different types of economic distress. Specifically, the lower-tail factor (F_t^{10}) captures "Main Street" stagnation—collapsing industrial production and real economic output. However, bond risk premia do not necessarily spike simply because physical output slows. Instead, I demonstrate that the upper-tail macro factor (F_t^{90}) serves as a superior predictor. Unlike F_t^{10}, this upper-tail factor mathematically aggregates extreme volatility across collapsing asset prices and broad financial distress. It isolates a comprehensive, data-driven latent measure of systemic "macro-panic", capturing the late-cycle stagflationary shocks and surging interest rates that manual stress indices often miss. Ultimately, pairing this macro-panic factor with the level macro factors (Ludvigson and Ng, 2009) and the financial yield curve factor (Cochrane and Piazzesi, 2005) produces the dominant out-of-sample predictive performance across the entire quantile spectrum.

Methodologically, capturing these shifting regimes requires moving beyond linear frameworks. Recent years have seen a surge in the application of machine learning methods

within empirical macrofinance (Bianchi et al., 2021; Chen et al., 2024). Specifically, tree-based algorithms like Random Forests have proven highly effective in forecasting complex financial volatility and asset returns (Luong and Dokuchaev, 2018; Gupta et al., 2019). While linear quantile regressions attempt to model distribution extremes, they frequently suffer from severe out-of-sample extrapolation errors when confronted with highly non-linear market shocks. To test whether specific predictors drive specific tails, I employ an in-sample Quantile LASSO. The results confirm that extreme quantiles select distinctly smaller sets of macro-financial variables compared to central quantiles, proving that crisis regimes are governed by forces that differ from normal market conditions. However, in out-of-sample forecasting, applying these quantile-specific variables yields limited predictive gains. As for the out-of-sample forecast, I utilize a Quantile Random Forest (QRF) (Meinshausen, 2006). By using recursive decision trees, the QRF inherently captures non-linear crisis thresholds and sudden regime shifts, natively identifying the macro-panic thresholds needed to predict sudden spikes in the bond risk premium.

To clarify the empirical design, the forecasting exercise targets one-year-ahead excess bond returns, $xr_{t+1}^{(n)}$, across two-to-ten-year maturities. The evaluation proceeds in two stages. First, in-sample analysis contrasts mean-selected (OLS) against quantile-specific (Quantile LASSO) predictors to identify shifting macroeconomic risk drivers at the extremes. Second, out-of-sample predictions are generated via a recursive expanding window across a dense quantile grid ($\tau \in [0.01, 0.99]$) comparing QRF and linear quantile forecasts. To straightforwardly isolate the value of tail information, the out-of-sample exercise compares five competing models: a naive empirical quantile baseline; a "financial only" model using the Cochrane and Piazzesi (2005) yield-curve factor; a "macro + financial" model adding standard macroeconomic principal components (Ludvigson and Ng, 2009); the fully "quantile factor augmented" model incorporating F_t^{90}; and a Quantile LASSO benchmark. Forecast accuracy is then evaluated using the asymmetric mean check loss to map exactly where conditional predictability breaks down across the term structure.

Furthermore, my analysis of the cyclicality of bond risk premia forecasts demonstrates clear regime differences across the business cycle. The correlation between the forecasts and Indus-

trial Production is heavily countercyclical during recessions, yet mildly cyclical (or acyclical) during economic expansions, highlighting the asymmetric nature of risk pricing. Crucially, forecasting the extremes uncovers an underlying term-structure duality. Results suggest that upper-tail risks at the short end of the curve are potentially driven by acute macro-liquidity panics (the "dash for cash"), whereas lower-tail risks at the long end are anchored by structural safe-haven demand (the "flight to safety"). The fact that short-term tail risks are tightly anchored to transient macro-policy cycles and immediate intermediary constraints potentially explains why the QRF dominates at shorter maturities. However, as these immediate liquidity panics dilute over longer horizons, this advantage narrows. At the 10-year horizon, simple empirical quantiles become highly competitive. These findings coincide with Ang and Piazzesi (2003), who demonstrated that macroeconomic factors primarily explain movements at the short end and the middle of the yield curve, illustrating how transient macroeconomic signals eventually give way to distant, slow-moving structural shocks (Cieslak and Povala, 2015). Crucially, these findings are not merely artifacts of the recent pandemic. Extensive robustness checks confirm that excluding the extreme COVID-19 period from the out-of-sample evaluation leaves the predictive dominance of the macro-panic factor and the QRF model unchanged, while yielding deeper insights into the nature of tail risks.

The remainder of the article is structured as follows. Section 2 lays out the construction of bond risk premia, the extraction of macroeconomic and quantile factors, and the econometric framework. Section 3 conducts in-sample factor selection, defines the macro-panic factor, and presents out-of-sample forecast performances across model specifications. Section 4 discusses the cyclicality and term structure dynamics of the bond risk premia forecasts. Section 5 conducts robustness check by excluding the COVID-19 period. Finally, Section 6 concludes.

2 Methodology

2.1 Bond Excess Returns

The forecast target of this study is the excess return on zero-coupon Treasury bonds, formulated in line with Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009). Denote the log price and the continuously compounded yield of a zero-coupon bond at time t (with maturity at $t+n$ and payoff of one dollar) by $p_t^{(n)}$ and $y_t^{(n)} = -\frac{1}{n}p_t^{(n)}$ respectively. The log excess return to holding an n-year bond for one year (from t to $t+1$, with remaining maturity $n-1$) is denoted by:

$$xr_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)} \tag{1}$$

where $y_t^{(1)}$ signifies the yield on a 1-year bond at time t. Throughout the article, the main forecast targets are the extreme conditional quantiles of $xr_{t+1}^{(n)}$ across both short-term and long-term maturity horizons.

2.2 Construction of Predictive Factors

To forecast the distribution of bond risk premia, I construct three distinct sets of predictive factors: a yield-curve financial factor, level macroeconomic factors, and upper-tail quantile macroeconomic factors.

Financial Factor (CP_t): Following Cochrane and Piazzesi (2005), the financial factor is designed to capture the predictive information embedded entirely within the yield curve. The log forward rate for an n-year bond at time t is defined as $f_t^{(n)} = p_t^{(n-1)} - p_t^{(n)}$. To exploit the information across the entire maturity spectrum (up to 10 years), the single financial factor CP_t is constructed as the fitted value from a regression of the average excess bond return

onto the 1-year yield and the forward rates:

$$\overline{rx}_{t+1} = \frac{1}{9}\sum_{n=2}^{10} xr_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(2)} + \cdots + \gamma_{10} f_t^{(10)} + \overline{\varepsilon}_{t+1}$$

$$CP_t = \hat{\boldsymbol{\gamma}}' \boldsymbol{f}_t \tag{2}$$

Level Macroeconomic Factors (F_t): To augment the yield curve information, I extract broad macroeconomic aggregates from the high-dimensional FRED-MD database (McCracken and Ng, 2016). The level factors (F_t) are extracted via Principal Component Analysis (PCA), targeting the conditional mean of the macroeconomic panel following the methodology of Ludvigson and Ng (2009).

Quantile Macroeconomic Factors (F_t^τ): Because traditional PCA inherently targets the mean, it naturally smooths over the extreme volatilities that characterize severe economic shocks. To capture these threshold effects, I extract quantile-specific macroeconomic factors following Chen et al. (2021). Rather than targeting average conditions, this methodology extracts latent factors that represent the tail behavior of the macro-panel. Specifically, I isolate the upper-tail "macro-panic" factor, F_t^{90}, which serves as a proxy for systemic financial distress and binding intermediary constraints. This will be further discussed in section 3.2.2.

2.3 Econometric Framework

With the predictors defined, the objective shifts to modeling the conditional quantiles of excess bond returns. The τ-th conditional quantile of $xr_{t+1}^{(n)}$ given information set $\mathcal{I}_t$ is denoted as $Q_\tau(xr_{t+1}^{(n)}|\mathcal{I}_t)$. Forecast accuracy is evaluated using the asymmetric check-loss function:

$$\rho_\tau(\mathbf{e}) = \begin{cases} \tau \cdot e, & e \geq 0 \\ (\tau - 1) \cdot e, & e < 0 \end{cases} \tag{3}$$

where $e = xr_{t+1}^{(n)} - \hat{Q}_\tau(xr_{t+1}^{(n)}|\mathcal{I}_t)$. This function heavily penalizes absolute forecast errors that deviate from the target quantile τ. To comprehensively model these distributions, I compare three distinct econometric techniques.

2.3.1 Linear Quantile Regression and Quantile LASSO

The baseline approach assumes the τ-th quantile of bond risk premia is a linear combination of the predictors X_t. The standard linear quantile regression obtains estimates by minimizing the unpenalized check-loss over the training sample.

However, because different macro-financial forces may drive different tails of the distribution (e.g., a flight-to-safety versus a liquidity panic), the optimal subset of predictors might vary across τ. To conduct quantile-specific variable selection, I employ the L_1-penalized Quantile LASSO (Belloni and Chernozhukov, 2011):

$$\hat{\beta}^\tau = \arg\min_\beta \frac{1}{T}\sum_{t=0}^{T} \rho_\tau(xr_{t+1}^{(n)} - X_t\beta) + \lambda \sum_{j=1}^{k} |\beta_j| \tag{4}$$

where λ is the tuning parameter selected via cross-validation. To evaluate the explanatory power of these quantile-selected models against a naive empirical quantile baseline, one can compute the Pseudo-R^2:

$$R_\tau^2 = 1 - \frac{\sum_t \rho_\tau\left(xr_{t+1}^{(n)} - \hat{Q}_\tau(xr_{t+1}^{(n)}|X_t)\right)}{\sum_t \rho_\tau\left(xr_{t+1}^{(n)} - \hat{Q}_\tau(xr_{t+1}^{(n)})\right)} \tag{5}$$

where the denominator represents the check-loss of the unconditional empirical quantile. Positive Pseudo-R^2 indicates better performance compared to the baseline.

2.3.2 Quantile Random Forest (QRF)

While Quantile LASSO accommodates quantile-specific predictor selection, it remains constrained by its linearity assumption. During severe economic crises, structural breaks and

threshold effects govern market pricing. To natively capture this non-linearity without relying on unstable sparse selection, I implement the Quantile Random Forest (Meinshausen, 2006).

Unlike standard random forests which average the conditional mean across trees, QRF retains the value of all observations within the leaf nodes to construct a conditional empirical distribution function. Let $w_i(X_t)$ represent the aggregated weight of historical observation i given the current predictor state X_t. The QRF estimates the conditional distribution as:

$$\hat{F}(y|X_t) = \sum_{i=1}^{T} w_i(X_t)\mathbf{1}(xr_{i+1}^{(n)} < y) \tag{6}$$

The τ-th conditional quantile forecast is then obtained by inverting this distribution:

$$\hat{Q}_\tau(xr_{t+1}^{(n)}|X_t) = \inf\{y : \hat{F}(y|X_t) \geq \tau\} \tag{7}$$

By utilizing recursive binary splitting, QRF seamlessly interacts the financial yield factors (CP_t) with the latent macro-panic factor (F_t^{90}). It intrinsically identifies the crisis thresholds needed to forecast sudden structural spikes in bond risk premia.

3 Empirical Application

3.1 Data and Sample

The empirical analysis utilizes monthly data spanning from August 1971 to December 2023. The bond yield data is sourced from the non-parametrically kernel-smoothed zero-coupon Treasury bond yield curve reconstructed by Liu and Wu (2021). This continuous yield curve allows for the extraction of bond risk premia across a wide range of maturities, from 1 to 10 years. The macroeconomic predictors are derived from the FRED-MD database (McCracken and Ng, 2016), which contains 126 monthly-updated macroeconomic time series

categorized into distinct economic dimensions (e.g., output, labor, prices, stock market). Detailed listings of the policy-relevant variables and their stationary transformations are provided in Appendix A.

To evaluate out-of-sample predictability, the dataset is split into an initial in-sample training window from August 1971 to January 2000 (accounting for 55% of the sample). The out-of-sample evaluation period runs from January 2000 to December 2023. Forecasts are generated recursively: before each 12-month-ahead prediction is made, all financial and macroeconomic factors are dynamically re-estimated using the expanding informational window.

3.2 In-Sample Analysis: Mean-Targeted Selection

Following the traditional mean-targeted predictive framework, I first analyze the in-sample predictive power of the extracted factors using ordinary least squares (OLS). The exact form of the fully augmented mean regression is specified as:

$$xr_{t+1}^{(n)} = \beta_0 + \boldsymbol{\beta}_1' \boldsymbol{F}_t + \beta_2 CP_t + \beta_3 F_t^{90} + \epsilon_{t+1} \tag{8}$$

where the target variables are the 1-year ahead excess bond returns of various maturities. To account for the MA(12) structure in the overlapping annualized continuously-compounded returns, standard errors are corrected using the Newey and West (1987) HAC estimator with an 18-month lag following Ludvigson and Ng (2009). Table 1 presents the regression estimates across multiple maturities.

3.2.1 Level Macroeconomic Factors

I begin with the macroeconomic factors, which are selected via principal component analysis (PCA) over the FRED-MD dataset within the in-sample window and the optimal number of principal components is determined by the information criteria of Bai and Ng (2002).

Table 1: Regression of Monthly Excess Bond Returns on Lagged Predictors

		F_{1t}	F_{2t}	F_{3t}	F_{4t}	F_{6t}^2	F_{6t}^3	CP_t	F_t^{90}	R^2
$rx_{t+1}^{(2)}$	(a)							**0.271**		0.320
								(6.558)		
	(b)	**0.023**	**0.027**	**0.026**	**0.024**	**-0.007**	**-0.002**			0.376
		(5.015)	(3.606)	(4.659)	(3.143)	(-3.219)	(-6.067)			
	(c)	**0.015**	0.011	**0.017**	**0.022**	**-0.007**	**-0.002**	**0.214**		0.500
		(3.232)	(1.465)	(2.572)	(3.613)	(-3.883)	(-6.133)	(4.688)		
	(d)	**0.011**	0.006	**0.020**	**0.021**	**-0.006**	**-0.001**	**0.214**	**0.114**	0.514
		(1.976)	(0.970)	(2.650)	(3.607)	(-3.536)	(-6.600)	(4.732)	(1.725)	
$rx_{t+1}^{(3)}$	(a)							**0.052**		0.312
								(6.891)		
	(b)	**0.022**	**0.047**	**0.048**	**0.041**	**-0.012**	**-0.004**			0.311
		(2.746)	(3.558)	(4.295)	(2.788)	(-3.205)	(-7.112)			
	(c)	**0.022**	0.018	**0.026**	**0.032**	**-0.012**	**-0.003**	**0.414**		0.495
		(2.879)	(1.335)	(2.301)	(3.191)	(-4.021)	(-6.459)	(5.071)		
	(d)	0.014	0.010	**0.031**	**0.030**	**-0.010**	**-0.003**	**0.414**	**0.209**	0.504
		(1.528)	(0.800)	(2.466)	(3.096)	(-3.541)	(-6.573)	(5.091)	(1.776)	
$rx_{t+1}^{(4)}$	(a)							**0.713**		0.381
								(7.427)		
	(b)	**0.025**	**0.066**	**0.063**	**0.051**	**-0.017**	**-0.005**			0.291
		(2.317)	(3.630)	(4.301)	(2.594)	(-3.396)	(-7.792)			
	(c)	**0.025**	0.023	**0.031**	**0.039**	**-0.017**	**-0.004**	**0.603**		0.499
		(2.483)	(1.270)	(2.127)	(3.006)	(-4.420)	(-7.002)	(5.303)		
	(d)	0.013	0.011	**0.039**	**0.035**	**-0.014**	**-0.004**	**0.602**	**0.299**	0.509
		(1.110)	(0.686)	(2.347)	(2.826)	(-3.769)	(-6.808)	(5.317)	(1.897)	
$rx_{t+1}^{(5)}$	(a)							**0.870**		0.381
								(7.650)		
	(b)	**0.026**	**0.080**	**0.074**	**0.057**	**-0.021**	**-0.006**			0.271
		(2.046)	(3.578)	(4.280)	(2.376)	(-3.591)	(-8.249)			
	(c)	**0.026**	0.026	**0.035**	**0.042**	**-0.021**	**-0.005**	**0.747**		0.485
		(2.227)	(1.197)	(1.965)	(2.758)	(-4.584)	(-7.322)	(5.312)		
	(d)	0.012	0.013	**0.043**	**0.037**	**-0.018**	**-0.004**	**0.746**	**0.364**	0.495
		(0.860)	(0.607)	(2.236)	(2.476)	(-3.880)	(-6.671)	(5.323)	(1.944)	
$rx_{t+1}^{(7)}$	(a)							**1.202**		0.396
								(7.747)		
	(b)	0.029	**0.111**	**0.095**	**0.069**	**-0.030**	**-0.009**			0.259
		(1.693)	(3.791)	(4.400)	(2.108)	(-3.924)	(-8.997)			
	(c)	0.022	0.036	0.040	**0.047**	**-0.030**	**-0.007**	**1.051**		0.489
		(1.958)	(1.228)	(1.757)	(2.341)	(-4.253)	(-8.05)	(5.011)		
	(d)	0.008	0.016	**0.052**	**0.040**	**-0.026**	**-0.006**	**1.048**	**0.530**	0.500
		(0.451)	(0.556)	(2.184)	(2.017)	(-4.134)	(-6.709)	(5.046)	(2.129)	
$rx_{t+1}^{(10)}$	(a)							**1.572**		0.368
								(6.629)		
	(b)	0.029	**0.144**	**0.111**	0.072	**-0.042**	**-0.012**			0.220
		(1.121)	(3.534)	(3.727)	(1.559)	(-4.051)	(-8.061)			
	(c)	0.029	0.044	0.037	0.043	**-0.043**	**-0.011**	**1.410**		0.445
		(1.228)	(1.367)	(1.348)	(1.534)	(-5.243)	(-7.334)	(4.472)		
	(d)	-0.003	0.012	**0.057**	0.032	**-0.035**	**-0.008**	**1.407**	**0.824**	0.460
		(-0.121)	(0.298)	(1.731)	(1.085)	(-3.959)	(-5.778)	(4.531)	(2.274)	

Note: This table shows estimates of in-sample OLS regressions of excess bond returns on different specifications of lagged predictors in equation (8). The target variable $xr_{t+1}^{(n)}$ is the excess log return on the n-year zero coupon Treasury bond. The F_t are selected (non-linear transformed) principal components estimated from FRED-MD. The CP_t factors are Cochrane and Piazzesi (2005) forward rate factors, where CP_t is the overall forward rate factor. F^{90} is the 90% quantile factor of the FRED-MD. Newey and West (1987) corrected t-statistics with 18-month lag reported in parentheses. Coefficients significant at 5% or better marked bold. All numbers rounded to three decimal places.

After obtaining the initial six optimal components (F_{1t} through F_{6t}), it is useful to understand the underlying economic variations they capture. Appendix B details the top underlying FRED-MD variables most highly correlated with each principal component. Broadly, the loadings indicate that F_{1t} captures real activity and employment, F_{2t} reflects term spreads, F_{4t} encapsulates interest rates and the housing market, F_{3t} tracks aggregate inflation pressures, and F_{6t} summarizes asset prices and stock market dynamics.

Following Ludvigson and Ng (2009), I then generate the squared and cubed terms of these selected components (for example, F_{6t}^2 and F_{6t}^3) to exploit potential non-linear effects. The extended principal components then enter the model selection procedure by regressing target variables over an exhaustive combination of the level, squared, and cubed terms.

The specification generating the lowest BIC is then recorded. Without using the financial factors, the selected model contains F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}^2, and F_{6t}^3. Model (b) in Table 1 displays the performance of this macro-only specification. Notice that while the exact model selected for excess bond returns of different terms can vary (the number of 5%-significant factors decreases as the bond term grows), the specification selected with regard to $xr_{t+1}^{(2)}$ is taken as the benchmark for evaluation.

To proceed, the *overall* financial factor CP_t enters the model selection procedure alongside the extended set of macro-panel principal components. The new specification with the lowest BIC contains CP_t, F_{1t}, F_{3t}, F_{4t}, F_{6t}^2, and F_{6t}^3. In particular, after CP_t enters the regression (Model c), F_{2t} loses its predictive power—its Newey-West corrected coefficient is no longer significant at 5%—and is dropped from the optimal subset. Combining the financial factor with the level macroeconomic factors significantly improves the overall model fit (R^2), demonstrating that macroeconomic aggregates contain predictive information not fully spanned by the yield curve itself.

3.2.2 Tail Macroeconomic Factors: Uncovering the Macro-Panic

While the level factors provide a robust baseline during normal economic expansions, they naturally smooth over the sudden, non-linear market shocks that trigger structural breaks. To explicitly capture the asymmetric nature of business cycles and isolate the dynamics of severe economic distress, I augment the traditional predictor set with quantile-specific macroeconomic factors. Following the procedure proposed by Chen et al. (2021), I extract these quantile factors from the comprehensive FRED-MD panel.

Instead of targeting the conditional mean of the macro variables, this methodology targets the extremes of the macroeconomic distribution, allowing me to extract both a lower-tail factor (F_t^{10}) and upper-tail factors ($F_t^{90}, F_t^{95}, F_t^{99}$). It is important to note that the number of quantile factors extracted from the macroeconomic panel varies across the quantile spectrum, as the optimal number is decided through rank minimization rather than standard information criteria (Chen et al., 2021). A detailed breakdown of the number of selected factors per quantile τ is provided in Appendix C. Notice there are two 90% quantile factors, but the second one is omitted due to insignificance at in-sample selection, and thus all of the analyzed quantile factors above are scalar.

To understand the different economic states these tail factors capture, Table 2 presents their unconditional correlations with Industrial Production (IP) and standard level factors. An analysis of these correlations reveals why the upper-tail macro factor (F_t^{90}) serves as a superior predictor of bond risk premia compared to the lower-tail factor (F_t^{10}). The lower-tail factor is heavily dominated by traditional output contractions, evidenced by its massive negative correlation with both overall Industrial Production (-0.474) and the real-activity mean factor, F_{1t} (-0.328). Therefore, F_t^{10} represents a traditional, demand-driven "Main Street" recession where physical output collapses (not reported in Table 1 due to insignificance).

However, bond risk premia do not necessarily spike simply because factories slow down. During a pure output contraction, the Federal Reserve typically cuts rates, which lowers overall bond yields but does not necessarily cause investors to panic and demand a massive

Table 2: Correlation between Industrial Production (IP) and Predictive Factors

	IP	CP_t	F_{1t}	F_{3t}	F_{4t}	F_{6t}	F_t^{10}	F_t^{90}	F_t^{95}	F_t^{99}
IP	1.000									
CP_t	-0.070	1.000								
F_{1t}	0.140	0.075	1.000							
F_{3t}	0.099	0.119	0.149	1.000						
F_{4t}	-0.059	0.067	-0.188	-0.038	1.000					
F_{6t}	0.137	-0.133	0.057	0.007	0.023	1.000				
F_t^{10}	-0.474	-0.151	-0.328	-0.199	0.182	0.067	1.000			
F_t^{90}	-0.203	0.189	0.185	0.097	0.045	-0.305	-0.457	1.000		
F_t^{95}	-0.112	0.187	0.203	0.107	0.011	-0.300	-0.548	0.948	1.000	
F_t^{99}	-0.115	0.151	0.149	0.087	0.019	-0.262	-0.459	0.829	0.892	1.000

Note: The table displays the correlation between Industrial Production (IP) and the predictive factors: the financial factor CP_t, the mean macroeconomic factors $F_{1t}, F_{3t}, F_{4t}, F_{6t}$, the lower-tail factor F_t^{10}, and the upper-tail factors $F_t^{90}, F_t^{95}, F_t^{99}$.

risk premium. Instead, it is the upper-tail macro factor that serves as the dominant predictor of extreme market stress. Crucially, the model specifically selects the 90% upper-tail factor (F_t^{90}) over the more extreme quantiles (95% or 99%) because it provides an optimal signal-to-noise balance. The 90th percentile effectively isolates periods of sustained systemic distress, evidenced by its stronger countercyclical correlation with persistent real-economy variables like IP (-0.208 for F_t^{90} compared to just -0.114 for F_t^{99}).

In contrast to the demand-driven stagnation of F_t^{10}, the F_t^{90} factor exhibits the distinct signature of a late-cycle stagflationary shock or "Wall Street" panic. It correlates positively with inflation (F_{3t}) and interest rates (F_{4t}), while correlating negatively with the stock market (F_{6t}, at -0.305) and industrial output. This indicates that F_t^{90} captures periods of macro-financial tightening—where elevated price pressures and rising rates trigger broad financial market distress. Because bond risk premia are actually compensation for bearing inflation, rate volatility, and liquidity risks, F_t^{90} satisfactorily isolates the structural environments that force bond risk premia to widen or shrink. By anchoring to the 90th percentile, the model captures these underlying structural shifts while avoiding the extreme 95% and 99% tails, which are increasingly dominated by idiosyncratic noise, fleeting market anomalies, and data sparsity.

By mathematically aggregating extreme volatility across the panel's variables, F_t^{90} tracks broad financial distress and collapsing asset prices. Augmenting the model specification with F_t^{90} (specification d in Table 1) provides the exact countercyclical, financial-stress component required to forecast extreme spikes in the bond risk premium (results for F_t^{95} and F_t^{99} are not shown due to inferior in-sample R^2 compared to F_t^{90}).

3.3 In-Sample Analysis: Quantile-Specific Selection

The preceding mean-regression framework follows the spirit of Ludvigson and Ng (2009) by selecting variables that best explain the average variation in bond risk premia. However, since the primary objective of this study is to forecast the entire distribution of excess returns, the optimal subset of predictors might vary across different quantiles.

To formally investigate this regime-dependent predictability and distinguish it from conventional mean-selection, I perform an in-sample variable selection targeting specific quantiles of the benchmark 2-year excess bond return ($xr_{t+1}^{(2)}$) using the L_1-penalized Quantile LASSO (Belloni and Chernozhukov, 2011). Crucially, to avoid severe multicollinearity and confounding that destabilizes the L_1 penalty, the pre-computed non-linear macroeconomic transformations (e.g., F_{6t}^2 and F_{6t}^3) are excluded from the selection pool; the algorithm is fed only the base level predictors. Because the Quantile LASSO algorithm can exhibit instability and does not inherently choose the exact same set of variables in every iteration, I employ a robust resampling procedure to maintain tractability. Specifically, for each target quantile, I repeat the Quantile LASSO procedure 100 times. In each iteration, I utilize a 10-fold cross-validated, quantile-specific tuning parameter (λ) to penalize the check-loss function. A predictor is ultimately selected for a given quantile if it is chosen by the algorithm in more than 50% of the repetitions.

The full tables detailing the selected variables across the quantile spectrum for $xr_{t+1}^{(2)}$ are provided in Appendix D. The in-sample selection results confirm that extreme quantiles consistently select distinctly different, and often much smaller, subsets of macro-financial

variables compared to central quantiles. This empirical evidence validates the hypothesis that tail risks are driven by state-dependent mechanisms. However, while Quantile LASSO provides valuable inferential insights into variable importance across the distribution, forcing these highly non-linear market shocks into sparse, linear quantile specifications yields limited out-of-sample predictive gains. This will be discussed further in the following section.

3.4 Out-of-Sample Forecast Evaluation

3.4.1 Forecasting Benchmarks and Evaluation Metric

To rigorously evaluate out-of-sample predictability, I compare the performance of five candidate models: (1) the unconditional Empirical Quantile (EQ) which serves as a naive historical baseline, (2) the "financial only" model which relies purely on the yield-curve factor CP_t (Cochrane and Piazzesi, 2005), (3) the "macro + financial" model which adds the selected level macroeconomic factors (Ludvigson and Ng, 2009), (4) the "quantile factor augmented" model, which further incorporates the upper-tail macro panic factor F_t^{90}, and (5) the Quantile-LASSO selected model (Q-LASSO).

It is important to note a distinction in the exact predictor sets used depending on the forecasting algorithm. While the baseline linear quantile regressions must explicitly utilize the fully extended mean-selected predictors (including the non-linear transformations F_{6t}^2 and F_{6t}^3 identified in Section 3.2), the Quantile Random Forest natively captures complex non-linear effects and variable interactions through its recursive tree structure. Therefore, to avoid unnecessary confounding and over-parameterization, the QRF models are fed only the non-transformed base level terms.

Forecast accuracy is evaluated using the standard asymmetric check-loss function. It is mathematically crucial to note that raw check-loss values cannot be directly compared across different quantiles. By construction, the check-loss exhibits an inverse U-shape across the quantile spectrum, inherently generating higher loss values near the median and lower values

at the extreme tails, regardless of model accuracy. Therefore, performance must be evaluated relatively (as ratios) against a prespecified benchmark at each specific quantile (e.g. the "financial only" model).

To formally test whether the improvements in these relative check loss ratios are statistically significant, I employ the one-sided Diebold-Mariano (DM) test. Because the target variable utilizes overlapping annualized returns, the forecast errors inherently exhibit an MA(12) structure. Consequently, the DM test statistics are evaluated using Newey and West (1987) HAC standard errors with an 18-month lag. The raw test statistics and corresponding p-values are reported in Appendix G.

3.4.2 The Failure of Linear Models

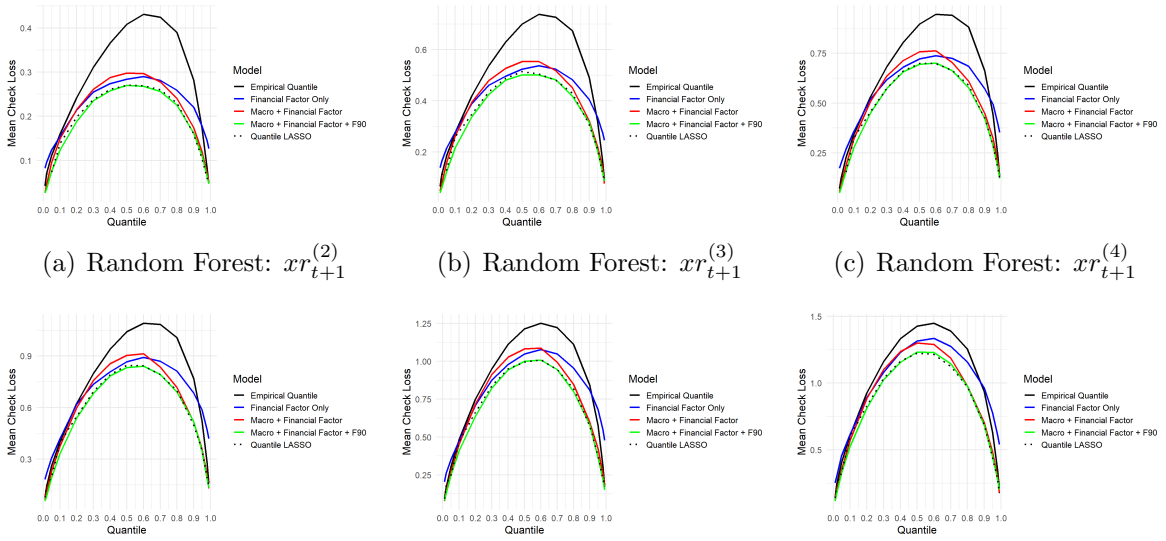
Before detailing the relative performance of the predictor sets, it is necessary to establish the choice of the forecasting algorithm. Standard linear quantile regressions yield highly volatile out-of-sample predictions and generally produce larger check-losses compared to tree-based machine learning models. The full out-of-sample performance tables and plots for the linear models are provided in Appendix E.

In reality, risk premia spike abruptly when macro conditions breach critical thresholds and intermediary constraints bind. Linear models assume a constant marginal effect, leading to severe extrapolation errors during crises.The Quantile Random Forest (QRF) consistently outperforms linear quantile models because bond risk premia react non-linearly to macroeconomic distress—by using recursive decision trees, the QRF inherently captures these sudden regime shifts.

3.4.3 The Value of the Macro-Panic Factor

Evaluating the QRF forecasts reveals distinct dynamics across the predictive distribution. To rigorously assess the out-of-sample contribution of the macroeconomic predictors, Table 3

Figure 1: Mean Check Loss Distribution over Quantiles and Terms (Quantile Random Forest)



(a) Random Forest: $xr_{t+1}^{(2)}$ (b) Random Forest: $xr_{t+1}^{(3)}$ (c) Random Forest: $xr_{t+1}^{(4)}$

(d) Random Forest: $xr_{t+1}^{(5)}$ (e) Random Forest: $xr_{t+1}^{(7)}$ (f) Random Forest: $xr_{t+1}^{(10)}$

Note: Figures above demonstrate the out-of-sample mean check loss over quantiles for Quantile Random Forest (number of trees n=500). Target variables are excess log bond returns of maturity 2 to 10 years.

reports the relative mean check-loss of the two macro-augmented specifications scaled by the benchmark "financial only" model. Cells containing numbers greater than one (marked gray) indicate that the benchmark model yielded a lower forecast loss at that specific quantile.

As demonstrated on the left side of Table 3, when evaluating the "macro + financial" model, the specification significantly outperforms the benchmark at the extreme tails (values well below 1). However, this dominance fails near the central quantiles (ranging from 0.15 to 0.60), where the relative check loss ratios frequently exceed 1. This dynamic is visually apparent in Figure 1, where the "macro + financial" model (solid red line) dips below the "financial only" model (solid blue line) at the tails but often crosses above it near the median. This suggests that while yield curve information is optimal for pricing average conditions, standard macroeconomic level factors alone are insufficient to achieve full predictive dominance across the entire distribution.

Augmenting the model with the upper-tail macro factor, F_t^{90}, significantly transforms this predictive landscape. As shown on the right side of Table 3, the inclusion of this latent

macro-panic" measure—creating the quantile factor augmented" model—achieves full predictive dominance across the entire quantile spectrum and across all bond maturities. Every single relative check loss ratio falls below 1, completely clearing the underperforming gray cells. More importantly, the Diebold-Mariano tests (denoted by asterisks) confirm that this outperformance is highly statistically significant at the extreme tails (e.g., $\tau \leq 0.05$ and $\tau \geq 0.90$), where the macro-panic factor successfully captures distinct tail patterns. Visually, in Figure 1, this is represented by the solid green line uniformly shifting downward, lying cleanly below both the baseline and mean-factor specifications. Furthermore, the overall scaled relative performance indices shrink significantly, with the most prominent improvements concentrated in the extreme lower quantiles of short-term bonds (e.g., dropping from 0.36 to 0.26 at the 1% quantile for $xr_{t+1}^{(2)}$).

Table 3: Relative Check Loss Performance over Quantiles (F_t^{90})
(Quantile Random Forest - Baseline: Financial Only)

	Macro + Financial						**Quantile Factor Augmented**					
Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
0.01	0.36*	0.42**	0.38**	0.35**	0.42**	0.54**	0.26*	0.26**	0.29**	0.33**	0.42**	0.55**
0.02	0.55*	0.47**	0.46**	0.46**	0.60**	0.67**	0.39*	0.38*	0.44*	0.45**	0.51**	0.56**
0.05	0.76**	0.74**	0.72*	0.74*	0.75*	0.82*	0.60**	0.63*	0.61*	0.65*	0.69**	0.76**
0.1	0.92	0.92	0.90	0.89	0.94	0.95	0.80*	0.80*	0.75*	0.79*	0.86	0.89*
0.2	1.00	0.99	0.97	0.99	1.03	0.99	0.87*	0.85*	0.84	0.88	0.95	0.92*
0.3	1.02	1.04	1.05	1.04	1.04	1.01	0.92	0.93	0.94	0.94	0.94	0.93
0.4	1.06	1.06	1.05	1.06	1.03	1.01	0.95	0.96	0.96	0.98	0.94	0.93
0.5	1.05	1.05	1.05	1.05	1.03	1.00	0.95	0.96	0.96	0.97	0.96	0.94*
0.6	1.03	1.04	1.04	1.04	1.01	0.97	0.94**	0.96**	0.96*	0.96*	0.94*	0.92**
0.7	0.99	0.98	0.99	0.96	0.96*	0.92**	0.91***	0.91***	0.92***	0.92***	0.92***	0.87***
0.8	0.93**	0.92**	0.90***	0.89***	0.86***	0.85***	0.89***	0.88***	0.86***	0.85***	0.83***	0.84***
0.9	0.82***	0.81***	0.77***	0.75***	0.75***	0.75***	0.77***	0.78***	0.75***	0.73***	0.75***	0.74***
0.95	0.63***	0.70***	0.65***	0.61***	0.58***	0.64***	0.68***	0.70***	0.64***	0.66***	0.57***	0.60***
0.98	0.51**	0.48***	0.52***	0.55***	0.45***	0.46***	0.48**	0.48***	0.51***	0.44***	0.44***	0.40***
0.99	0.34**	0.35**	0.42***	0.37***	0.32***	0.32***	0.35**	0.38***	0.31***	0.33***	0.37***	0.34***

Note: The table shows the relative out-of-sample mean check loss performance of the two different model specifications ("macro + financial" and "quantile factor augmented") relative to the "financial only" model with Quantile Random Forest (number of trees n=500). Scaling the mean check loss of the "macro + financial" and "quantile factor augmented" model by the mean check loss of the "financial only" model generates numbers in the cells on the left hand and right hand respectively. All numbers have been rounded to two decimal places. The number is greater than 1 (marked gray) if the mean check loss of the scaled model is greater than that of the "financial factor only" model. Asterisks denote the statistical significance of a one-sided Diebold-Mariano test, evaluated using Newey-West standard errors with an 18-month lag, indicating that the target model yields a significantly lower out-of-sample check loss than the "financial only" baseline (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

3.4.4 Q-LASSO vs. Mean-Selected Predictors

A natural methodological question is whether a dynamic, quantile-specific variable selection algorithm (Q-LASSO) outperforms the static "quantile factor augmented" model when both are fed into the QRF. A visual inspection of Figure 1 reveals that the check-loss curve of the Q-LASSO model (dashed black line) and the "quantile factor augmented" model (solid green line) appear nearly overlapping across almost the entire spectrum of quantiles and maturities. However, to rigorously determine which method is superior, Table 4 reports the precise relative out-of-sample check-loss of the Q-LASSO model against the fully augmented specification.

Table 4: Relative Check Loss Performance over Quantiles (Quantile Random Forest – Q-LASSO vs. Quantile Factor Augmented)

Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
0.01	1.56	1.59	1.42	1.37	1.16	1.20
0.02	1.94*	1.80*	1.71	1.73	1.47*	1.39*
0.05	0.97	1.05	0.97	1.02	1.00	1.01
0.10	1.12**	1.20*	1.17*	1.19*	1.15*	1.15**
0.20	1.05***	1.03**	1.02	1.02*	1.04***	1.02**
0.30	1.01**	1.03***	1.00	1.02**	1.02**	1.01
0.40	1.01*	1.01	1.01	1.01	1.01	1.01
0.50	1.00	1.02***	1.01	1.02**	0.99	0.99
0.60	1.00	1.01	1.00	1.00	1.00	0.99
0.70	1.02*	1.00	1.00	1.00	1.00	0.98
0.80	1.02	1.03**	1.02*	1.02	1.02**	0.99
0.90	0.99	0.97	1.00	0.96	0.99	1.00
0.95	0.94	1.03	1.02	1.01	1.02	0.97
0.98	0.94	0.95	0.95	1.03	0.97	1.00
0.99	1.04	1.00	1.03	1.08*	1.15	0.86

Note: Entries report the ratio of out-of-sample mean check loss from the Quantile LASSO model to that of the "quantile factor augmented" model. Values below one indicate that the Quantile LASSO improves upon the "quantile factor augmented" model and are marked gray. Estimation is performed using the Quantile Random Forest (number of trees n=500). Both models utilize solely the base linear predictors to avoid confounding. Asterisks denote the statistical significance of a one-sided Diebold-Mariano test, evaluated using Newey-West standard errors with an 18-month lag, indicating that the "quantile factor augmented" QRF model yields a significantly lower out-of-sample check loss than the Q-LASSO baseline (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

The results demonstrate that forcing non-linear market shocks into sparse, linear pre-selections via Q-LASSO generally degrades out-of-sample performance, even when non-linear confounding terms are excluded from the selection pool. Despite their visual overlap in the plots, across the vast majority of maturities and quantiles, the precise relative ratios are greater than one, with the Diebold-Mariano tests confirming that this underperformance is frequently statistically significant at the extremes (raw statistics can be found in Appendix G). Since the QRF natively performs feature selection and handles redundancy through random

subspace splitting, providing it with the robust baseline of the "quantile factor augmented" model proves more computationally efficient and predicatively superior than relying on the linear Q-LASSO algorithm.

3.4.5 Term-Structure Predictability Decay

While the "quantile factor augmented" model dominates at the short end of the yield curve, an important term-structure duality emerges as bond maturity increases. This decay in overall predictability is visually evident across the subpanels of Figure 1. Tracking the forecast performance from the 2-year maturity (subfigure 2(a)) to the 10-year maturity (subfigure 2(f)), the absolute level of mean check-loss scales upward dramatically across all model specifications. Consequently, the relative performance gap between the sophisticated QRF models and the simple Empirical Quantile (EQ) narrows significantly for long-term bonds. The EQ, despite being a naive unconditional historical baseline, becomes a highly competitive and robust forecasting alternative at the 10-year horizon.

This reflects a decay in macroeconomic predictability along the yield curve, which is in line with Ang and Piazzesi (2003); Cieslak and Povala (2015); Bianchi et al. (2021). Short-term tail risks are tightly anchored to current macro-policy cycles and acute liquidity panics, making them highly predictable using current macroeconomic data. Conversely, long-term risk premia are driven by structural safe-haven demand and distant, unforecastable shocks, such as shifts in global savings gluts or long-run inflation expectations. At these longer horizons, current macro data contains very little predictive signal and the unconditional EQ somehow effectively anchors to long-run structural realities.

4 Cyclicality and Term Structure Dynamics

Traditional macroeconomic studies frequently document that average bond risk premia fluctuate countercyclically with the broader economy (Bansal and Yaron, 2004; Ludvigson and

Ng, 2009). During economic downturns, aggregate risk aversion rises, causing the expected compensation for holding duration risk to increase. However, by shifting the forecasting target from the conditional mean to the extreme quantiles, a much richer, bidirectional narrative emerges. Severe economic shocks do not merely shift the average risk premium, they violently alter its distribution, triggering simultaneous but opposing market mechanisms. Accordingly, this section analyzes the cyclical behavior of the out-of-sample Quantile Random Forest (QRF) forecasts to disentangle these distinct crisis dynamics.

4.1 The Dual Nature of Crisis Pricing and Asymmetric Shocks

Figure 2 and Figure 3 juxtapose the standardized 12-month moving averages of the QRF bond risk premia forecasts at the lower tail ($\tau = 0.05$) and upper tail ($\tau = 0.95$) against Industrial Production (IP) growth. A striking regime-dependent pattern immediately emerges across both extremes: the forecasts exhibit mildly procyclical or acyclical behavior during economic expansions, but pivot to become considerably countercyclical during recessions.

At first glance, it may appear contradictory that both the extreme low forecasts and the extreme high forecasts are negatively correlated with IP. However, this illustrates the difference between a general mean-shift in risk pricing and shock-induced distributional widening. During a prolonged economic contraction, such as the 2008 Great Recession, the explosion of aggregate risk aversion potentially pushes the entire predictive distribution of bond risk premia upward. This universal upward shift mathematically dominates the correlation metrics, resulting in a strongly negative correlation with IP for both tails over the full sample.

However, the true duality of the tails is revealed during sudden, violent shocks—most notably the March 2020 crash. Unlike the prolonged 2008 crisis, the brief but severe 2020 recession triggered a highly asymmetric and sequential response. As Industrial Production plummeted, the upper-tail forecast ($\tau = 0.95$) moved sharply higher immediately, reflecting a liquidity-driven "dash for cash" where intermediary constraints bound and the maximum required premium spiked. Simultaneously, the lower-tail forecast ($\tau = 0.05$) remained remarkably

anchored, showing very little immediate downward movement. This dynamic suggests that during an abrupt, severe shock, the acute phase of the panic is primarily led by a scramble for liquidity rather than an immediate flock into safe-haven.

Furthermore, this initial shock catalyzed a massive widening of the gap between the two tails in the post-2020 period. Following the immediate liquidity panic, the lower-tail forecast ($\tau = 0.05$) eventually plunged deep into negative territory. This delayed plunge reflects a massive, subsequent "flight to safety" alongside the overwhelming force of central bank interventions and quantitative easing. By forecasting the entire distribution, the QRF model successfully captures this modern shock-induced timeline: an initial liquidity panic that pulls the upper bound higher, followed by a delayed safe-haven scramble that drastically widens the overall distribution of expected returns, a nuance that linear mean-models miss entirely.

4.2 Decay and Dominance Along the Term Structure

Beyond general countercyclicality, comparing the 5th and 95th quantiles reveals a divergence in how macroeconomic shocks propagate across the yield curve. As the maturity of the bond grows, the countercyclical correlations of the two tails move in opposite directions, directly informing the relative importance of liquidity versus duration shocks.

The upper-tail "dash for cash" is a short-end phenomenon. When intermediary constraints bind and macro-panic sets in, investors liquidate assets treated as cash-equivalents, which are primarily short-to-medium-term Treasuries. Because this liquidity premium is most acute at shorter maturities, the $\tau = 0.95$ forecast is highly sensitive to macroeconomic distress at the 2-year bond, exhibiting a strong recessionary correlation of -0.604 (Figure 4(a)). However, as maturity increases to the 10-year bond, distant structural forces dilute this immediate liquidity panic, causing the countercyclical correlation to weaken by a quarter to -0.465 (Figure 4(f)).

Conversely, the lower-tail "flight to safety" is essentially a duration play. When the real economy stalls and investors anticipate aggressive central bank rate cuts, they crowd into

longer-term bonds to lock in yields and maximize capital gains. Because the long end of the curve serves as, up to now, the ultimate safe haven, the compression of the term premium ($\tau = 0.05$) becomes more tightly coupled to the real economic downturn as maturity increases. Consequently, the recessionary correlation of the lower tail strengthens dramatically as we move along the term structure: shifting from -0.479 at the 2-year horizon (Figure 3(a)) to a dominant -0.766 at the 10-year horizon (Figure 3(f)). Ultimately, these contrasting term structure dynamics firmly indicates that short-term tail risks are driven by acute macro-liquidity panics, while long-term tail risks are heavily anchored by structural safe-haven demand.

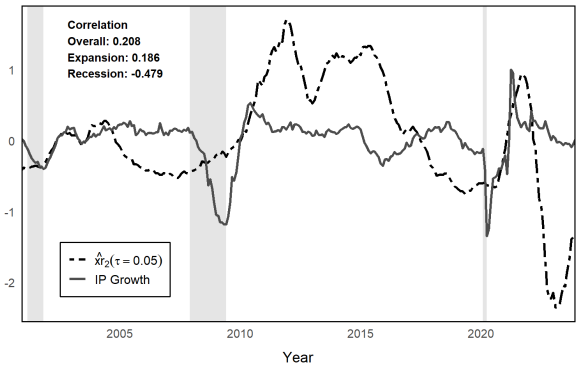
5 Robustness Check: Prior to COVID-19

To ensure that the out-of-sample forecasting results are not uniquely driven by the extreme, exogenous shock of the COVID-19 pandemic, I conduct a robustness check by truncating the evaluation period to end in December 2019, with the corresponding plots and tables provided in Appendix F. The results of this pre-COVID evaluation yield several important insights regarding both model performance and the economic nature of tail risks.

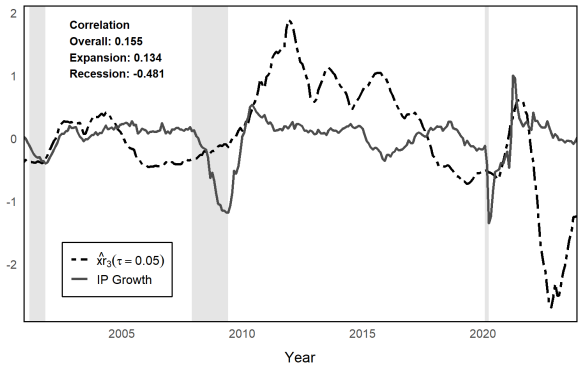
First, comparing the overall mean check loss in the truncated sample (Figure 5) to the full sample reveals that absolute forecast errors are noticeably smaller prior to 2020. Because the COVID-19 shock, unlike a standard business cycle unwinding, was an unprecedented black-swan event that drastically inflated market volatility and forecast errors universally, removing it underscores the baseline predictability of the models. Furthermore, the performance gap between the model-based forecasts and the naive Empirical Quantile widens significantly in the pre-COVID sample. This indicates that standard macro-financial variables contained a stronger, more reliable predictive signal during conventional business cycles, whereas the exogenous and sudden nature of the pandemic shock challenged standard predictive frameworks.

Second, the performance of the "quantile factor augmented" model remains highly robust at

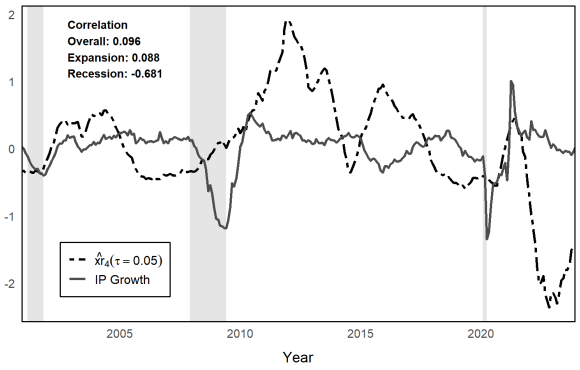
Figure 2: Bond Risk Premium Forecast ($\tau = 0.05$) vs. IP Growth (standardized)



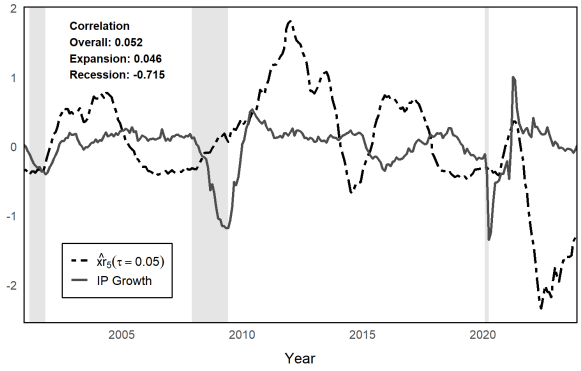
(a) $\hat{xr}^{(2)}_{t+1}(\tau = 0.05)$ vs. IP Growth



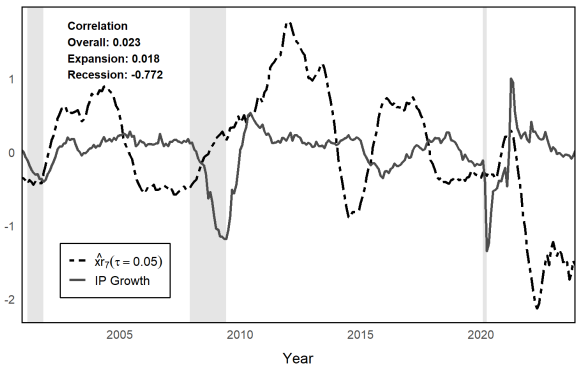
(b) $\hat{xr}^{(3)}_{t+1}(\tau = 0.05)$ vs. IP Growth



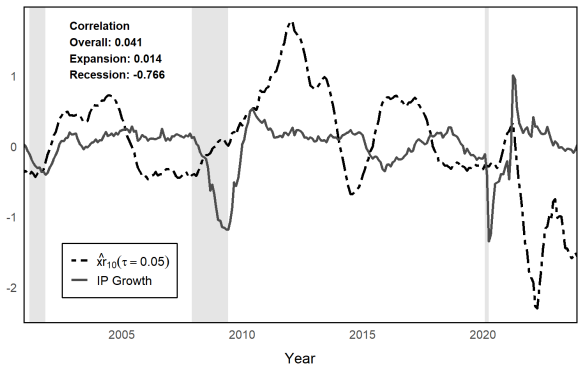
(c) $\hat{xr}^{(4)}_{t+1}(\tau = 0.05)$ vs. IP Growth



(d) $\hat{xr}^{(5)}_{t+1}(\tau = 0.05)$ vs. IP Growth



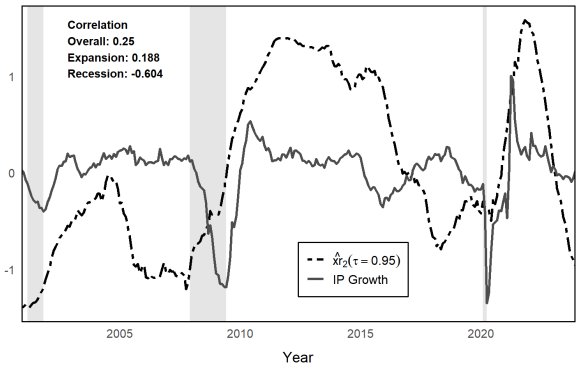
(e) $\hat{xr}^{(7)}_{t+1}(\tau = 0.05)$ vs. IP Growth



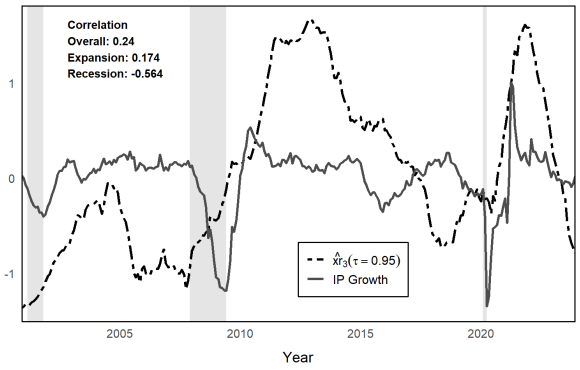
(f) $\hat{xr}^{(10)}_{t+1}(\tau = 0.05)$ vs. IP Growth

Note: Figures above demonstrate the out-of-sample predictions of excess log bond returns of maturity 2 to 10 years at $\tau = 0.05$ (generated by Quantile Random Forest) with the quantile factor model specification versus same period industrial production (IP) growth, the correlations of the two time series shown on the top left corner of each subfigure. The plots use 12-month moving averages of the standardized predictions and IP growth to make visual comparison, and the shaded areas denote recessions periods according to the National Bureau of Economic Research.

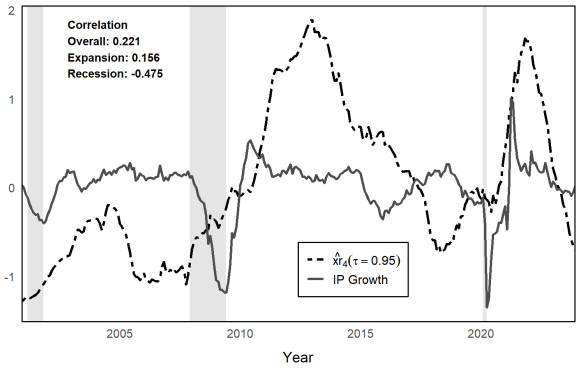
Figure 3: Bond Risk Premium Forecast ($\tau = 0.95$) vs. IP Growth (standardized)



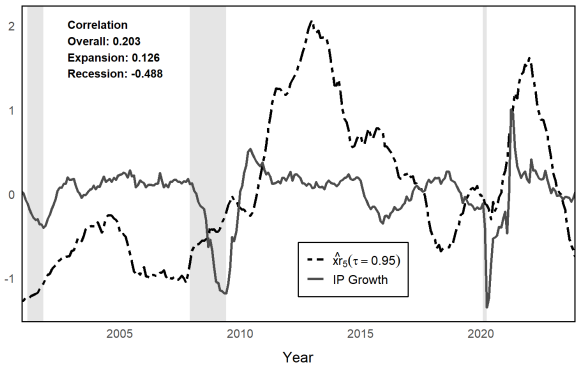
(a) $\hat{xr}^{(2)}_{t+1}(\tau = 0.95)$ vs. IP Growth



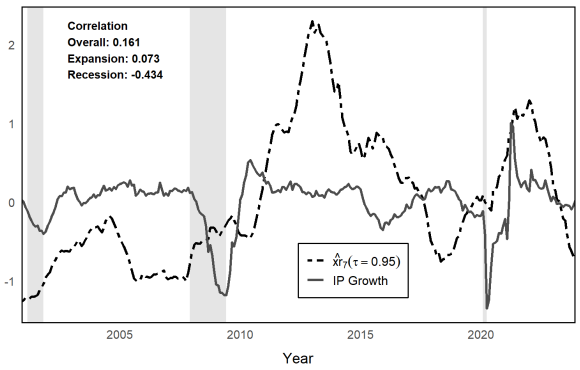
(b) $\hat{xr}^{(3)}_{t+1}(\tau = 0.95)$ vs. IP Growth



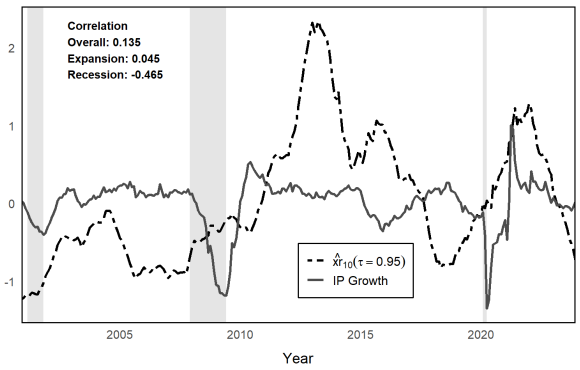
(c) $\hat{xr}^{(4)}_{t+1}(\tau = 0.95)$ vs. IP Growth



(d) $\hat{xr}^{(5)}_{t+1}(\tau = 0.95)$ vs. IP Growth



(e) $\hat{xr}^{(7)}_{t+1}(\tau = 0.95)$ vs. IP Growth



(f) $\hat{xr}^{(10)}_{t+1}(\tau = 0.95)$ vs. IP Growth

Note: Figures above demonstrate the out-of-sample predictions of excess log bond returns of maturity 2 to 10 years at $\tau = 0.95$ (generated by Quantile Random Forest) with the quantile factor model specification versus same period industrial production (IP) growth, the correlations of the two time series shown on the top left corner of each subfigure. The plots use 12-month moving averages of the standardized predictions and IP growth to make visual comparison, and the shaded areas denote recessions periods according to the National Bureau of Economic Research.

the extreme tails, though it exhibits negligible underperformance relative to the "financial only" baseline at central quantiles, as seen in the right panel of Table 10. It is important to clarify the mechanics behind this relative ratio: incorporating the extreme macro-panic factor (F_t^{90}) actually lowers the absolute check loss across the entire quantile spectrum.

Interestingly, excluding the COVID-19 period clarifies the distinct economic mechanisms governing the upper and lower tails. As shown in Table 10, the yield-curve-based "financial only" model performs much better at predicting the lower tail than the upper tail. This asymmetry aligns with economic theory: the yield curve aggregates expectations of future monetary policy and growth, making it highly informative for structural, slow-moving "flight-to-safety" regimes. Conversely, upper-tail "dash-for-cash" events are often driven by sudden liquidity frictions, margin calls, and dealer constraints (He et al., 2022), which are inherently black-swan events that the shape of the yield curve cannot fully anticipate.

Finally, an analysis of the standardized forecast correlations with Industrial Production (IP) growth (Figures 7 and 8) reveals how the pandemic distorted specific term-structure dynamics. In the pre-COVID sample, the $\tau = 0.05$ (flight-to-safety) forecasts exhibit a remarkably uniform and massive negative correlation with IP during recessions across all maturities (e.g., -0.942 at the 2-year and -0.871 at the 10-year). However, when the COVID-19 period is included, a stark divergence emerges: the negative correlation almost halved for short-term bonds but only weakened slightly in magnitude for long-term bonds. This maturity-specific distortion is well-documented in the literature surrounding the March 2020 market intervention. By instantly slashing the federal funds rate to the zero lower bound (ZLB) and issuing aggressive forward guidance, the Federal Reserve effectively pinned short-term yields. As demonstrated by Swanson and Williams (2014) and Vissing-Jorgensen (2021), reaching the ZLB causes short-term risk premia to become highly compressed and detached from variations in real macroeconomic output. Long-term bonds, while still affected by the overarching market interventions, retained more of their structural risk profile, allowing them to maintain a stronger (albeit slightly dampened) covariation with the extreme collapse in physical output, thus absorbing safe-haven flows and structural uncertainties of the pandemic shock (Duffie, 2020; Haddad et al., 2021).

6 Conclusion

For decades, the empirical asset pricing literature has relied on mean-targeted models to forecast bond risk premia, implicitly assuming that excess returns respond linearly to the business cycle. This paper demonstrates that such an approach potentially masks the true nature of tail risk in the bond market. By moving beyond the conditional mean to forecast the entire distribution of excess returns, I show that crisis dynamics are governed by threshold effects and sudden financial frictions that traditional linear models systematically overlook.

The central empirical contribution of this study is the identification of a latent, data-driven "macro-panic" factor. Extracted from the upper tail (F_t^{90}) of a large macroeconomic panel, this factor tracks the specific environments where systemic volatility rises and intermediary constraints tend to bind. When combined with standard yield and level-macro predictors, this macro-panic factor significantly enhances predictability across the entire quantile spectrum. Methodologically, I demonstrate that the Quantile Random Forest (QRF) natively captures these non-linear regime shifts and crisis thresholds. As a result, the QRF circumvents the need for preliminary, in-sample variable selection via Quantile LASSO—a procedure that is often time-consuming and generates unstable model selection. Instead, pairing the QRF directly with standard, mean-selected predictors can efficiently and robustly forecast severe market distress.

Furthermore, the analysis reveals that risk pricing is highly state-dependent, exhibiting distinct regime-dependent cyclicality that alters from mildly cyclical during expansions to heavily countercyclical during recessions. Crucially, forecasting the extreme tails uncovers a divergence in how modern macroeconomic shocks propagate across the yield curve. Results suggest that upper-tail risks at the short end of the curve are likely driven by acute macro-liquidity panics (a "dash for cash"), whereas lower-tail risks at the long end are heavily anchored by structural safe-haven demand (a "flight to safety"). As these liquidity panics dilute over longer horizons, a decay in macroeconomic predictability occurs. While the conditional machine learning model dominates at shorter maturities, naive empirical quantiles become highly competitive at the 10-year horizon, illustrating how transient macro-financial

signals eventually give way to unforecastable structural shocks.

Finally, by tracking the entire predictive distribution during abrupt market breaks like the March 2020 crash, this paper reveals the sequential nature of modern financial crises. Rather than a simultaneous bidirectional movement, sudden severe shocks tend to be initially led by a scramble for liquidity that forces the upper bound of required premia higher. This is followed by a delayed flock into safe-haven at the lower bound, drastically widening the overall distribution of expected returns. While this paper establishes the predictability and term-structure dynamics of these extreme tail movements, uncovering how this shock-induced distributional widening subsequently transmits back into the real economy remains a critical avenue for future research, the resolution of which will further clarify the macro-finance link and aid policymakers during structural market breaks.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–1289.

Adrian, T., R. K. Crump, and E. Moench (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics 110*(1), 110–138.

Adrian, T. and N. Liang (2016). Monetary policy, financial conditions, and financial stability.

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics 50*(4), 745–787.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance 59*(4), 1481–1509.

Bauer, M. D. and J. D. Hamilton (2018). Robust bond risk premia. *The Review of Financial Studies 31*(2), 399–448.

Belloni, A. and V. Chernozhukov (2011). L1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*(1), 82–130.

Bernanke, B. (2013). Long-term interest rates - speech given at the annual macroecoonomics conference. *Federal Reserve Bank of San Francisco*.

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premiums with machine learning. *The Review of Financial Studies 34*(2), 1046–1089.

Brainard, L. (2021, March). Some preliminary financial stability lessons from the COVID-19 shock. Speech at the 2021 Annual Washington Conference, Institute of International Bankers (via webcast). Board of Governors of the Federal Reserve System.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market liquidity and funding liquidity. *The Review of Financial Studies 22*(6), 2201–2238.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), 205–251.

Chauvet, M. and Z. Senyuz (2016). A dynamic factor model of the yield curve components as a predictor of the economy. *International Journal of Forecasting 32*(2), 324–343.

Chen, L., J. J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910.

Chen, L., M. Pelger, and J. Zhu (2024). Deep learning in asset pricing. *Management Science 70*(2), 714–750.

Cieslak, A. and P. Povala (2015). Expected returns in treasury bonds. *The Review of Financial Studies 28*(10), 2859–2901.

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Review 95*(1), 138–160.

Duffie, D. (2020). Still the world's safe haven? redesigning the u.s. treasury market after the covid-19 crisis. Technical report, Hutchins Center Working Paper No. 62, Brookings Institution.

Eriksen, J. N. (2017). Expected business conditions and bond risk premia. *Journal of Financial and Quantitative Analysis 52*(4), 1667–1703.

Estrella, A. and F. S. Mishkin (1998). Predicting us recessions: Financial variables as leading indicators. *Review of Economics and Statistics 80*(1), 45–61.

Gupta, R., C. Pierdzioch, A. J. Vivian, and M. E. Wohar (2019). The predictive value of inequality measures for stock returns: An analysis of long-span uk data using quantile random forests. *Finance Research Letters 29*, 315–322.

Haddad, V., A. Moreira, and T. Muir (2021). When selling becomes a viral infection: Disruptions in debt markets in the covid-19 crisis and the fed's response. *The Review of Financial Studies 34*(11), 5309–5351.

Hamilton, J. D. and D. H. Kim (2002). A re-examination of the predictability of the yield spread for real economic activity. *Journal of Money, Credit, and Banking 34*(2), 340–360.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *The American Economic Review 103*(2), 732–770.

He, Z., S. Nagel, and Z. Song (2022). Treasury inconvenience yields during the covid-19 crisis. *Journal of Financial Economics 143*(1), 57–79.

Ireland, P. N. (2015). Monetary policy, bond risk premia, and the economy. *Journal of Monetary Economics 76*, 124–140.

Liu, Y. and J. C. Wu (2021). Reconstructing the yield curve. *Journal of Financial Economics 142*(3), 1395–1425.

Longstaff, F. A. (2002). The flight-to-liquidity premium in us treasury bond prices.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

Luong, C. and N. Dokuchaev (2018). Forecasting of realised volatility with the random forests algorithm. *Journal of Risk and Financial Management 11*(4), 61.

McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research 7*(6).

Nagel, S. (2016). The liquidity premium of near-money assets. *The Quarterly Journal of Economics 131*(4), 1927–1971.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica 55*(3), 703–708.

Rudebusch, G. D., B. P. Sack, and E. T. Swanson (2006). Macroeconomic implications of changes in the term premium. *FRB of San Francisco Working Paper* (2006-46).

Rudebusch, G. D. and J. C. Williams (2009). Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business & Economic Statistics 27*(4), 492–503.

Swanson, E. T. and J. C. Williams (2014). Measuring the effect of the zero lower bound on medium- and longer-term interest rates. *American Economic Review 104*(10), 3154–3185.

Vissing-Jorgensen, A. (2021). The treasury market in spring 2020 and the response of the federal reserve. *Journal of Monetary Economics 124*, 19–47.

Waller, C. J. (2024). Thoughts on quantitative tightening, including remarks on the paper "quantitative tightening around the globe: What have we learned?". Remarks at the 2024 U.S. Monetary Policy Forum, New York, New York. Board of Governors of the Federal Reserve System.

Wright, J. H. (2006). The yield curve and predicting recessions. *Finance and Economics Discussion Series, FRB Washington, D.C.*.

A Macroeconomic Data and Transformations

This appendix details the 126 macroeconomic variables from the FRED-MD dataset used in the empirical analysis, along with their respective stationary transformations.

Table 5: Policy Relevant Variables and Transformation Scheme

Series	Abbreviation	Transformation
Industrial Production	IP	$\Delta log(x_t)$
Real Personal Income	RPI	$\Delta log(x_t)$
Real Personal Income ex Transfer Receipt	RPIx	$\Delta log(x_t)$
Real Manufacturing and Trade Industries Sales	MTS	$\Delta log(x_t)$
Unfilled Orders for Durable Goods	UO	$\Delta log(x_t)$
Retail and Food Services Sales	RFS	$\Delta log(x_t)$
M2 Monetary Stock	M2	$\Delta^2 log(x_t)$
10-Year Treasury Rte	TB	Δx_t
Consumer Price Index	CPI	$\Delta^2 log(x_t)$
Consumer Price Index ex Food	CPIx	$\Delta^2 log(x_t)$
S&P 500	SNP	$\Delta log(x_t)$

B Selected Factors' Correlation with FRED-MD

Table 6: Top 8 variables for Selected Factors by In-Sample Absolute Correlation

	1	2	3	4	5	6	7	8
PC1	USGOOD 0.882	PAYEMS 0.858	IPMANSICS 0.858	MANEMP 0.855	INDPRO 0.853	CUMFNS 0.837	DMANEMP 0.835	IPDMAT 0.811
PC2	BAAFFM 0.738	AAAFFM 0.712	T10YFFM 0.686	T5YFFM 0.674	BAA 0.603	TB6SMFFM 0.558	AAA 0.546	T1YFFM 0.545
PC3	CUSR0000SAC 0.693	CUSR0000SA0L2 0.691	CPIAUCSL 0.677	CUSR0000SA0L5 0.675	DNDGRG3M 0.675	PCEPI 0.647	CPIULFSL 0.525	CPITRNSL 0.472
PC4	GS1 0.500	TB6MS 0.479	AAA 0.477	GS5 0.476	BAA 0.467	GS10 0.447	TB3MS 0.437	PERMITNE 0.432
PC5	T1YFFM 0.544	TB6SMFFM 0.544	TB3SMFFM 0.528	T5YFFM 0.468	GS10 0.462	GS5 0.461	T10YFFM 0.459	GS1 0.454
PC6	S&P PE 0.450	S&P 500 0.443	S&P div 0.441	DPCERA3M 0.431	RETAILx 0.400	TOTRESNS 0.399	UEMPMEAN 0.379	NONBORRES 0.352
PC10	PAYEMS 0.694	USGOOD 0.687	MANEMP 0.672	INDPRO 0.658	DMANEMP 0.650	PERMIT 0.646	IPMANSICS 0.639	IPMAT 0.635
PC90	HOUST 0.499	PERMIT 0.472	INDPRO 0.470	USTRADE 0.467	IPMANSICS 0.465	HOUSTS 0.465	PAYEMS 0.455	CUMFNS 0.454

C Number of Extracted Quantile Factors

The optimal number of quantile factors extracted from the macroeconomic panel varies across the quantile spectrum.

Table 7: Number of Selected Quantile Factors

τ	**Number of Quantile Factors**
0.01	1
0.05	1
0.10	1
0.25	4
0.50	5
0.75	4
0.90	2
0.95	1
0.99	1

Note: The number of quantile factors indicate the number of diagonal entries of the $L' \cdot L$ that is greater than the first entry (times a constant that grows with N) — rank minimization method. L is the $N \times r$ loading matrix of the quantile factor of the corresponding τ.

D Quantile LASSO Selected Variables

The following Table 8 detail the specific macro-financial variables selected by the Quantile LASSO procedure across different quantiles for the benchmark 2-year excess bond return and the longer term 10-year excess bond return as comparison.

Table 8: Quantile-Specific Model Selection: Target $xr_{t+1}^{(2)}$

Quantile	Selected Variables
0.01	Intercept only
0.02	F_{1t}, F_{4t}, F_{6t}
0.05	CP_t, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{90}
0.10	CP_t, F_{3t}, F_{4t}, F_{6t}
0.20	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{90}
0.30	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{90}
0.40	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{90}
0.50	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{90}
0.60	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{10}, F_t^{90}
0.70	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{90}
0.80	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{90}
0.90	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{10}, F_t^{90}
0.95	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{90}
0.98	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{10}, F_t^{90}
0.99	CP_t, F_{3t}, F_{4t}, F_{6t}, F_t^{10}, F_t^{90}

Notes: The table reports selected variables based on 100 repetitions of 10-fold cross-validated Quantile LASSO. Variables listed are selected in at least 50 out of 100 repetitions. Target variables are the conditional quantiles of $xr_{t+1}^{(2)}$. The scalar financial factor is shown in blue, level macro factors in black, and lower- and upper-tail macro quantile factors in red.

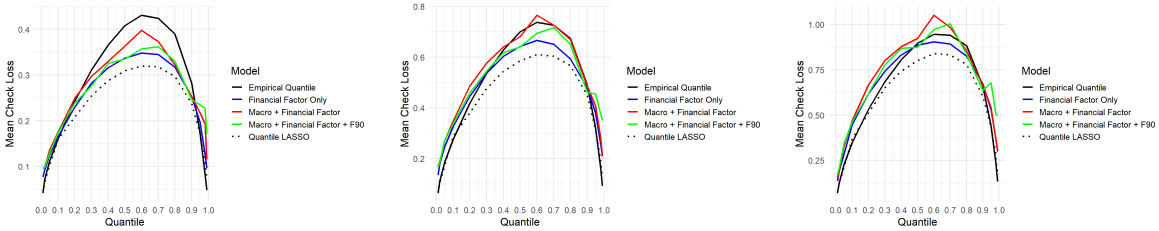
Table 9: Quantile-Specific Model Selection: Target $xr_{t+1}^{(10)}$

Quantile	Selected Variables
0.01	F_t^{90}
0.02	CP_t, F_{6t}
0.05	CP_t, F_{6t}, F_t^{10}
0.10	CP_t, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{90}
0.20	CP_t, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{10}, F_t^{90}
0.30	CP_t, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{10}, F_t^{90}
0.40	CP_t, F_{2t}, F_{3t}, F_{4t}, F_{6t}, F_t^{10}, F_t^{90}
0.50	CP_t, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{10}, F_t^{90}
0.60	CP_t, F_{1t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{10}, F_t^{90}
0.70	CP_t, F_{1t}, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{10}, F_t^{90}
0.80	CP_t, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{90}
0.90	CP_t, F_{2t}, F_{3t}, F_{4t}, F_{5t}, F_{6t}, F_t^{10}, F_t^{90}
0.95	CP_t, F_{2t}, F_{6t}, F_t^{10}
0.98	CP_t, F_{2t}, F_{6t}, F_t^{10}
0.99	CP_t, F_{2t}, F_{6t}, F_t^{10}

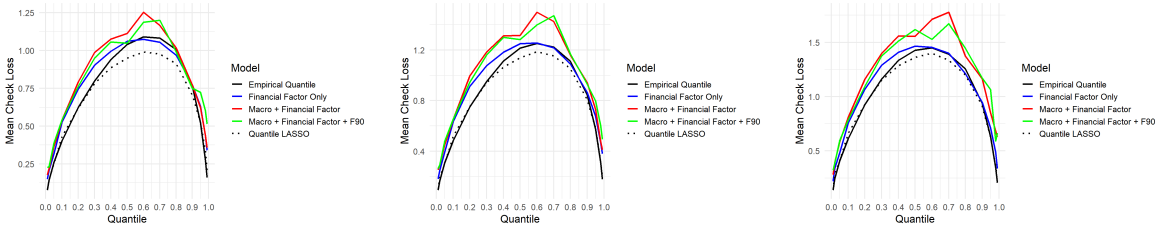
Notes: The table reports selected variables based on 100 repetitions of 10-fold cross-validated Quantile LASSO. Variables listed are selected in at least 50 out of 100 repetitions. Target variables are the conditional quantiles of $xr_{t+1}^{(10)}$. The scalar financial factor is shown in blue, level macro factors in black, and lower- and upper-tail macro quantile factors in red.

E Plots for Linear Quantile Forecast Check Loss

Figure 4: Mean Check Loss Distribution over Quantiles and Terms (Linear Quantile Forecast)



(a) Linear Model: $xr_{t+1}^{(2)}$ (b) Linear Model: $xr_{t+1}^{(3)}$ (c) Linear Model: $xr_{t+1}^{(4)}$



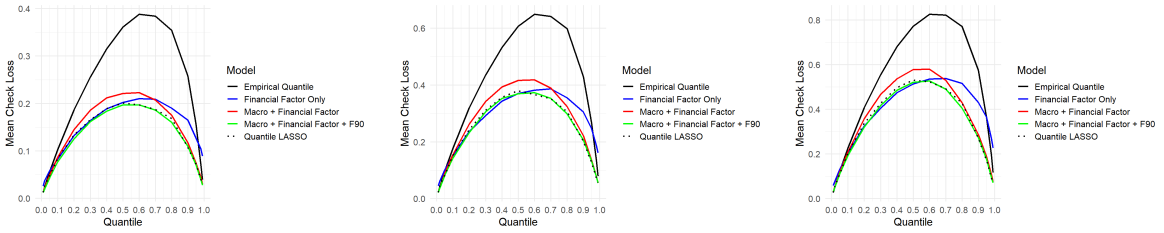
(d) Linear Model: $xr_{t+1}^{(5)}$ (e) Linear Model: $xr_{t+1}^{(7)}$ (f) Linear Model: $xr_{t+1}^{(10)}$

Note: Figures above demonstrate the out-of-sample mean check loss over quantiles for linear quantile regression. Target variables are excess log bond returns of maturity 2 to 10 years.

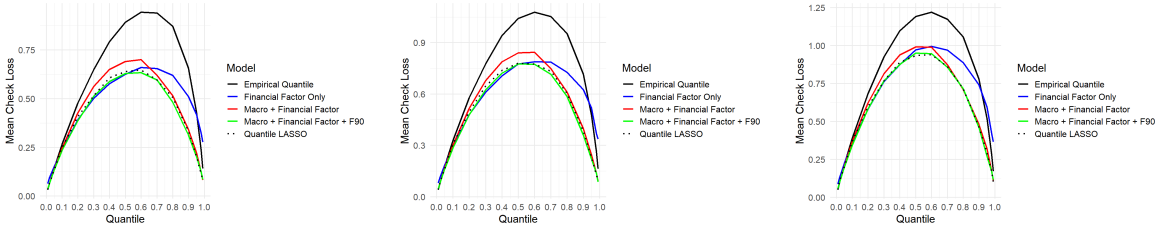
F Robustness Check - Forecast Excluding COVID-19

This appendix displays the main results of the article but with out-of-sample period up to 2019:12 – strictly excludes the COVID period. Figure 5 shows the mean check loss distribution of Quantile Random Forest forecast while Figure 6 demonstrates that of linear quantile forecasts. Table 10 and 11 are robust-sample counterparts of 3 and 4.

Figure 5: Mean Check Loss Distribution over Quantiles and Terms (Quantile Random Forest - COVID Excluded)



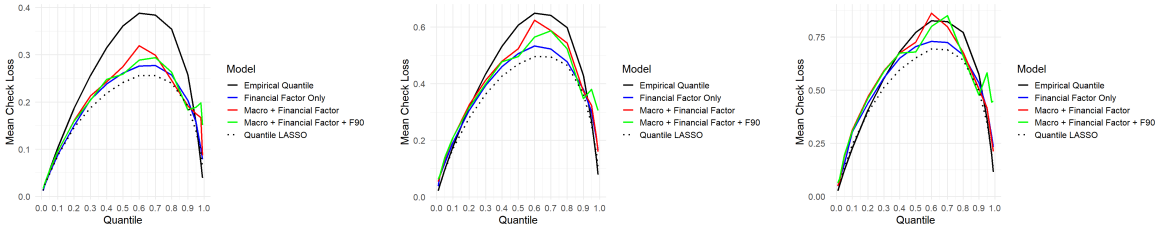
(a) Random Forest: $xr_{t+1}^{(2)}$ (b) Random Forest: $xr_{t+1}^{(3)}$ (c) Random Forest: $xr_{t+1}^{(4)}$



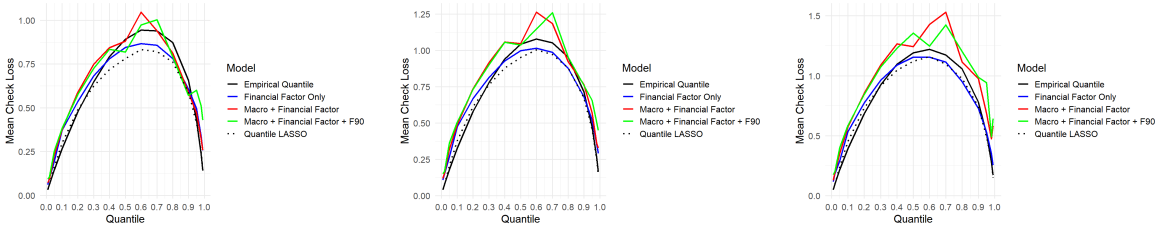
(d) Random Forest: $xr_{t+1}^{(5)}$ (e) Random Forest: $xr_{t+1}^{(7)}$ (f) Random Forest: $xr_{t+1}^{(10)}$

Note: Figures above demonstrate the out-of-sample mean check loss over quantiles for Quantile Random Forest (number of trees n=500). Target variables are excess log bond returns of maturity 2 to 10 years. The forecast period is up to 2019:12.

Figure 6: Mean Check Loss Distribution over Quantiles and Terms (Linear Quantile Forecast - COVID Excluded)



(a) Linear Model: $xr_{t+1}^{(2)}$ (b) Linear Model: $xr_{t+1}^{(3)}$ (c) Linear Model: $xr_{t+1}^{(4)}$



(d) Linear Model: $xr_{t+1}^{(5)}$ (e) Linear Model: $xr_{t+1}^{(7)}$ (f) Linear Model: $xr_{t+1}^{(10)}$

Note: Figures above demonstrate the out-of-sample mean check loss over quantiles for linear quantile regression. Target variables are excess log bond returns of maturity 2 to 10 years. The forecast period is up to 2019:12.

Table 10: Relative Check Loss Performance over Quantiles (F_t^{90})
(Quantile Random Forest - Baseline: Financial Only)

	Macro + Financial						Quantile Factor Augmented					
Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
0.01	0.50**	0.47**	0.52**	0.58**	0.57**	0.64**	0.46**	0.47**	0.50**	0.56**	0.56***	0.65**
0.02	0.61**	0.63**	0.70**	0.68**	0.71**	0.76**	0.53**	0.61**	0.66**	0.68**	0.69**	0.76**
0.05	0.84	0.88	0.91	0.92	0.93	0.92	0.74*	0.81*	0.85*	0.89	0.88*	0.89*
0.1	1.05	1.02	1.02	1.03	1.02	1.03	0.93	0.93	0.95	0.98	0.97	0.96
0.2	1.10	1.11	1.10	1.09	1.07	1.07	0.95	0.98	0.99	0.98	0.99	1.00
0.3	1.13	1.17	1.15	1.12	1.11	1.07	0.99	1.04	1.03	1.02	1.02	1.01
0.4	1.12	1.15	1.13	1.13	1.12	1.07	0.97	1.03	1.02	1.02	1.02	1.00
0.5	1.09	1.12	1.12	1.11	1.08	1.02	0.98	1.00	1.01	1.01	1.00	0.98
0.6	1.06	1.09	1.08	1.06	1.07	1.00	0.94*	0.98	0.99	0.96	0.98	0.95
0.7	0.99	1.01	0.98	0.95	0.95	0.90*	0.89**	0.91**	0.91**	0.91*	0.91**	0.89**
0.8	0.92*	0.91*	0.83***	0.83***	0.84**	0.80***	0.84***	0.84***	0.78***	0.78***	0.81***	0.81***
0.9	0.71***	0.73***	0.66***	0.67***	0.64***	0.65***	0.67***	0.68***	0.63***	0.61***	0.57***	0.62***
0.95	0.58***	0.55***	0.52***	0.53***	0.47***	0.54***	0.56***	0.54***	0.49***	0.48***	0.43***	0.46***
0.98	0.43***	0.42***	0.40***	0.38***	0.37***	0.41***	0.38***	0.41***	0.35***	0.39***	0.38***	0.40***
0.99	0.34***	0.35***	0.33***	0.30***	0.27***	0.30***	0.31***	0.34***	0.31***	0.31***	0.26***	0.32***

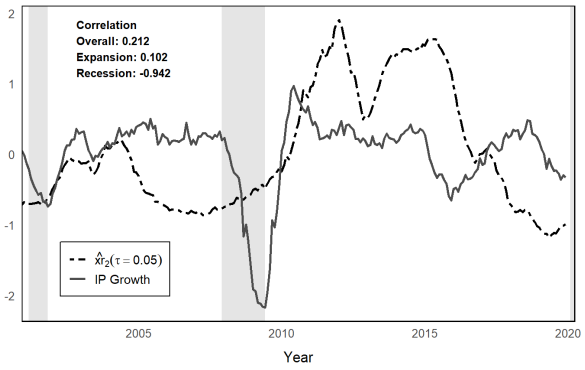
Note: The table shows the relative out-of-sample mean check loss performance of the two different model specifications ("macro + financial" and "quantile factor augmented") relative to the "financial only" model with Quantile Random Forest (number of trees n=500). Scaling the mean check loss of the "macro + financial" and "quantile factor augmented" model by the mean check loss of the "financial only" model generates numbers in the cells on the left hand and right hand respectively. All numbers have been rounded to two decimal places. The number is greater than 1 (marked gray) if the mean check loss of the scaled model is greater than that of the "financial factor only" model. Asterisks denote the statistical significance of a one-sided Diebold-Mariano test, evaluated using Newey-West standard errors with an 18-month lag, indicating that the target model yields a significantly lower out-of-sample check loss than the "financial only" baseline (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Out-of-sample up to 2019:12.

Table 11: Relative Check Loss Performance over Quantiles
(Quantile Random Forest – Q-LASSO vs. Quantile Factor Augmented)

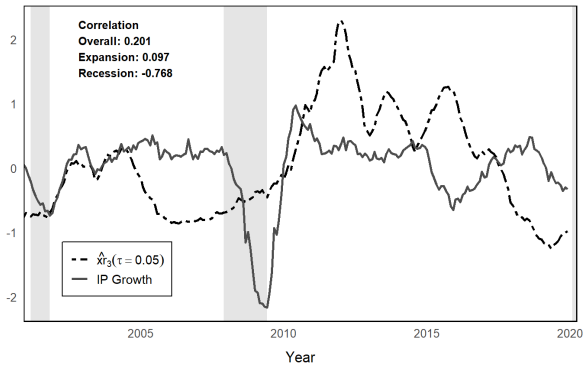
Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
0.01	1.08	1.05	0.90	0.89	0.89	0.86
0.02	1.35***	1.24*	1.10	1.02	1.06	1.08
0.05	1.09***	1.01	1.00	1.02	1.02	1.01
0.10	1.09**	1.08**	1.06*	1.06**	1.08**	1.07**
0.20	1.07***	1.05**	1.05***	1.06***	1.04**	1.03**
0.30	1.02**	1.02*	1.02*	1.00	1.04***	1.00
0.40	1.03*	1.01*	1.03**	1.03***	1.02**	1.01
0.50	1.02*	1.03**	1.02**	1.01	1.01	0.98
0.60	1.00	0.98	0.99	1.02	1.00	0.99
0.70	1.01	1.00	1.00	1.00	1.02	1.00
0.80	1.05**	1.03	1.05**	1.05**	1.03*	1.00
0.90	0.97	0.98	0.99	1.07	1.06	1.00
0.95	1.01	0.96	0.98	1.00	1.03	1.08*
0.98	1.00	0.95	0.98	0.92	0.91	0.98
0.99	1.14**	1.04	0.99	0.94	1.12*	0.88

Note: Entries report the ratio of out-of-sample mean check loss from the Quantile LASSO model to that of the "quantile factor augmented" model. Values below one indicate that the Quantile LASSO improves upon the "quantile factor augmented" model and are marked gray. Estimation is performed using the Quantile Random Forest (number of trees n=500). Both models utilize solely the base linear predictors to avoid confounding. Asterisks denote the statistical significance of a one-sided Diebold-Mariano test, evaluated using Newey-West standard errors with an 18-month lag, indicating that the "quantile factor augmented" QRF model yields a significantly lower out-of-sample check loss than the Q-LASSO baseline (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Out-of-sample up to 2019:12.

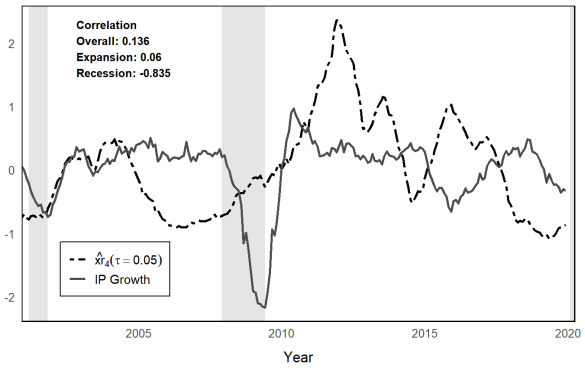
Figure 7: Bond Risk Premium Forecast ($\tau = 0.05$) vs. IP Growth (standardized - COVID Excluded)



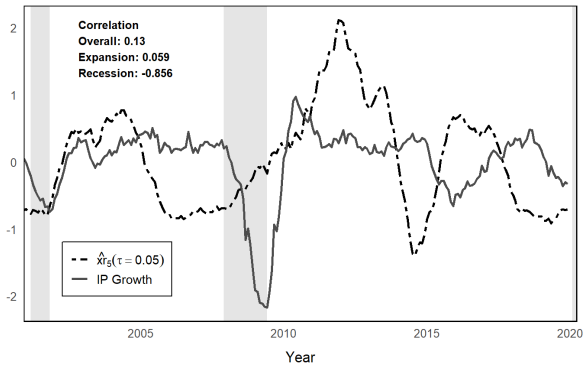
(a) $\hat{xr}^{(2)}_{t+1}(\tau = 0.05)$ vs. IP Growth



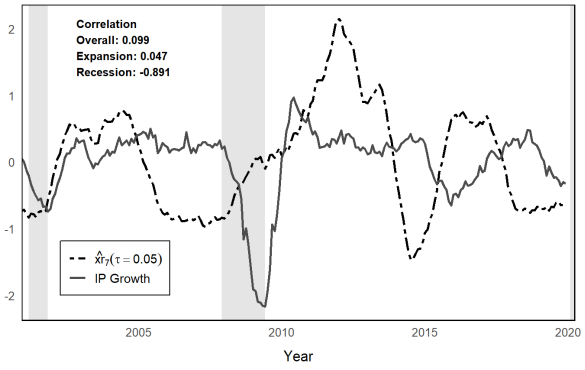
(b) $\hat{xr}^{(3)}_{t+1}(\tau = 0.05)$ vs. IP Growth



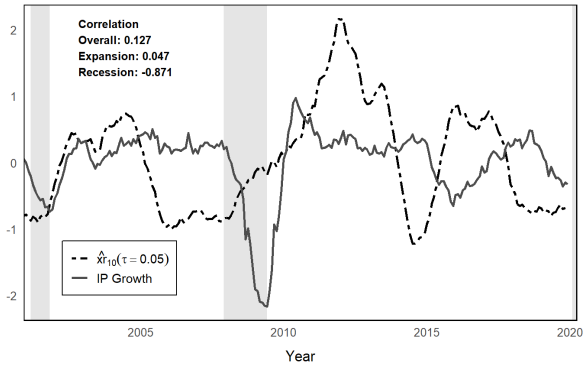
(c) $\hat{xr}^{(4)}_{t+1}(\tau = 0.05)$ vs. IP Growth



(d) $\hat{xr}^{(5)}_{t+1}(\tau = 0.05)$ vs. IP Growth



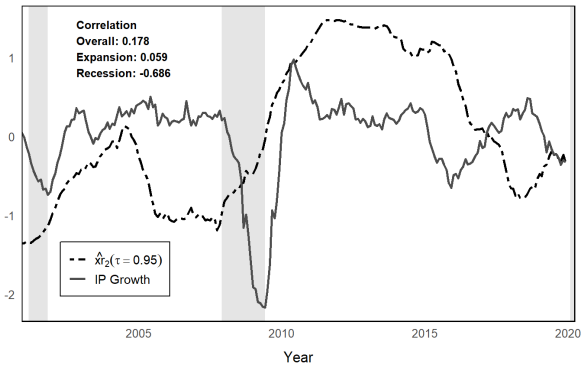
(e) $\hat{xr}^{(7)}_{t+1}(\tau = 0.05)$ vs. IP Growth



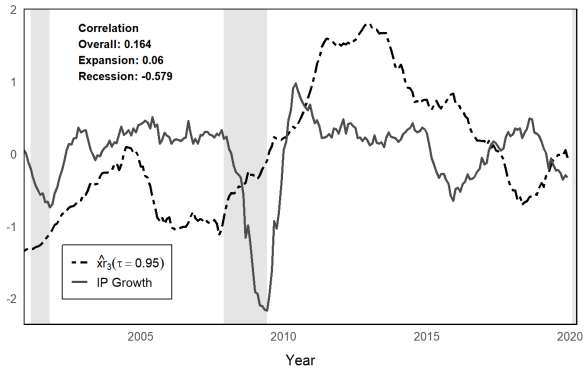
(f) $\hat{xr}^{(10)}_{t+1}(\tau = 0.05)$ vs. IP Growth

Note: Figures above demonstrate the out-of-sample predictions of excess log bond returns of maturity 2 to 10 years at $\tau = 0.05$ (generated by Quantile Random Forest) with the quantile factor model specification versus same period industrial production (IP) growth, the correlations of the two time series shown on the top left corner of each subfigure. The plots use 12-month moving averages of the standardized predictions and IP growth to make visual comparison, and the shaded areas denote recessions periods according to the National Bureau of Economic Research. Out-of-sample up to 2019:12.

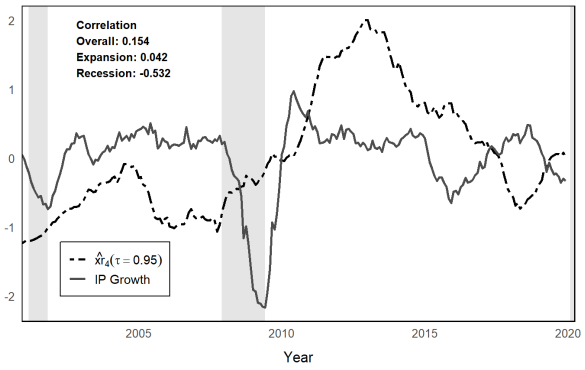
Figure 8: Bond Risk Premium Forecast ($\tau = 0.95$) vs. IP Growth (standardized - COVID Excluded)



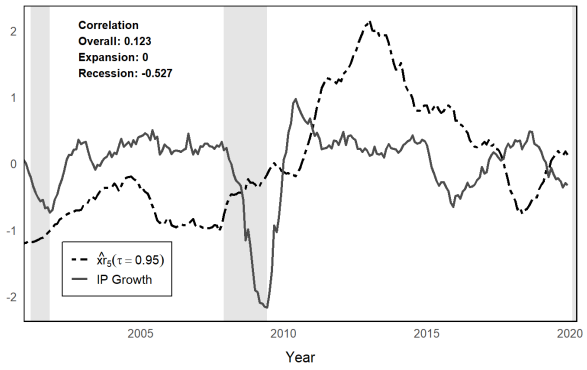
(a) $\hat{xr}^{(2)}_{t+1}(\tau = 0.95)$ vs. IP Growth



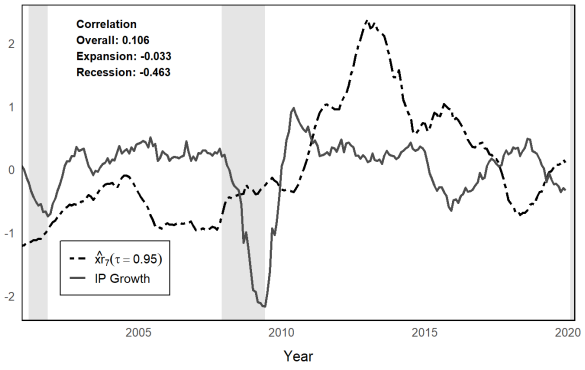
(b) $\hat{xr}^{(3)}_{t+1}(\tau = 0.95)$ vs. IP Growth



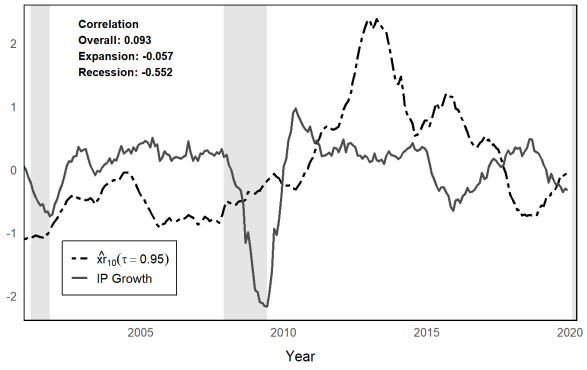
(c) $\hat{xr}^{(4)}_{t+1}(\tau = 0.95)$ vs. IP Growth



(d) $\hat{xr}^{(5)}_{t+1}(\tau = 0.95)$ vs. IP Growth



(e) $\hat{xr}^{(7)}_{t+1}(\tau = 0.95)$ vs. IP Growth



(f) $\hat{xr}^{(10)}_{t+1}(\tau = 0.95)$ vs. IP Growth

Note: Figures above demonstrate the out-of-sample predictions of excess log bond returns of maturity 2 to 10 years at $\tau = 0.95$ (generated by Quantile Random Forest) with the quantile factor model specification versus same period industrial production (IP) growth, the correlations of the two time series shown on the top left corner of each subfigure. The plots use 12-month moving averages of the standardized predictions and IP growth to make visual comparison, and the shaded areas denote recessions periods according to the National Bureau of Economic Research. Out-of-sample up to 2019:12.

G Diebold-Mariano Test Statistics

Table 12: Full Out-of-Sample Diebold-Mariano Test Statistics (Target Models vs. Financial Only Baseline)

Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
Panel A: Macro + Financial vs. Financial Only						
0.01	-1.598 (0.056)	-1.755 (0.040)	-1.731 (0.042)	-1.740 (0.041)	-1.878 (0.031)	-1.769 (0.039)
0.02	-1.604 (0.055)	-1.696 (0.046)	-1.748 (0.041)	-1.789 (0.037)	-1.853 (0.032)	-1.989 (0.024)
0.05	-1.775 (0.039)	-1.711 (0.044)	-1.652 (0.050)	-1.639 (0.051)	-1.589 (0.057)	-1.548 (0.061)
0.10	-0.439 (0.331)	-0.685 (0.247)	-1.207 (0.114)	-0.998 (0.160)	-0.551 (0.291)	-0.970 (0.167)
0.20	0.091 (0.536)	0.187 (0.574)	-0.302 (0.381)	-0.475 (0.318)	0.230 (0.591)	-0.173 (0.431)
0.30	0.312 (0.622)	0.581 (0.719)	0.476 (0.683)	0.554 (0.710)	0.958 (0.831)	0.452 (0.674)
0.40	0.735 (0.769)	1.104 (0.865)	1.074 (0.858)	1.293 (0.902)	1.098 (0.863)	0.123 (0.549)
0.50	0.849 (0.802)	1.138 (0.872)	1.194 (0.883)	1.028 (0.848)	0.815 (0.792)	-0.304 (0.381)
0.60	0.509 (0.694)	0.659 (0.745)	0.783 (0.783)	0.553 (0.710)	0.225 (0.589)	-0.798 (0.213)
0.70	-0.294 (0.385)	-0.245 (0.403)	-0.687 (0.246)	-0.964 (0.168)	-1.360 (0.087)	-1.812 (0.036)
0.80	-1.679 (0.047)	-1.662 (0.049)	-2.688 (0.004)	-2.458 (0.007)	-2.342 (0.010)	-3.705 (0.000)
0.90	-2.992 (0.002)	-2.943 (0.002)	-2.903 (0.002)	-3.159 (0.001)	-3.146 (0.001)	-3.419 (0.000)
0.95	-2.852 (0.002)	-3.051 (0.001)	-3.323 (0.001)	-3.234 (0.001)	-3.286 (0.001)	-3.191 (0.001)
0.98	-2.311 (0.011)	-2.694 (0.004)	-3.221 (0.001)	-2.853 (0.002)	-3.269 (0.001)	-2.843 (0.002)
0.99	-2.312 (0.011)	-2.190 (0.015)	-2.536 (0.006)	-2.594 (0.005)	-2.830 (0.002)	-2.468 (0.007)
Panel B: Quantile Factor Augmented vs. Financial Only						
0.01	-1.602 (0.055)	-1.699 (0.045)	-1.667 (0.048)	-1.707 (0.044)	-1.883 (0.030)	-1.749 (0.041)
0.02	-1.579 (0.058)	-1.609 (0.054)	-1.620 (0.053)	-1.674 (0.048)	-1.835 (0.034)	-1.889 (0.030)
0.05	-1.729 (0.042)	-1.588 (0.057)	-1.610 (0.054)	-1.587 (0.057)	-1.661 (0.049)	-1.705 (0.045)
0.10	-1.522 (0.065)	-1.468 (0.072)	-1.583 (0.057)	-1.344 (0.090)	-1.241 (0.108)	-1.624 (0.053)
0.20	-1.404 (0.081)	-1.377 (0.085)	-1.223 (0.111)	-1.193 (0.117)	-1.183 (0.119)	-1.315 (0.095)
0.30	-1.191 (0.117)	-1.041 (0.149)	-0.935 (0.175)	-1.023 (0.154)	-1.043 (0.149)	-0.988 (0.162)
0.40	-1.135 (0.129)	-0.635 (0.263)	-0.783 (0.217)	-0.678 (0.249)	-0.840 (0.201)	-1.255 (0.105)
0.50	-1.138 (0.128)	-1.051 (0.147)	-1.032 (0.152)	-0.949 (0.172)	-1.257 (0.105)	-1.653 (0.050)
0.60	-1.959 (0.026)	-1.703 (0.045)	-1.317 (0.094)	-1.460 (0.073)	-1.568 (0.059)	-2.063 (0.020)
0.70	-2.640 (0.004)	-2.447 (0.007)	-2.590 (0.005)	-2.479 (0.007)	-2.794 (0.003)	-2.865 (0.002)
0.80	-3.070 (0.001)	-3.428 (0.000)	-3.920 (0.000)	-3.510 (0.000)	-3.912 (0.000)	-4.534 (0.000)
0.90	-3.462 (0.000)	-3.588 (0.000)	-3.787 (0.000)	-3.971 (0.000)	-4.243 (0.000)	-4.111 (0.000)
0.95	-2.896 (0.002)	-3.036 (0.001)	-3.715 (0.000)	-3.613 (0.000)	-3.549 (0.000)	-3.622 (0.000)
0.98	-2.252 (0.013)	-2.563 (0.005)	-3.046 (0.001)	-2.703 (0.004)	-2.912 (0.002)	-2.737 (0.003)
0.99	-2.331 (0.010)	-2.385 (0.009)	-2.655 (0.004)	-2.657 (0.004)	-2.587 (0.005)	-2.574 (0.005)

Note: The table reports the raw Diebold-Mariano test statistics with corresponding one-sided p-values in parentheses. The loss differential is defined as $d_t = \text{Loss}_{\text{Target}} - \text{Loss}_{\text{Financial}}$. Thus, a negative test statistic indicates that the target model (Macro + Financial in Panel A; Quantile Factor Augmented in Panel B) yields a lower out-of-sample check loss than the baseline Financial Only model. The test is evaluated using Newey-West standard errors with an 18-month lag.

Table 13: Pre-COVID Out-of-Sample Diebold-Mariano Test Statistics (Target Models vs. Financial Only Baseline)

Panel A: Macro + Financial vs. Financial Only						
Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
0.01	-2.033 (0.022)	-1.948 (0.026)	-2.015 (0.023)	-2.221 (0.014)	-2.268 (0.012)	-1.950 (0.026)
0.02	-1.679 (0.047)	-1.724 (0.043)	-1.851 (0.033)	-2.151 (0.016)	-1.935 (0.027)	-1.702 (0.045)
0.05	-0.796 (0.214)	-0.847 (0.199)	-0.794 (0.214)	-0.760 (0.224)	-0.751 (0.227)	-0.852 (0.197)
0.10	0.313 (0.623)	0.206 (0.582)	0.307 (0.621)	0.403 (0.656)	0.357 (0.639)	0.428 (0.665)
0.20	0.827 (0.796)	1.074 (0.858)	1.278 (0.899)	1.417 (0.921)	1.152 (0.875)	1.353 (0.911)
0.30	1.163 (0.877)	1.648 (0.950)	2.093 (0.981)	2.186 (0.985)	1.983 (0.976)	1.425 (0.922)
0.40	1.077 (0.859)	1.823 (0.965)	2.255 (0.987)	2.355 (0.990)	2.122 (0.983)	1.455 (0.927)
0.50	1.032 (0.848)	1.699 (0.955)	1.945 (0.974)	1.821 (0.965)	1.434 (0.924)	0.419 (0.662)
0.60	0.833 (0.797)	1.357 (0.912)	1.328 (0.907)	0.971 (0.834)	1.008 (0.843)	-0.059 (0.476)
0.70	-0.119 (0.453)	0.107 (0.542)	-0.289 (0.386)	-0.900 (0.185)	-0.797 (0.213)	-1.622 (0.053)
0.80	-1.409 (0.080)	-1.639 (0.051)	-3.078 (0.001)	-2.750 (0.003)	-2.207 (0.014)	-3.043 (0.001)
0.90	-2.851 (0.002)	-3.019 (0.001)	-3.198 (0.001)	-3.030 (0.001)	-3.594 (0.000)	-3.421 (0.000)
0.95	-2.785 (0.003)	-3.152 (0.001)	-3.654 (0.000)	-3.348 (0.000)	-3.404 (0.000)	-3.274 (0.001)
0.98	-2.799 (0.003)	-3.368 (0.000)	-3.605 (0.000)	-3.628 (0.000)	-3.133 (0.001)	-3.115 (0.001)
0.99	-2.831 (0.003)	-3.422 (0.000)	-3.883 (0.000)	-3.700 (0.000)	-3.042 (0.001)	-3.092 (0.001)
Panel B: Quantile Factor Augmented vs. Financial Only						
0.01	-2.139 (0.017)	-2.051 (0.021)	-2.107 (0.018)	-2.321 (0.011)	-2.391 (0.009)	-2.008 (0.023)
0.02	-1.914 (0.028)	-1.890 (0.030)	-2.065 (0.020)	-2.304 (0.011)	-2.178 (0.015)	-1.823 (0.035)
0.05	-1.326 (0.093)	-1.421 (0.078)	-1.341 (0.091)	-1.227 (0.111)	-1.410 (0.080)	-1.374 (0.085)
0.10	-0.544 (0.293)	-0.707 (0.240)	-0.678 (0.249)	-0.297 (0.383)	-0.550 (0.291)	-0.601 (0.274)
0.20	-0.584 (0.280)	-0.256 (0.399)	-0.243 (0.404)	-0.289 (0.386)	-0.149 (0.441)	0.032 (0.513)
0.30	-0.200 (0.421)	0.771 (0.779)	0.705 (0.759)	0.635 (0.737)	0.622 (0.733)	0.199 (0.579)
0.40	-0.441 (0.330)	0.724 (0.765)	0.556 (0.710)	0.517 (0.697)	0.746 (0.772)	0.109 (0.543)
0.50	-0.448 (0.327)	-0.017 (0.493)	0.351 (0.637)	0.197 (0.578)	-0.049 (0.480)	-0.585 (0.280)
0.60	-1.320 (0.094)	-0.498 (0.310)	-0.324 (0.373)	-0.875 (0.191)	-0.351 (0.363)	-1.010 (0.157)
0.70	-2.100 (0.018)	-1.759 (0.040)	-1.729 (0.043)	-1.636 (0.052)	-1.736 (0.042)	-2.181 (0.015)
0.80	-2.377 (0.009)	-2.798 (0.003)	-3.702 (0.000)	-3.644 (0.000)	-3.340 (0.000)	-3.542 (0.000)
0.90	-3.033 (0.001)	-3.266 (0.001)	-3.701 (0.000)	-3.662 (0.000)	-4.006 (0.000)	-3.661 (0.000)
0.95	-2.833 (0.003)	-3.455 (0.000)	-3.854 (0.000)	-3.545 (0.000)	-3.658 (0.000)	-3.521 (0.000)
0.98	-2.880 (0.002)	-3.452 (0.000)	-3.679 (0.000)	-3.604 (0.000)	-3.074 (0.001)	-3.017 (0.001)
0.99	-2.859 (0.002)	-3.526 (0.000)	-3.891 (0.000)	-3.649 (0.000)	-3.105 (0.001)	-3.033 (0.001)

Note: The table reports the raw Diebold-Mariano test statistics with corresponding one-sided p-values in parentheses, based on an out-of-sample period up to 2019:12 (COVID period excluded). The loss differential is defined as $d_t = \text{Loss}_{\text{Target}} - \text{Loss}_{\text{Financial}}$. Thus, a negative test statistic indicates that the target model (Macro + Financial in Panel A; Quantile Factor Augmented in Panel B) yields a lower out-of-sample check loss than the baseline Financial Only model. The test is evaluated using Newey-West standard errors with an 18-month lag.

Table 14: Full Out-of-Sample Diebold-Mariano Test Statistics (Quantile Factor Augmented vs. Q-LASSO)

Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
0.01	-1.053 (0.147)	-1.007 (0.157)	-0.850 (0.198)	-0.760 (0.224)	-0.582 (0.280)	-0.533 (0.297)
0.02	-1.426 (0.077)	-1.424 (0.078)	-1.233 (0.109)	-1.271 (0.102)	-1.371 (0.086)	-1.425 (0.078)
0.05	0.682 (0.752)	-0.890 (0.187)	0.839 (0.799)	-0.836 (0.202)	0.126 (0.550)	-0.307 (0.380)
0.10	-1.708 (0.044)	-1.594 (0.056)	-1.403 (0.081)	-1.515 (0.065)	-1.642 (0.051)	-1.932 (0.027)
0.20	-2.697 (0.004)	-1.835 (0.034)	-0.886 (0.188)	-1.451 (0.074)	-2.535 (0.006)	-1.800 (0.036)
0.30	-1.757 (0.040)	-2.624 (0.005)	-0.258 (0.398)	-2.294 (0.011)	-2.171 (0.015)	-0.954 (0.170)
0.40	-1.320 (0.094)	-1.142 (0.127)	-0.505 (0.307)	-1.013 (0.156)	-0.679 (0.249)	-0.620 (0.268)
0.50	-0.261 (0.397)	-2.536 (0.006)	-0.760 (0.224)	-1.915 (0.028)	0.696 (0.756)	0.694 (0.756)
0.60	-0.074 (0.470)	-0.409 (0.341)	0.305 (0.620)	-0.150 (0.440)	-0.131 (0.448)	1.120 (0.868)
0.70	-1.298 (0.098)	0.184 (0.573)	-0.162 (0.436)	0.235 (0.593)	-0.033 (0.487)	1.413 (0.921)
0.80	-1.121 (0.132)	-2.218 (0.014)	-1.527 (0.064)	-0.868 (0.193)	-1.809 (0.036)	0.313 (0.623)
0.90	0.730 (0.767)	1.800 (0.964)	0.042 (0.517)	1.569 (0.941)	1.070 (0.857)	-0.104 (0.459)
0.95	0.936 (0.825)	-1.112 (0.133)	-1.013 (0.156)	-0.555 (0.290)	-0.885 (0.189)	1.334 (0.908)
0.98	1.437 (0.924)	1.130 (0.870)	1.055 (0.854)	-0.261 (0.397)	0.631 (0.736)	-0.059 (0.477)
0.99	-0.555 (0.290)	0.001 (0.500)	-0.558 (0.289)	-1.378 (0.085)	-0.789 (0.215)	1.065 (0.856)

Note: The table reports the raw Diebold-Mariano test statistics with corresponding one-sided p-values in parentheses. The loss differential is defined as $d_t = \text{Loss}_{\text{Target}} - \text{Loss}_{\text{Baseline}}$, where the target is the Quantile Factor Augmented model and the baseline is the Q-LASSO model. Thus, a negative test statistic indicates that the Quantile Factor Augmented model yields a lower out-of-sample check loss than the linear Q-LASSO alternative. The test is evaluated using Newey-West standard errors with an 18-month lag.

Table 15: Pre-COVID Out-of-Sample Diebold-Mariano Test Statistics (Quantile Factor Augmented vs. Q-LASSO)

Quantile	$xr_{t+1}^{(2)}$	$xr_{t+1}^{(3)}$	$xr_{t+1}^{(4)}$	$xr_{t+1}^{(5)}$	$xr_{t+1}^{(7)}$	$xr_{t+1}^{(10)}$
0.01	-0.829 (0.204)	-0.262 (0.397)	0.813 (0.791)	0.911 (0.818)	0.773 (0.780)	0.951 (0.829)
0.02	-2.997 (0.002)	-1.573 (0.058)	-0.902 (0.184)	-0.174 (0.431)	-0.771 (0.221)	-0.818 (0.207)
0.05	-3.229 (0.001)	-0.311 (0.378)	-0.087 (0.465)	-1.257 (0.105)	-0.789 (0.216)	-0.470 (0.319)
0.10	-2.310 (0.011)	-1.992 (0.024)	-1.597 (0.056)	-1.671 (0.048)	-2.169 (0.016)	-2.361 (0.010)
0.20	-3.163 (0.001)	-2.333 (0.010)	-2.788 (0.003)	-3.588 (0.000)	-2.323 (0.011)	-1.934 (0.027)
0.30	-1.971 (0.025)	-1.422 (0.078)	-1.560 (0.060)	-0.693 (0.244)	-2.428 (0.008)	-0.024 (0.490)
0.40	-1.619 (0.053)	-1.350 (0.089)	-2.041 (0.021)	-2.372 (0.009)	-1.689 (0.046)	-1.053 (0.147)
0.50	-1.422 (0.078)	-2.020 (0.022)	-1.689 (0.046)	-1.147 (0.126)	-0.687 (0.246)	1.299 (0.902)
0.60	-0.211 (0.417)	1.185 (0.881)	0.494 (0.689)	-1.269 (0.103)	-0.245 (0.403)	0.408 (0.658)
0.70	-0.489 (0.313)	0.256 (0.601)	-0.215 (0.415)	-0.031 (0.488)	-0.893 (0.186)	-0.148 (0.441)
0.80	-2.059 (0.020)	-1.003 (0.159)	-2.089 (0.019)	-1.901 (0.029)	-1.576 (0.058)	0.124 (0.549)
0.90	0.648 (0.741)	0.607 (0.728)	0.511 (0.695)	-0.923 (0.178)	-1.047 (0.148)	-0.125 (0.450)
0.95	-0.313 (0.377)	1.257 (0.895)	0.724 (0.765)	0.135 (0.554)	-0.740 (0.230)	-1.575 (0.058)
0.98	0.306 (0.620)	0.760 (0.776)	0.445 (0.672)	1.197 (0.884)	1.092 (0.862)	0.759 (0.776)
0.99	-2.126 (0.017)	-0.806 (0.210)	0.361 (0.641)	1.394 (0.918)	-1.577 (0.058)	1.915 (0.972)

Note: The table reports the raw Diebold-Mariano test statistics with corresponding one-sided p-values in parentheses, based on an out-of-sample period up to 2019:12 (COVID period excluded). The loss differential is defined as $d_t = \text{Loss}_{\text{Target}} - \text{Loss}_{\text{Baseline}}$, where the target is the Quantile Factor Augmented model and the baseline is the Q-LASSO model. Thus, a negative test statistic indicates that the Quantile Factor Augmented model yields a lower out-of-sample check loss than the linear Q-LASSO alternative. The test is evaluated using Newey-West standard errors with an 18-month lag.