Volatility Forecasting Factors*

Marco Cinquetti[†] Seok Young Hong[‡] Ingmar Nolte[§] Sandra Nolte[¶]

This version: September 18, 2025

Abstract

This paper examines the predictive role of high-frequency factor volatilities in modeling the volatility of individual stocks. We develop a dynamic forecasting framework that selects the most informative factor-specific realized volatility from a large cross-section of asset pricing anomalies. Embedded in a log-linear specification, the model integrates both market-wide and idiosyncratic components, allowing for a flexible representation of volatility dynamics. We prove the strong consistency of our selection procedure, and show that our selection rule asymptotically identifies the factor that truly drives volatility. We further show how measurement errors affect the adaptive selection. Empirical results based on a broad universe of U.S. equities demonstrate that the proposed method significantly outperforms standard benchmarks, both statistically and economically. The findings underscore the importance of incorporating high-frequency cross-sectional information in volatility modeling, offering a scalable and interpretable approach to understanding time-varying risks in equity markets.

Keywords: Realized volatility; Factor models; High-frequency data; Volatility forecasting
JEL Classification: C51, C53, C55, C58, G17

*We thank Nicola Fusari, Yifan Li, Aleksey Kolokolov, Olga Kolokolova, Manh Pham, Roberto Renò, Shifan Yu, Xinyi Zhang (discussant), as well as participants at the Financial Econometrics Conference in Honour of Stephen Taylor and the 2nd Financial Fraud, Misconduct and Market Manipulation Conference, and seminars at various institutions, for helpful comments and discussions.

†Corresponding author. Lancaster University Management School, Lancaster, United Kingdom; m.cinquetti1@lancaster.ac.uk

‡Nanyang Technological University, Singapore; seokyoung.hong@ntu.edu.sg

§Lancaster University Management School, Lancaster, United Kingdom; i.nolte@lancaster.ac.uk

¶Lancaster University Management School, Lancaster, United Kingdom; s.nolte@lancaster.ac.uk

1 Introduction

Volatility plays a fundamental role in finance as it underlies the risk-reward dynamics defining modern financial theories, from investment decision-making to monetary policies. Despite its centrality, the modeling of volatility remains largely uninformed about the multivariate frameworks that dominate empirical asset pricing. From the first factor model of Sharpe (1964) to the "zoo" of factors in Harvey et al. (2016), the literature on financial returns has long embraced systematic, cross-sectional structures; in contrast, volatility has mainly been modeled as an idiosyncratic, asset-specific process. As Bollerslev (2022) notes, volatility modeling has been slow to internalize the inherently multivariate nature of market dynamics, often ignoring the factor structures that drive both return co-movements and correlated risk exposures. The result is a methodological gap: whereas expected returns are modeled via linear exposures to observable or latent factors, volatility is typically forecasted without acknowledging systematic drivers.

Any volatility modeling framework ultimately hinges on how precisely variance can be measured. Early contributions such as the ARCH model of Engle (1982) and its generalized version by Bollerslev (1986) introduced conditional heteroskedasticity as a time-varying property of return series. Parallel to these, the stochastic volatility models by Taylor (1982) offered a latent process formulation. The introduction of high-frequency financial data marked a fundamental shift, enabling, under ideal sampling conditions, nonparametric estimation of risk with realized volatility (RV) measures. However, the discrete and discontinuous nature of financial markets poses challenges for the unbiased and consistent estimation of realized volatility: multi-scale estimators (Zhang et al., 2005, Nolte and Voev, 2012) and pre-averaging methods (Christensen et al., 2014) address market microstructure noise; bipower and multipower estimators (Barndorff-Nielsen and Shephard, 2004), truncated estimators (Mancini, 2009) and some combinations thereof (Corsi et al., 2010) focus on jumps; recent contributions (Andersen et al., 2023, Li et al., 2025) provide robust estimation in presence of short-lived extreme price movements. These advancements have made it possible to recover the latent volatility process efficiently, providing a reliable instrument for volatility forecasting models.

One of the most influential frameworks to leverage realized volatility is the Heterogeneous Autoregressive model of Corsi (2009). The HAR model parsimoniously captures the long-memory behavior of volatility by including lagged daily, weekly, and monthly realized volatilities as regressors, approximating the effect of a broader lag distribution in a simple linear specification. Volatility forecasting has further improved with extensions such as the semivariance HAR (SHAR) model of Patton and Sheppard (2015), allowing for conditional asymmetries through signed returns, or the HARQ of Bollerslev et al. (2016), introducing realized quarticity to account for temporal variations in RV measurement errors. Despite these refinements, a foundational limitation remains: HAR models are inherently univariate, thus agnostic to any systematic risk factors.

The limitations of univariate volatility models have motivated a shift towards specifications incorporating cross-sectional information. A prominent example is the market-HAR model introduced by Hizmeri et al. (2022), which augments the standard framework with market-level realized (co)variances and semi(co)variances. Building on similar reasoning, the multiplicative volatility factor (MVF) model proposed by Ding et al. (2025) presents a parsimonious structure in which the realized variance of each stock is expressed as the product of a latent common volatility factor and an idiosyncratic residual. Empirically, both models demonstrate significant gains in forecast accuracy, confirming a stylized fact: stock volatilities co-move over time, often driven by aggregated shocks rather than isolated firm-level events. This insight is well-grounded in the empirical literature. Early studies by Engle et al. (1988) and Calvet et al. (2006) document volatility spillovers across markets and asset classes. Complementing this, Herskovic et al. (2016) identify a latent common idiosyncratic volatility factor that explains a significant fraction of cross-sectional volatility dispersion. More recent works emphasize the role of firm-level linkages: Herskovic et al. (2020) show that firms embedded in central positions within economic networks exhibit stronger volatility co-movement. At the macro level, Bollerslev et al. (2018) and Engle and Campos-Martins (2023) provide evidence of global volatility factors driving fluctuations in equity markets, reinforcing the importance of modeling volatility beyond the firm-specific scale.

The observed heterogeneity in volatility patterns extends beyond a single driver. Figure 1 displays the annualized realized volatility of the high-frequency market (MKT) factor against the other four Fama and French (2015) factors, namely size (SMB), value (HML), profitability (RMW), and investment (CMA). Every time series exhibits distinct and persistent dynamics over time, suggesting that volatility may originate from multiple, structurally distinct sources. Recent works by Barigozzi and Hallin (2017) and Kapadia et al. (2024) provide robust evidence that multiple volatility shocks arise from diverse economic sources such as styles, industries, and risk premia structures. In this context, the identification of the most informative factor variance must be flexible and dynamic, thereby motivating adaptive multi-factor frameworks that update their forecasting structure in response to shifting volatility regimes.

The present paper introduces a novel volatility model that extends the HAR structure by dynamically selecting the most informative components among a large cross-section of high-frequency asset pricing factors. Specifically, we construct daily realized variances based on second-level intraday returns for 287 observable anomalies, covering the full set proposed in Fama and French (2018), Chen and Zimmermann (2022) and Jensen et al. (2023). Methodologically, this approach contributes to the literature on high-dimensional forecasting by imposing structural sparsity through selection rather than shrinkage. Unlike LASSO-type

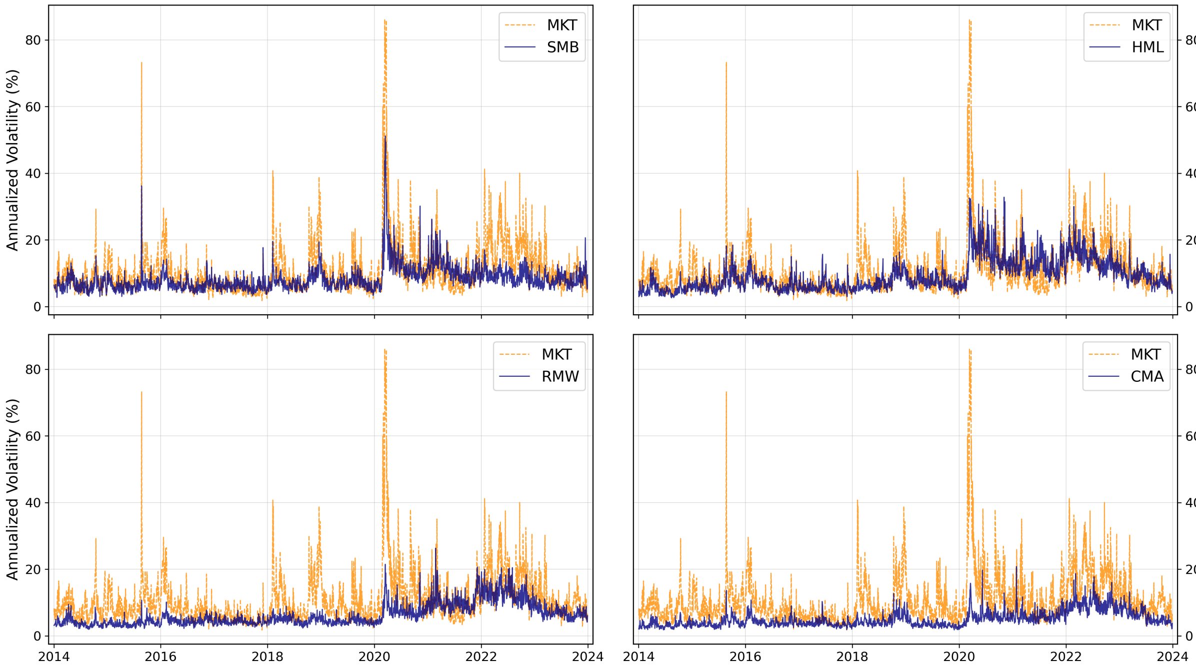

Figure 1: Annualized 5-minute realized volatility of Fama-French five factors (MKT, SMB, HML, RMW, CMA) from 2014 to 2023. The figure illustrates the heterogeneous dynamics of factor volatilities.

estimators (Chinco et al., 2019, Gu et al., 2020) or composite forecast models (Freyberger et al., 2020), which often yield opaque functional relationships or unstable inclusion patterns, our model remains interpretable, tractable, and grounded in economic theory. Moreover, it complements recent advances in volatility decomposition (Barigozzi and Hallin, 2020, Luciani and Veredas, 2015), network-based risk propagation (Zheng and Li, 2011), and regime-sensitive factor modeling (Asai et al., 2015, Atak and Kapetanios, 2013).

We evaluate forecasting performance across more than one thousand U.S. equities, markedly exceeding the cross-sectional coverage of typical studies in the volatility forecasting literature. The proposed model is benchmarked against several prominent volatility forecasting frameworks, delivering statistically significant improvements in forecast accuracy. Assets inherit volatility from common drivers: shocks to factor-level uncertainty propagate to individual variance through time-varying exposures, so realized factor volatility at multiple horizons add information beyond an asset's own history. The robustness of these gains is confirmed across multiple realized volatility estimators, forecasting horizons, and loss functions. These results establish the necessity of incorporating factor-specific volatility components in forecasting models, and show that informational efficiency depends not only on the precision of realized-variance measurement but also on the structural selection of its predictors.

The remainder of the paper is structured as follows. Section 2 outlines the econometric framework, model specification, and theoretical foundations of the log-HAR framework with adaptive factor volatilities. We prove the strong consistency of our selection procedure, where we show the quasi-likelihood (QLIKE)-based selection rule asymptotically identifies the factor that truly drives volatility. We also characterize how the mis-selection probability responds to measurement error. Section 3 describes the dataset construction and implementation details. Section 4 presents the forecasting results and factor analysis. Section 5 concludes.

2 Theoretical framework

2.1 Setup and background theory

Throughout the paper, we work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with a complete and right-continuous filtration. For each asset $i = 1, \ldots, N$, the log-price process $(P_{i,t})_{t\geq 0}$ is defined on this space, adapted to $(\mathcal{F}_t)$, and is an Itô semimartingale of the form

$$dP_{i,t} = \mu_{i,t}\, dt + \sigma_{i,t}\, dW_{i,t}; \qquad t \geq 0, \tag{1}$$

where the drift $\mu_{i,t}$ and instantaneous volatility $\sigma_{i,t}$ are $(\mathcal{F}_t)$-progressively measurable processes such that, for every $T < \infty$, $\int_0^T |\mu_{i,t}| dt < \infty$ a.s. and $\int_0^T \sigma_{i,t}^2 dt < \infty$ a.s., and $W_{i,t}$ is a standard $(\mathcal{F}_t)$-Brownian motion.

Writing the trading day $t \geq 1$ as the unit-length interval $(t-1, t]$, we define the latent integrated variance for asset i on day t as

$$IV_{i,t} := \int_{t-1}^{t} \sigma_{i,s}^2\, ds. \tag{2}$$

The integrated variance aggregates the entire intraday volatility path, and is the quantity we ultimately seek to forecast at daily and longer horizons. To maintain conciseness in the exposition, we abstract from discontinuous price moves here; jumps and other extreme price movements are incorporated through robust estimators in the empirical analysis presented in Section 4.

Although the integrated variance $IV_{i,t}$ is not observable, it is well known that high-frequency returns yield a consistent nonparametric estimator. Specifically, let $\Delta = 1/M$ be the intraday sampling interval and index observations by $j = 1, \ldots, M$. On the equally spaced intraday grid, the j-th intraday return is $r_{i,t,j} = P_{i,t-1+j\Delta} - P_{i,t-1+(j-1)\Delta}$. Under the continuous semimartingale assumption and in the absence of market microstructure noise, the realized variance

$$RV_{i,t} := \sum_{j=1}^{M} r_{i,t,j}^2 = \sum_{j=1}^{M} \left(P_{i,t-1+j\Delta} - P_{i,t-1+(j-1)\Delta}\right)^2 \tag{3}$$

converges in probability to $IV_{i,t}$ as the mesh $\Delta \to 0$ (i.e. $M \to \infty$); the analogous joint convergence holds in multivariate sense for the realized covariance matrix, and hence for any fixed linear portfolio. If there was a jump component in (1), then (3) would converge in probability to the quadratic variation, which incorporates the jump variations in addition to the integrated variance. See Theorems 4.47 and 4.52 of Jacod and Shiryaev (2003) for details. The limiting distribution of the realized variance has been widely investigated in the literature under diverse conditions, see Aït-Sahalia and Jacod (2014) for a comprehensive exposition.

We construct asset pricing factors from the same universe of stocks as the individual assets. Let $r_{k,t}$ denote the daily log return on an asset pricing factor $k \in \mathcal{K}$, with $\mathcal{K}$ denoting the finite index set corresponding to the set of candidate factors (e.g., MKT, SMB, HML).[1] Writing $r_{k,t,j}$ for the j-th intraday log return of factor k computed on the equally spaced grid $\{t-1+\Delta, t-1+2\Delta, \dots, t\}$ with mesh $\Delta = 1/M$, we estimate the factor volatility nonparametrically by the realized variance

$$FRV_{k,t} := \sum_{j=1}^{M} r_{k,t,j}^2, \tag{4}$$

the *realized factor variance*. By continuous mapping, the consistency and limiting distribution results extend to the realized variance of the factor return, provided the sampling is free of microstructure noise and the factor is implemented as a self-financing, tradable linear portfolio with weights that are $\mathcal{F}_{t-1}$-measurable and of bounded variation on $(t-1, t]$.

The asset- and factor-level realized variances defined above constitute the observable building blocks for the theoretical framework developed in the next subsections.

2.2 Adaptive Factor-Driven Volatility Models

We propose a volatility forecasting framework that augments the heterogeneous autoregressive (HAR) structure with multiplicative factor components. To ensure the non-negativity of the variance, to alleviate right-skewness and heavy tails, and to render the multiplicative decomposition additive, we adopt a log-linear specification and work with $\log RV_{i,t}$, i.e., the log realized variance of stock i on day t.

We model $\log RV_{i,t}$ as the linear combination of three log predictor blocks. First, the market proxy, which is defined as the cross-sectional average of realized variances across all stocks

$$CRV_t := \frac{1}{N} \sum_{i=1}^{N} RV_{i,t}, \tag{5}$$

hereafter referred to as *common realized variance* (CRV). Second, the stock-specific component

$$\xi_{i,t} := \frac{RV_{i,t}}{CRV_t}, \tag{6}$$

defined as the multiplicative residual of stock i's variance relative to the cross-section. Third,

[1] All factors examined are diversified in the sense of Ross (2013), so the associated portfolios bear negligible firm-specific risk. This feature distinguishes our analysis from Herskovic et al. (2016) and related studies that focus on commonality in the volatility of firm-specific returns.

the factor specific component $FRV_{k^\star_{i,t}}$, which is the *realized factor variance* (4) corresponding to the high-frequency factor $k^\star_{i,t} \in \mathcal{K}$, where $k^\star_{i,t}$ is chosen adaptively as the factor whose realized variance lags yield the best in-sample explanatory power for stock i.[2] See (8) below and the discussion that follows.

Each component adds to the model daily, weekly, and monthly lags. Specifically, for each stock i, day t and forecasting horizon h, the log-realized variance $\log RV_{i,t+h}$ follows the linear model

$$\begin{aligned}
\log RV_{i,t+h} &= \beta_0 + \beta^{(d)}_{CRV} \log CRV^{(d)}_t + \beta^{(w)}_{CRV} \log CRV^{(w)}_t + \beta^{(m)}_{CRV} \log CRV^{(m)}_t \\
&\quad + \beta^{(d)}_{\xi} \log \xi^{(d)}_{i,t} + \beta^{(w)}_{\xi} \log \xi^{(w)}_{i,t} + \beta^{(m)}_{\xi} \log \xi^{(m)}_{i,t} \\
&\quad + \beta^{(d)}_{k^\star} \log FRV^{(d)}_{k^\star_{i,t}} + \beta^{(w)}_{k^\star} \log FRV^{(w)}_{k^\star_{i,t}} + \beta^{(m)}_{k^\star} \log FRV^{(m)}_{k^\star_{i,t}} + \varepsilon_{i,t+h} \\
&= \beta_0 + \beta^\top_{\text{CRV}} \underbrace{\begin{bmatrix} \log CRV^{(d)}_t \\ \log CRV^{(w)}_t \\ \log CRV^{(m)}_t \end{bmatrix}}_{\text{Market/Common block}} + \beta^\top_{\xi} \underbrace{\begin{bmatrix} \log \xi^{(d)}_{i,t} \\ \log \xi^{(w)}_{i,t} \\ \log \xi^{(m)}_{i,t} \end{bmatrix}}_{\text{Stock block}} + \beta^\top_{k^\star} \underbrace{\begin{bmatrix} \log FRV^{(d)}_{k^\star_{i,t}} \\ \log FRV^{(w)}_{k^\star_{i,t}} \\ \log FRV^{(m)}_{k^\star_{i,t}} \end{bmatrix}}_{\text{Factor block}} + \varepsilon_{i,t+h},
\end{aligned} \tag{7}$$

where the coefficients $\beta_{CRV} = (\beta^{(d)}_{CRV}, \beta^{(w)}_{CRV}, \beta^{(m)}_{CRV})$, $\beta_\xi = (\beta^{(d)}_\xi, \beta^{(w)}_\xi, \beta^{(m)}_\xi)$ and $\beta_{k^\star} = (\beta^{(d)}_{k^\star}, \beta^{(w)}_{k^\star}, \beta^{(m)}_{k^\star})$ are estimated by the OLS over a rolling estimation window of fixed length L. The superscripts (d), (w), (m) refer to the daily $(x^{(d)}_t = x_t)$, weekly $(x^{(w)}_t = \frac{1}{5}\sum_{\tau=0}^{4} x_{t-\tau})$, and monthly $(x^{(m)}_t = \frac{1}{22}\sum_{\tau=0}^{21} x_{t-\tau})$ averages, respectively, for $x_t = CRV_t$, $\xi_{i,t}$, or $FRV_{k^\star_{i,t}}$.

For a selection window of length S, the high-frequency factor to be chosen by $k^\star_{i,t}$ is the one delivering the *best in-sample quasi-likelihood* (QLIKE) performance. That is, we choose the factor that minimizes the QLIKE loss

$$k^\star_{i,t} := \underset{k \in \mathcal{K}}{\arg\min} \left\{ \frac{1}{S} \sum_{s=t-S+1}^{t} \left[\log\left(\frac{\widehat{RV}^{\,k}_{i,s}}{RV_{i,s}} \right) + \frac{RV_{i,s}}{\widehat{RV}^{\,k}_{i,s}} - 1 \right] \right\}, \tag{8}$$

where $RV_{i,s}$ is the measured realized variance and $\widehat{RV}^{\,k}_{i,s}$ is the forecasted (with candidate factor k) realized variance of stock i on day s of the selection period.[3] The forecasts $\widehat{RV}^{\,k}_{i,s}$ are calculated using the estimated coefficients of the proposed model, which includes the factor block, daily, weekly, and monthly lags of $\log FRV_{k^\star_{i,t}}$, alongside the CRV and ξ blocks.

[2]Equation (4) defines $FRV_{k,t}$ as the realized variance of factor k at date t. When the selected factor index is $k^\star_{i,t}$ (chosen for stock i at date t), the corresponding series is $FRV_{k^\star_{i,t},t}$. To avoid reporting the same time index, we henceforth write $FRV_{k^\star_{i,t}}$

[3]As a robustness check, we also implement the selection rule using the RMSE loss metric. The resulting forecasts (reported in Appendix C) are qualitatively similar, indicating that our adaptive factor choice is not sensitive to the specific loss function employed.

Remark. The estimation window L and the selection window S serve different purposes. The former is used to estimate the coefficients in (7) while the latter is used to score candidate factors in (8). In our specification, we impose $S \leq L$ so that the S evaluation observations lie within the L-day estimation sample, ensuring that the QLIKE criterion is computed in-sample at t under common (fixed) parameter estimates.

The final specification (7) preserves the interpretability of the standard HAR structure while expanding its information set to include an extensive set of high-frequency, economically grounded realized factor variances. By combining a log-linear specification with adaptive factor selection, the model remains tractable and scalable for large cross-sections yet flexible enough to capture multi-factor volatility dynamics, effectively unifying autoregressive persistence with a multiplicative factor structure.

To motivate the model structure, we document persistence in the key variance components. As established by Ding et al. (2025) and Herskovic et al. (2016), both the cross-sectional realized variance (CRV) and the multiplicative residual component exhibit substantial long-memory characteristics, reflecting persistent temporal dependencies in asset-specific and market-level volatility dynamics. We extend this analysis to the realized factor variances (FRV) used in our framework. We investigate whether a similar degree of persistence characterizes the realized variances of the high-frequency asset pricing factors included in our forecasting framework.

For each factor in our dataset, we compute the first-order autocorrelation of daily realized variances. Figure 2 presents the empirical distribution of these autocorrelations, while Table 1 reports its summary statistics. Consistent with the hypothesis of long-memory behavior, we find that factor-level realized variances are highly persistent, with a median autocorrelation of 0.815 and an interquartile range spanning from 0.776 to 0.846. These findings corroborate the presence of a strong autoregressive structure in factor volatilities, and support the inclusion of FRV as a dynamic predictor in volatility forecasting models.

The single-factor framework naturally generalizes to the two or three most informative factor variances for each stock. Given the candidate set $\mathcal{K}$, let $\mathcal{K}^{\star}_{i,t} = \{k^{\star}_{i,t,1}, k^{\star}_{i,t,2}, \ldots, k^{\star}_{i,t,K^{\star}}\} \subseteq \mathcal{K}$ denote the ordered set of factors selected for stock i at date t, where $K^{\star} = \mathrm{card}(\mathcal{K}^{\star}_{i,t}) \in \{2, 3\}$ is the number of selected factors. The elements are ordered by in-sample QLIKE loss with lowest first so that $k^{\star}_{i,t,1}$ is the top-ranked factor.

In this multi-factor framework, the forecasting equation (7) becomes

$$\log RV_{i,t+h} = \beta_0 + \sum_{z\in\{d,w,m\}} \left(\beta^{(z)}_{CRV} \log CRV^{(z)}_t + \beta^{(z)}_{\xi} \log \xi^{(z)}_{i,t} + \sum_{k\in\mathcal{K}^{\star}_{i,t}} \beta^{(z)}_k \log FRV^{(z)}_{k,t} \right) + \varepsilon_{i,t+h}. \tag{9}$$

Table 1: Summary statistics of first-order autocorrelation for factor-level realized variances. The reported values are the 25% quantile (Q1), the median, and the 75% quantile (Q3) of the autocorrelation across the factors under construction.

	Q1	Median	Q3
$\text{corr}(\text{FRV}_t, \text{FRV}_{t-1})$	0.776	0.815	0.846

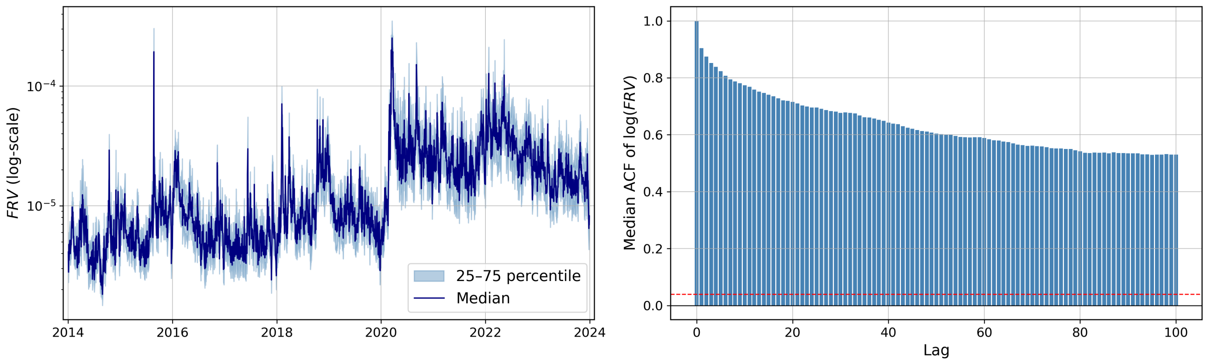

Figure 2: Median persistence of factor realized variances. The left panel plots the daily first-order autocorrelation with its interquartile range (25–75%), while the right panel shows the median autocorrelation function across all high-frequency factors. The red dashed line marks the 5% critical value.

We restrict $K^\star$ to $\{2, 3\}$ in order to balance explanatory power, estimation stability, and interpretability; because realized factor variances share considerable information, performance saturates with only a few top factors. The selection routine proceeds iteratively, searching for the next volatility predictor that adds the most explanatory power. First, we scan the full universe of candidate series and retain the factor $k^\star_{i,t,1}$ that delivers the lowest in-sample QLIKE (or RMSE) loss over the rolling selection window S. Holding $k^\star_{i,t,1}$ fixed, we re-estimate the model after adding each remaining candidate one at a time and select the factor $k^\star_{i,t,2}$ that achieves the largest additional loss reduction; this completes the specification with $K^\star = 2$. If a third predictor is allowed, we repeat the exercise once more: evaluate every unused candidate conditional on $\{k^\star_{i,t,1}, k^\star_{i,t,2}\}$, and retain $k^\star_{i,t,3}$, i.e., the factor that yields the largest additional loss reduction. Thus, each additional factor variance is chosen strictly for the marginal improvement it brings.

Forecasts for time $t + h$ are formed at the close of day t using CRV_t, $\xi_{i,t}$, and $FRV_{k,t}$, $k \in \mathcal{K}^\star_{i,t}$, together with their weekly and monthly aggregates. All quantities are $\mathcal{F}_t$-measurable, so the procedure uses only information available by date t and does not use those beyond t.

2.3 Model Validity I: Adaptive Selection

We now establish the theoretical foundations of the log-HAR framework with adaptive factor volatilities introduced in Subsections 2.1 and 2.2. We prove in Theorem 1 that the QLIKE-based selector consistently identifies the true predictive factor, verifying the asymptotic correctness of the selection rule. We also show how the selection reliability is affected by the variance of measurement error proxied by realized quarticity in Theorem 2. The results motivate and support the validity of our use of robust realized-variance estimators and the factor-volatility block.

We suppose the true data generating process (DGP) for the log integrated variance is

$$\log(IV_{i,t}) = \beta_0 + \beta^{CRV} \log(CIV_t) + \beta^{\xi} \log(\xi_{i,t}^{IV}) + \sum_{k \in \mathcal{K}} \gamma_{k,t} \cdot \log(FIV_{k,t}) + \varepsilon_{i,t}, \tag{10}$$

where $CIV_t := N^{-1} \sum_i IV_{i,t}$, $\xi_{i,t}^{IV} := IV_{i,t}/CIV_t$, and

$$FIV_{k,t} := \int_{t-1}^{t} \sigma_{k,s}^2 \, ds \tag{11}$$

where $\sigma_{k,s}^2$ is the instantaneous variance of the factor's return at time s and $\gamma_{k,t}$ is the loading on $\log(FIV_{k,t})$, nonzero only for the unique active factor on the window, see Assumption A below. The observed quantities $\log(CRV_t)$, $\log(\xi_{i,t})$ and $\log(FRV_{k,t})$ are proxies for these latent terms. If factor k is a tradable linear portfolio with predictable weights $w_{k,s}$ on constituents with instantaneous covariance Σ_s, then $\sigma_{k,s}^2 = w_{k,s}' \Sigma_s w_{k,s}$ and $FIV_{k,t} = \int_{t-1}^{t} w_{k,s}' \Sigma_s w_{k,s} \, ds$. This justifies the interpretation of $FRV_{k,t}$ as a high-frequency proxy for the latent quantity in the data generating process.

Notations and Preliminaries. We introduce some notations we shall use throughout. Fix a stock i, a forecast origin t, and a forecasting horizon h, and recall that $\mathcal{K}$ is the index set for the candidate factors. In our empirical application $K = 287$, where $K = \text{card}(\mathcal{K}) < \infty$. For a candidate factor $k \in \mathcal{K}$, we define the parameter vector of the corresponding model as $\theta := (\beta_0, \beta_{CRV}, \beta_{\xi}, \beta_k) \in \Theta_k$ and the vector of predictors available at time s as $X_{i,s}^k := \big(\log(CRV_s), \log(\xi_{i,s}), \log(FRV_{k,s})\big)$, augmented with their weekly and monthly aggregates. We denote by $m(X_{i,s}^k; \theta)$ the log-linear predictor for stock i including factor k, so that

$$\log RV_{i,s+h} \;=\; m(X_{i,s}^k; \theta) \;+\; \varepsilon_{i,s+h}, \qquad s \in \{t-S, \dots, t-1\}, \tag{12}$$

cf. (7). Let the level target and its implied forecast be, respectively, $V_{i,s} := RV_{i,s+h} > 0$ and $\widehat{V}_{i,s}^k(\theta) := \exp\{\, m(X_{i,s}^k; \theta) \,\} > 0$. We refer to the single-observation QLIKE loss using

$$\ell\Big(V_{i,s}, X_{i,s}^k; \theta\Big) \;:=\; \log\left(\frac{\widehat{V}_{i,s}^k(\theta)}{V_{i,s}}\right) + \frac{V_{i,s}}{\widehat{V}_{i,s}^k(\theta)} - 1, \tag{13}$$

which is well-defined since $V_{i,s} > 0$ and $\widehat{V}_{i,s}^k(\theta) > 0$, and write the sample loss

$$L_{i,t}(k; \theta) \;=\; \frac{1}{S} \sum_{s=t-S}^{t-1} \ell\Big(V_{i,s}, X_{i,s}^k; \theta\Big). \tag{14}$$

Finally, let

$$\mathcal{L}_{i,t}(k) \;:=\; \inf_{\theta\in\Theta_k} \mathbb{E}\Big[\,\ell\Big(V_{i,s}, X^k_{i,s};\theta\Big)\;|\,\mathcal{F}_{t-1}\Big]$$

denote the minimized (conditional) population loss for factor $k \in \mathcal{K}$.

We impose the following regularity conditions:

Assumption A.

A1. *For each stock i and time t, there is a single true active factor, $k^\star_{i,t}$, which remains unchanged over $\{t-S,\ldots,t-1\}$. That is, the set of true non-zero coefficients $\mathcal{S}_{i,t} = \{k \in \mathcal{K}; \gamma_{k,t} \neq 0\}$ has only one element so that card$(\mathcal{S}_{i,t}) = 1$.*

A2. *$\gamma_{k^\star_{i,s}}$ is uniformly bounded away from zero by some positive constant C*

$$\inf_{s\in[t-S,t-1]} |\gamma_{k^\star_{i,s}}| \geq C > 0. \tag{15}$$

A3. *Let $\tilde{Z}_{i,s} = (1, \log(CRV_s), \log(\xi_{i,s}))$. Let $u_{k^\star_{i,s}}$ and $u_{k,s}$ be the residuals from the population regressions of $\log(FRV_{k^\star_{i,s}})$ and $\log(FRV_{k,s})$ on $\tilde{Z}_{i,s}$, respectively. Then, for some $\kappa \in [0,1)$,*

$$\max_{k\in\mathcal{K}\setminus\{k^\star_{i,s}\}} \Big|\mathrm{corr}(u_{k^\star_{i,s}}, u_{k,s})\Big| \leq \kappa < 1. \tag{16}$$

A4. *(i) $\{(V_{i,s}, X^k_{i,s})\}$ is strictly stationary and α-mixing with $\alpha(\ell) = O(\ell^{-c})$ for some $c > 2$; (ii) for each $k \in \mathcal{K}$, Θ_k is compact and $\theta \mapsto \ell(V_{i,s}, X^k_{i,s}, \theta)$ is measurable and continuous a.s.; (iii) there exists an envelope $M_{i,s}$ with $|\ell(V_{i,s}, X^k_{i,s}, \theta)| \leq M_{i,s}$ and $\mathbb{E}[M_{i,s}] < \infty$.*

Remark. Assumption A1 imposes a unique, locally stable volatility driver within the selection window, which is the minimal identification content needed for a single-factor selector. Assumption A2 is a standard signal-strength condition ensuring that the contribution of the true factor does not vanish on the window. Assumption A3 is a restricted non-collinearity requirement: after partialling out the common block $\tilde{Z}_{i,s} = (1, \log(CRV_s), \log(\xi_{i,s}))$, the innovation in the true factor cannot be replicated by any inactive factor; this rules out near-collinearity of the factor-specific signals once the common predictors are controlled for. Under the log-DGP in (10) and the QLIKE loss $\ell(\cdot)$, it follows that A2 and A3 jointly imply strict separation of the population risks: there exists $\Delta_{\min} > 0$ such that

$$\mathcal{L}_{i,t}(k) - \mathcal{L}_{i,t}(k^\star_{i,t}) \geq \Delta_{\min} \quad \text{for all} \quad k \in \mathcal{K} \setminus \{k^\star_{i,t}\}. \tag{17}$$

Assumption A4 provides a uniform law of large numbers over $k \in \mathcal{K}$ and $\theta \in \Theta_k$:

stationarity/mixing, compact parameter spaces, and an integrable envelope imply

$$\max_{k\in\mathcal{K}} \sup_{\theta\in\Theta_k} \left| L_{i,t}(k;\theta) - \mathbb{E}\left[\ell(V_{i,s}, X_{i,s}^k, \theta)\right] \right| \xrightarrow{p} 0, \tag{18}$$

so the sample criteria deviate uniformly little from their population counterparts, see Andrews (1987), Davidson (1994). Consequently, with K fixed and finite, the sample argmin equals the population argmin with probability tending to one, which is the main content of Theorem 1 below.

The strong consistency of our selection procedure is now formally presented. The theorem proves that our QLIKE-based selection rule asymptotically identifies the factor that truly drives volatility. With sufficiently long windows, the probability that the procedure picks the true factor converges to one, as desired.

Theorem 1. *Let $\mathcal{K}$ be the finite set of candidate factors. Suppose $\widehat{k}_{i,t}^{\star}$ is chosen at time t for stock i according to the minimum QLIKE loss criterion for the model with factor k. That is, for each $k \in \mathcal{K}$ and $\theta \in \Theta_k$, define*

$$L_{i,t}(k;\theta) := \frac{1}{S}\sum_{s=t-S}^{t-1} \ell(V_{i,s}, X_{i,s}^k, \theta) \qquad \textit{and} \qquad \hat{\theta}_k \in \arg\min_{\theta\in\Theta_k} L_{i,t}(k;\theta), \tag{19}$$

and let

$$\widehat{k}_{i,t}^{\star} \in \operatorname*{argmin}_{k\in\mathcal{K}} L_{i,t}(k;\hat{\theta}_k). \tag{20}$$

If Assumptions $A1 - A4$ hold, then the selection procedure identifies the true active factor with probability tending to one, i.e., as $S \to \infty$ with $S = o(T)$,

$$\mathbb{P}\Big(\widehat{k}_{i,t}^{\star} = k_{i,t}^{\star}\Big) \ \to \ 1, \tag{21}$$

where $k_{i,t}^{\star}$ denotes the unique active factor of (10) in the window.

Proof. See Appendix A. □

2.4 Model Validity II: Adaptive Selection under Measurement Error

We now examine how *measurement error in the factor-volatility block* affects the adaptive selection in Subsection 2.3. In our empirical implementation, realized factor variance $FRV_{k,t}$ is a noisy proxy for the latent factor integrated variance $FIV_{k,t}$, and the magnitude of this noise is time varying and captured by realized quarticity (FRQ). Theorem 2 below shows that measurement error *shrinks* the population loss gap that drives the selection: the higher the (conditional) variance of the measurement error for the *true* factor, the lower the probability that the QLIKE-based selector picks it. This provides a formal rationale for using robust high-frequency estimators and for including the factor-volatility block.

We impose the following conditions:

Assumption B. ***B1.*** *For each factor $k \in \mathcal{K}$ and each s,*

$$\log(FRV_{k,s}) = \log(FIV_{k,s}) + e_{k,s},$$

with $\mathbb{E}[e_{k,s} \mid \mathcal{F}_{s-1}] = 0$, $\mathrm{Var}(e_{k,s} \mid \mathcal{F}_{s-1}) = \sigma^2_{k,s}$, and $\mathbb{E}[|e_{k,s}|^{4+\delta}] < \infty$ for some $\delta > 0$. Moreover, $e_{k,s}$ is conditionally uncorrelated with $\tilde{Z}_{i,s} = (1, \log(CRV_s), \log(\xi_{i,s}))$ and with $\log(FIV_{k',s})$ for $k' \neq k$.

B2. *There exist positive constants $0 < c_- \leq c_+ < \infty$ such that*

$$c_- \, FIQ_{k,s} \;\leq\; \sigma^2_{k,s} \;\leq\; c_+ \, FIQ_{k,s},$$

where $FIQ_{k,s}$ is the factor integrated quarticity, for which $FRQ_{k,s}$ is consistent in the sense that $\sup_k \left|FRQ_{k,s}/FIQ_{k,s} - 1\right| = o_p(1)$ uniformly in k.

B3. *Let $\bar{\sigma}^2_{k^\star_{i,t}} := S^{-1} \sum_{s=t-S}^{t-1} \sigma^2_{k^\star_{i,s}}$. There exists $\bar{\sigma}^2_0 > 0$ such that $\bar{\sigma}^2_{k^\star_{i,t}} \leq \bar{\sigma}^2_0$, and $\bar{\sigma}^2_0$ can be made arbitrarily small.*

B4. *For each $k \in \mathcal{K}$, the population risk $R_{i,t}(k, \theta) := \mathbb{E}[\ell(V_{i,s}, X^k_{i,s}, \theta)]$ is twice continuously differentiable in the linear predictor $m(X^k_{i,s}; \theta)$ in a neighborhood of the optimum, with a uniform lower curvature bound: there exists $c_0 > 0$ such that for all k,*

$$R_{i,t}(k, \theta) - \inf_{\vartheta \in \Theta_k} R_{i,t}(k, \vartheta) \;\geq\; c_0 \, \mathbb{E}\Big[\big(m(X^k_{i,s}; \theta) - m(X^k_{i,s}; \theta^\circ_k)\big)^2\Big],$$

where $\theta^\circ_k \in \arg\min_{\vartheta \in \Theta_k} R_{i,t}(k, \vartheta)$ under latent inputs.

Remark. The assumptions are mild and specify a weak set of conditions under which the theory is valid. Assumption B1 specifies an errors-in-variable structure, and B2 ties the conditional variance of the error to quarticity. Assumption B3 bounds the window-average noise variance and B4 gives strong convexity of the population QLIKE risk in the log predictor.

We have the following result:

Theorem 2. *Under Assumptions A1-A4 and B1-B4, for*

$$\bar{\sigma}^2_{k^\star_{i,t}} := \frac{1}{S} \sum_{s=t-S}^{t-1} \sigma^2_{k^\star_{i,s}} \qquad \textit{and} \qquad H_{i,t}(S) := \max_{k \in \mathcal{K}} \sup_{\theta \in \Theta_k} \left|L_{i,t}(k; \theta) - R_{i,t}(k, \theta)\right|,$$

there exist a constant independent of S, denoted $C_1 > 0$, such that the following holds: for any incorrect $k \neq k^\star_{i,t}$,

$$R^\eta_{i,t}(k) - R^\eta_{i,t}(k^\star_{i,t}) \;\geq\; \Delta_{\min} - C_1\,\bar{\sigma}^2_{k^\star_{i,t}} + o(\bar{\sigma}^2_{k^\star_{i,t}}), \tag{22}$$

where $\Delta_{\min}$ is as in (17) and $R^\eta_{i,t}(\cdot)$ denotes the population risk evaluated with the noisy factor inputs. In addition, we have

$$\mathbb{P}\Big(\widehat{k}^\star_{i,t} = k^\star_{i,t}\Big) \;\geq\; \mathbb{P}\Big(H_{i,t}(S) \;<\; \tfrac{1}{2}\big[\Delta_{\min} - C_1\,\bar{\sigma}^2_{k^\star_{i,t}}\big]\Big). \tag{23}$$

In particular, if $\bar{\sigma}^2_{k^\star_{i,t}} < \Delta_{\min}/C_1$, then $R^\eta_{i,t}(k) > R^\eta_{i,t}(k^\star_{i,t})$ for all $k \neq k^\star_{i,t}$, and as $S \to \infty$ with $S = o(T)$, we have

$$\mathbb{P}\Big(\widehat{k}^\star_{i,t} = k^\star_{i,t}\Big) \;\to\; 1. \tag{24}$$

Proof. *See Appendix A.* □

3 Data

We assemble a comprehensive high-frequency dataset of U.S. equities to support the estimation of realized variance under infill asymptotics. The sample contains all CRSP common shares (share codes 10 or 11) listed on the NYSE, NASDAQ or AMEX (main exchange codes 1, 2 or 3) from January 2, 2014 to December 29, 2023, covering $T = 2516$ trading days and $N = 5370$ unique securities.

Intraday trades and quotes are sourced from the NYSE TAQ database via WRDS and restricted to regular trading hours (09:30–16:00 Eastern Time).[4] Following standard practice, we apply the filters of Barndorff-Nielsen et al. (2009) to remove observations outside regular hours, zero or negative prices, obvious quote errors, and extreme outliers. Additional safeguards remove FINRA Alternative Display Facility prints (exchange "D") and delete trades priced beyond the daily CRSP ask-high or bid-low. Prices are then sampled on a uniform one-second grid between 09:30:00 and 16:00:00 ET using the previous-tick method of Gençay et al. (2001). The resulting high-frequency returns can naturally be aggregated to any lower frequency, allowing a straightforward implementation for any realized variance estimator.

Daily opens, closes, shares outstanding, and delisting returns are obtained from CRSP. We match TAQ symbols to CRSP identifiers (PERMNOs) via the TAQ–CRSP Link table (covering about 98% of the universe).[5] We prioritize CRSP entries: each CRSP record is retained even if no intraday data exist that day, while unmatched TAQ observations are ignored.[6] To ensure

[4]We use the SAS code from Holden and Jacobsen (2014) to extract tick-by-tick transactions matched with contemporaneous bid–ask quotes from daily TAQ. Timestamps are recorded in milliseconds until mid-2015 and microseconds thereafter.

[5]Residual cases are matched by eight-digit CUSIP from the TAQ Master file.

[6]Unmatched observations involve fewer than 1% of stocks, almost all nano-caps, so they have little effect on weighted portfolio returns.

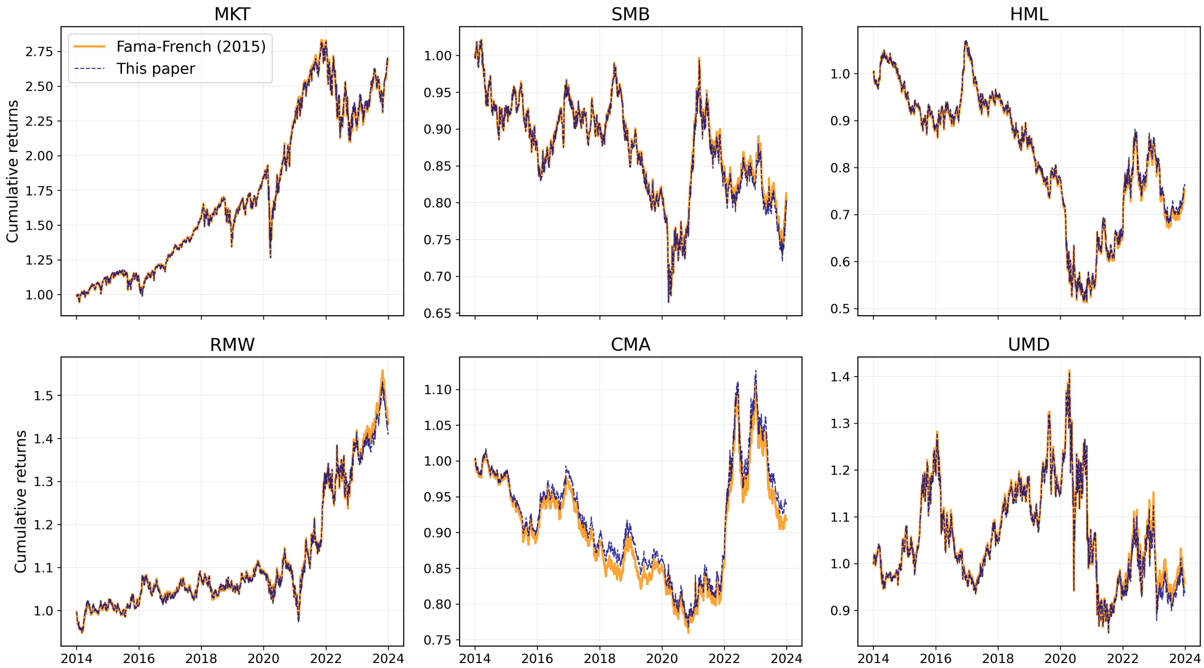

Figure 3: Cumulative gross daily returns of the Fama–French six factors using official data and our high-frequency replication. The orange line corresponds to daily returns sourced from the Kenneth R. French Data Library, while the blue dashed line represents our version constructed from 1-second returns aggregated to daily frequency.

consistency around corporate events, we overwrite the 09:30 and 16:00 TAQ prices with the CRSP open and close, and we adjust the closing return when delisting returns occur, in line with Hou et al. (2018).

3.1 High-frequency factors

Leveraging intraday stock information, we replicate a universe of $K = 287$ high-frequency factors. The first block contains the six canonical Market (MKT), Size (SMB), Value (HML), Profitability (RMW), Investment (CMA), and Momentum (UMD) factors, replicated to closely match the definitions in Fama and French (2018). The second block spans 281 characteristic-sorted portfolios drawn from the large collections of Jensen et al. (2023) (JKP) and Chen and Zimmermann (2022) (CZ).[7]

To replicate the Fama–French factors at high frequency, we follow the standard definitions with NYSE breakpoints, double-sorting on size and annual rebalancing in June. Adapting the procedure of Aït-Sahalia et al. (2020), we update value-weighted stock returns at 1-second frequency, and compute portfolio returns. Figure 3 shows that our daily aggregation of the replicated factors is virtually indistinguishable from the official data.

The remaining portfolios are constructed from a large cross-section of firm characteristics.[8] At

[7]As the Fama–French factors are widely used as benchmarks, we replicate them as faithfully as possible. By contrast, the broader zoo of factors is built under a uniform methodology to ensure comparability across signals.

[8]To avoid duplication where JKP and CZ provide conceptually similar characteristics, we retain the JKP implementation whenever the two series are empirically equivalent (return correlation > 95%) and drop signals with missing data, yielding 153 JKP and 128 CZ unique factors.

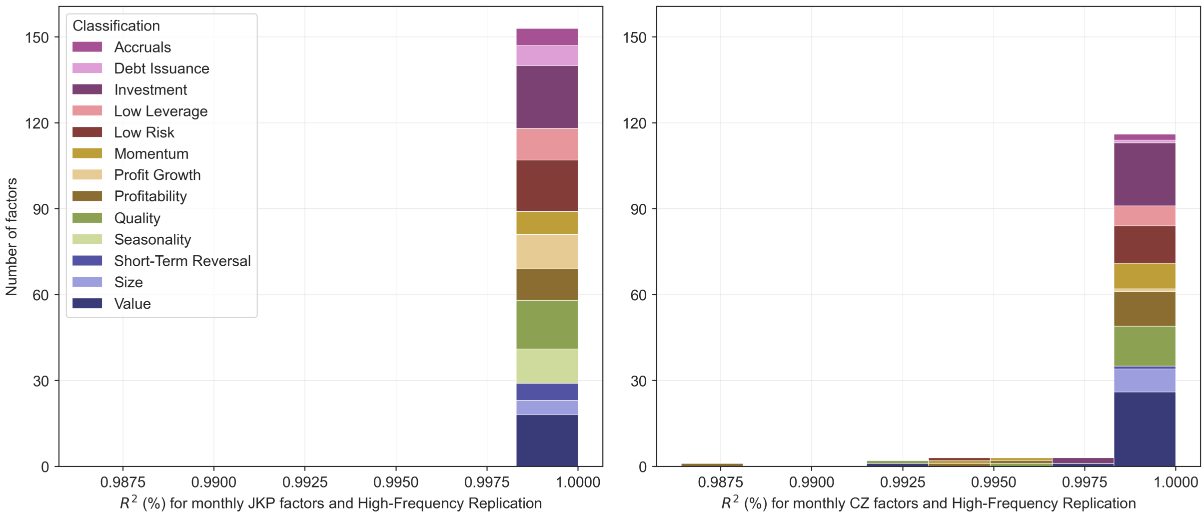

Figure 4: Histogram of R^2 from factor-level monthly return regressions. For each JKP (left) and selected CZ (right) factor, we regress our aggregated high-frequency portfolio returns on the original low-frequency portfolio over 2014–2023. Results are reported by economic cluster using the Jensen et al. (2023) taxonomy: CZ factors are mapped to clusters by the highest average correlation between their CAPM-residual returns and those of the JKP factors within each cluster. Higher R^2 indicates closer replication fidelity.

the end of each month, eligible stocks are sorted into terciles by the given characteristic. Following the empirical design in Jensen et al. (2023), we compute value-weighted returns for the top and bottom terciles and form a zero-investment high-minus-low spread held over the subsequent month. To validate our procedure, we compare our replication against the original low-frequency versions by examining the explanatory power of monthly return regressions. Figure 4 summarizes the comparison across JKP and CZ factors, demonstrating high fidelity for the whole replication.

A compendium of the sample factors and more details on the dataset are reported in Appendix B. To ensure comparability with the literature and computational tractability, we compute CRV_t and related quantities using the full-universe definitions as in Ding et al. (2025). As the factors and CRV_t are constructed from the same stock universe as the dependent variable, stock i mechanically enters both CRV_t and the factor realized variance $(FRV_{k,t})$ used to forecast $RV_{i,t+h}$. This own-observation inclusion has negligible impact on forecasts: the leave-one-out measure $CRV_t^{(-i)} = (N-1)^{-1}\sum_{j\neq i} RV_{j,t}$ differs from CRV_t by at most $O(1/N)$, and the factor portfolios are well diversified. Accordingly, we treat the effect as negligible.

4 Empirical evidence

The primary interest of the empirical analysis is the out-of-sample forecasting performance of the proposed model. The results of our framework are tested using twelve alternative realized volatility estimators, selected to encompass a broad spectrum of methodological approaches and robustness features. These include the classical realized variance (RV) of Andersen and Bollerslev (1998) at 5-minute and 1-minute frequency, and its 5-minute subsampled version. We consider the bipower variation (BPV) introduced by Barndorff-Nielsen and Shephard (2004) and its staggered definition (Andersen et al., 2007), the realized kernel developed by Barndorff-Nielsen et al. (2008), the truncated realized variance (TRV) of Mancini (2009) and the pre-averaged measures (PRV, PBV) of Christensen et al. (2014). We also include more recent contributions like the differenced-return variance (DV) of Andersen et al. (2023) and the nonparametric price duration variance (NPDV) proposed by Hong et al. (2023).

While all results are evaluated across the complete set of estimators to ensure robustness, we adopt the candlestick variance (or wick variance, WV) by Li et al. (2025) as the benchmark throughout the analysis due to its desirable properties. Specifically, the WV estimator provides resilience against microstructure noise and accommodates jumps and extreme price movements, remaining unbiased and robust under a wide range of market conditions. These features make it especially well-suited for high-frequency volatility estimation and motivate its primary role in our results.

Our forecasts are constructed using a recursive estimation approach. For each stock i and day t, we estimate the coefficients of the model in 9 by ordinary least squares over a rolling window of length $L = 1260$. We generate h-day-ahead forecasts for the log-realized variance, $\log \widehat{RV_{i,t+h}}$, using the estimated parameters and the daily, weekly, and monthly lags of the model explanatory components: the market-wide realized variance (CRV_t), the corresponding idiosyncratic component ($\xi_{i,t}$), and the log realized variance of the factors ($FRV_{k^\star_{i,t}}$) chosen for the best performance over the selection window $S = 252$.

We assess forecasts across three different forecasting horizons $h \in \{1, 5, 22\}$, corresponding to predictions for one trading day, week and month ahead. Forecast accuracy is assessed using the quasi-likelihood (QLIKE) loss function, defined by Patton (2011) and formulated by Bollerslev et al. (2016) as

$$QLIKE^i = \frac{1}{Q} \sum_{q=1}^{Q} \left(\log \frac{\widehat{RV}_q^i}{RV_q^i} + \frac{RV_q^i}{\widehat{RV}_q^i} - 1 \right), \tag{25}$$

where $\widehat{RV}_q^i$ and RV_q^i denote the forecasted and realized variances of stock i at horizon q, and Q is the number of out-of-sample observations.

4.1 Forecasting performance

We evaluate the forecast accuracy over the sample of $N = 1041$ CRSP stocks having less than 20% zero 5-minute returns between 2014 and 2023, with an evaluation period of $Q = 1213$ trading days. Realized variance is measured using the candlestick estimator of Li et al. (2025). Our factor-augmented specifications in (9) are denoted by HAR-1F, HAR-2F, and HAR-3F, corresponding respectively to $\mathcal{K}^\star = 1, 2, 3$. We compare their forecasting performance against widely adopted benchmarks, namely the standard HAR by Corsi (2009), the quarticity-adjusted

Table 2: Cross-sectional QLIKE loss distributions of different forecasting models and horizons, using the candlestick variance estimator of Li et al. (2025). Entries report the 25th percentile (Q1), mean, median (50th), and 75th percentile across $N = 1041$ stocks over an evaluation period of $Q = 1213$ trading days. In every row, bold entries indicate the lowest value.

	Forecasting models								
	HAR	SHAR	HARQ	HAR-MKT	MFV-CRV	MFV-PC1	HAR-1F	HAR-2F	HAR-3F
Panel A: $h = 1$									
Q1	0.1406	0.1414	0.1372	0.1415	0.1359	0.1340	0.1293	**0.1288**	**0.1288**
Median	0.1701	0.1738	0.1668	0.1728	0.1650	0.1632	0.1570	**0.1559**	0.1565
Mean	0.1816	0.1867	0.1779	0.1844	0.1767	0.1738	0.1674	0.1666	**0.1665**
Q3	0.2142	0.2218	0.2102	0.2175	0.2077	0.2041	0.1961	**0.1946**	0.1949
Panel B: $h = 5$									
Q1	0.2335	0.2427	0.2328	0.2344	0.2286	0.2228	0.2003	0.1950	**0.1904**
Median	0.2740	0.2827	0.2719	0.2747	0.2630	0.2563	0.2325	0.2246	**0.2202**
Mean	0.2900	0.2987	0.2865	0.2897	0.2768	0.2715	0.2455	0.2373	**0.2313**
Q3	0.3270	0.3363	0.3230	0.3269	0.3074	0.2990	0.2746	0.2656	**0.2573**
Panel C: $h = 22$									
Q1	0.4920	0.4988	0.4927	0.4877	0.4841	0.4860	0.3434	0.3096	**0.2654**
Median	0.5941	0.5967	0.5928	0.5834	0.5841	0.5797	0.3972	0.3527	**0.3039**
Mean	0.6179	0.6218	0.6165	0.6105	0.6065	0.6079	0.4133	0.3644	**0.3145**
Q3	0.7061	0.7163	0.7033	0.6997	0.6948	0.6993	0.4627	0.4094	**0.3498**

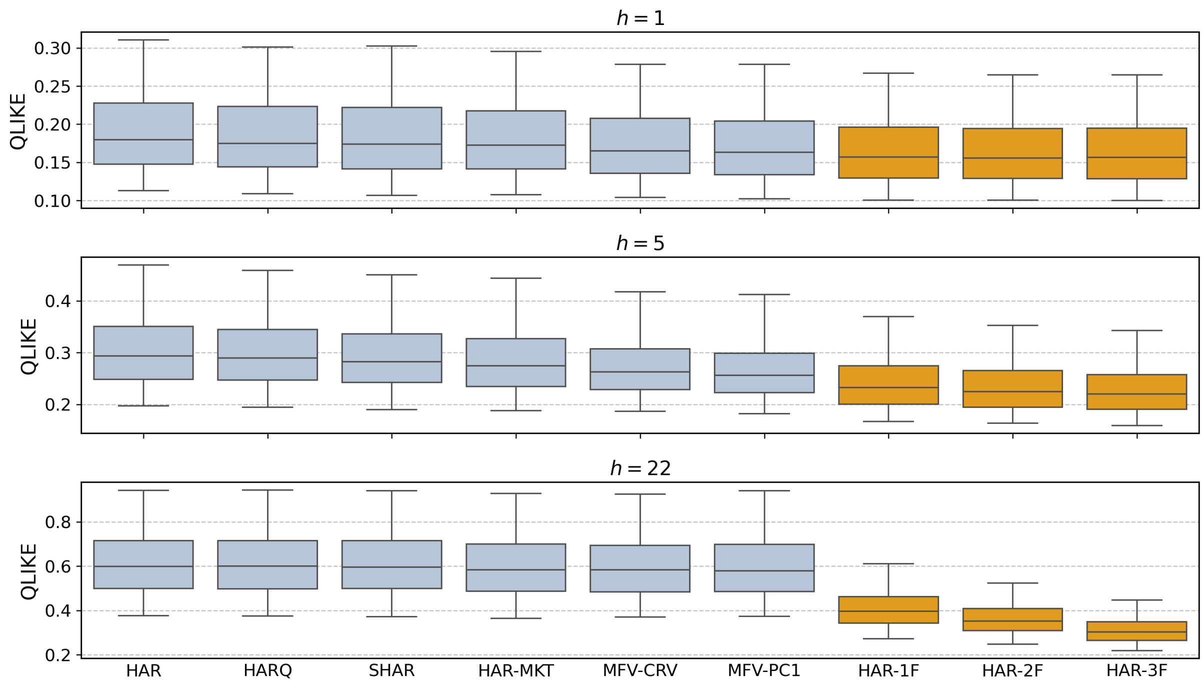

Figure 5: Distribution of QLIKE losses across models and forecast windows. Each boxplot summarizes the QLIKE values across all stocks using the WV estimator (Li et al., 2025). The models include traditional benchmarks (HAR, HARQ, SHAR), the market-HAR, the MFV model with common RV and PC1, the proposed one-factor model and its multi-factor extensions. Boxes represent the 25th, 50th, and 75th percentiles, while whiskers extend to the 1st and 99th percentiles.

HARQ of Bollerslev et al. (2016) and the asymmetric SHAR (Patton and Sheppard, 2015). We also include the market augmented HAR of Hizmeri et al. (2022) and the multiplicative volatility by Ding et al. (2025) with common RV (MFV-CRV) and first principal component (MFV-PC1).

In Table 2, the models we propose attain the lowest QLIKE across every quantile of the cross-section. The performance strengthens with the number of selected factors and with the length of the forecasting horizon, with the three-factor model lowering the average loss metric by up to 48% relative to the best alternative. Figure 5 complements these results by displaying the full distributions. The factor specifications exhibit a clear downward shift and tighter interquartile ranges relative to all benchmarks, with compressed upper tails that indicate robustness to outliers. The improvements are monotone, confirming that the inclusion of multiple volatility drivers yield lower and more stable forecasting loss. Our findings remain qualitatively unchanged across all alternative estimators considered[9].

We next evaluate the model performance at the individual stock level and assess the statistical relevance of the observed forecast improvements. Specifically, we compare the QLIKE loss of the HAR-1F model against that of each benchmark across the stock universe[10]. We quantify the fraction of stocks for which our factor model yields a lower QLIKE as the raw outperformance

$$\text{Out.perf.} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left\{QLIKE^{i}_{HAR-1F} < QLIKE^{i}_{bm}\right\}, \tag{26}$$

where $QLIKE^{i}_{HAR-1F}$ and $QLIKE^{i}_{bm}$ denote the QLIKE losses of stock i for the proposed model and the benchmark model, respectively. To further examine the statistical significance of the forecast differentials, we implement the Diebold and Mariano (2002) test for each stock. We evaluate the null hypothesis of equal predictive accuracy using a one-sided test on the difference between the QLIKE values of the proposed and benchmark models.[11] Finally, we report the proportion of stocks for which our model delivers statistically significant forecast improvements

$$\text{Sig.Out.perf.} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left\{p_i < \alpha\right\}, \tag{27}$$

where p_i is the p-value of the DM test for stock i, and $\alpha = 5\%$ is the significance threshold.

Table 3 reports the proportion of cases for which our factor augmented model statistically outperforms each benchmark. The significative outperformances show pervasive gains for our one-factor specification relative to the traditional benchmarks across any volatility estimator and forecasting horizon. For the weekly horizon, more than 95% of stocks exhibit significantly lower QLIKE using the HAR-1F rather than the traditional HAR model or its univariate variants. At the monthly horizon, the improvement increases to more than 99% in every case. At the daily horizon significance remains high but less pronounced, with differences depending on the

[9]Appendix C verifies robustness to the choice of realized-volatility estimator, forecasting horizon $h \in \{1, 5, 22\}$, and scoring rule.

[10]Figure 5 indicates that the multi-factor specifications (HAR-2F, HAR-3F) achieve further reductions in QLIKE relative to HAR-1F and the benchmarks: the one-factor outputs therefore provide a conservative result.

[11]The null and alternative hypotheses are $H_0 : \mathbb{E}(QLIKE_{HAR-1F,t} - QLIKE_{bm,t}) \geq 0$ against $H_1 : \mathbb{E}(QLIKE_{HAR-1F,t} - QLIKE_{bm,t}) < 0$. The test statistic is computed as $\bar{d}/\hat{\sigma}(\bar{d})$, where $\bar{d} = \sum_{q=1}^{Q} d_q/Q$ is the average loss differential $d_q = QLIKE_{HAR-1F,q} - QLIKE_{bm,q}$, and $\hat{\sigma}(\bar{q})$ is its heteroskedasticity-autocorrelation consistent (HAC) standard error.

Table 3: Significative outperformance proportion of the one-factor model (HAR-1F) over the benchmark models across forecasting horizons and volatility estimators between 2014 and 2023. Entries report the percentages of stocks for which the QLIKE loss of the HAR-1F model is significantly lower than the benchmark performance at the 5% significance level of the Diebold–Mariano test.

	Volatility estimators											
Benchmarks	RV5	$RV5_{ss}$	RV1	BPV	BPV_{stag}	RK	TRV	PRV	PBV	NPDV	DV	WV
Panel A: $h = 1$												
HAR	94.7	94.0	92.8	95.8	95.5	97.1	96.0	96.3	96.4	95.6	96.6	96.8
HARQ	64.2	87.9	91.9	96.0	95.9	97.0	95.2	92.5	92.0	94.8	95.5	88.8
SHAR	78.4	93.7	84.1	79.8	82.2	78.0	90.5	73.0	71.2	96.0	84.6	75.2
HAR-MKT	83.7	84.4	85.4	88.0	87.2	89.3	91.3	88.0	90.6	64.6	89.5	90.9
MFV-CRV	64.1	59.6	62.3	65.5	61.5	76.9	57.6	75.6	73.6	57.8	56.4	78.9
MFV-PC1	45.0	92.0	41.8	95.3	95.4	56.2	95.2	96.3	96.9	94.9	96.0	53.1
Panel B: $h = 5$												
HAR	98.2	97.8	97.7	98.1	98.0	97.9	97.6	97.9	97.6	99.1	98.5	99.2
HARQ	97.0	97.8	97.6	98.2	98.2	98.0	97.6	97.4	97.0	99.0	98.0	98.3
SHAR	97.9	97.7	97.0	97.3	97.1	97.3	96.9	96.0	95.1	98.4	97.0	96.7
HAR-MKT	95.1	95.7	96.7	95.3	95.5	95.1	96.2	93.7	93.3	96.6	96.9	97.7
MFV-CRV	84.9	83.0	87.4	85.8	84.8	87.1	86.7	85.3	85.1	93.3	87.9	93.1
MFV-PC1	77.0	92.6	79.0	93.3	93.0	78.2	92.0	94.0	93.7	96.8	93.2	85.1
Panel C: $h = 22$												
HAR	99.1	99.2	99.2	99.5	99.4	99.2	99.8	99.2	99.3	99.7	99.8	99.3
HARQ	99.1	99.2	99.2	99.5	99.5	99.2	99.8	99.2	99.2	99.6	99.8	99.3
SHAR	99.1	99.2	99.2	99.4	99.5	99.2	99.8	99.3	99.3	99.5	99.8	99.3
HAR-MKT	99.3	99.2	99.5	99.5	99.7	99.3	99.8	99.3	99.4	99.6	99.8	99.6
MFV-CRV	99.2	99.4	99.4	99.4	99.6	99.3	99.8	99.4	99.4	99.8	99.6	99.6
MFV-PC1	99.4	99.0	99.5	99.2	99.4	99.4	99.7	98.9	99.1	99.5	99.5	99.6

selected estimator. The pattern is stable across noise and jump robust measures, which indicates that the results are not driven by the choice of the realized specification. Overall, our model's superior statistical performance is robust across different estimators, benchmarks, and forecasting horizons: these results underscore the added value of flexible factor selection in capturing time-varying heterogeneity in volatility sources.

4.2 Return and volatility factor selections

We examine whether, within a broad factor universe with $K = 287$ characteristics, the same cross-sectional signal tends to drive both expected returns and realized variances for a given stock and day. For each (i, t), we compare the factor selected by an adaptive return model to the volatility driver selected by our volatility framework. The return selector is a one-factor specification with a cross-sectional component and a single characteristic factor return, chosen to minimize the sum of squared errors over the most recent $S = 252$ observations within a rolling window of length $L = 1260$. The volatility selector is the proposed HAR-1F model.

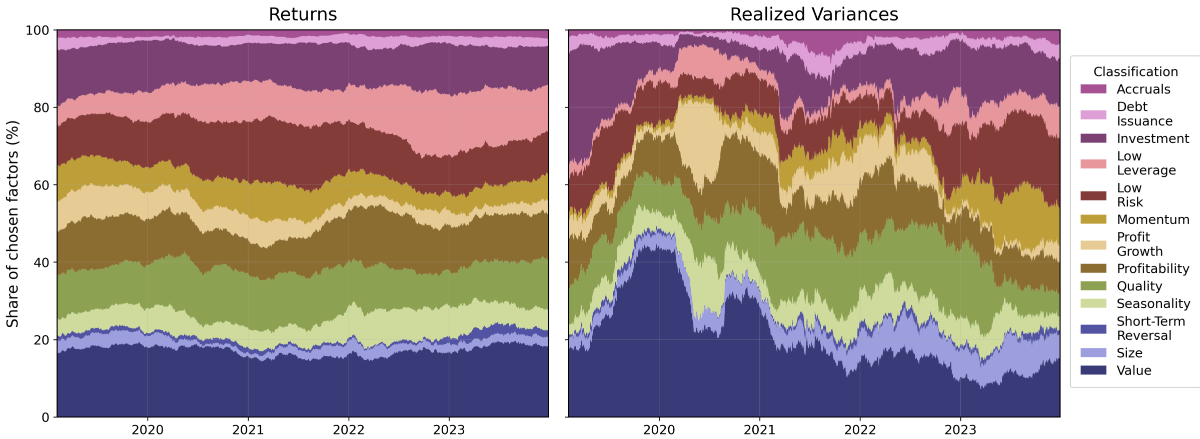

Figure 6: Time series of cluster shares implied by the adaptive selectors. Left: share of stocks whose return model selects a factor from each cluster. Right: corresponding shares for the HAR-1F volatility model using the candlestick estimator of Li et al. (2025). Stacked areas sum to 100% per day over a period of length $Q = 1213$ trading days.

We map each factor to one of thirteen economic clusters using the taxonomy of Jensen, Kelly and Pedersen (2023). JKP portfolios inherit their original labels, while CZ portfolios are assigned to clusters by the highest average correlation between their CAPM-residual returns and those of the JKP factors within each cluster. Figure 6 displays, for each trading day, the distribution across clusters of the factor selected by the return model (left) and by the HAR-1F volatility model (right), expressed as the share of stocks selecting a factor from each cluster; the stacked areas sum to 100% per day.

To quantify alignment at the stock–day level, let $k^R_{i,t}$ denote the factor chosen by the return selector and $k^V_{i,t}$ the factor chosen by the volatility selector, and let $c(k) \in \{1, \dots, 13\}$ map factors to clusters. Define the daily proportions

$$p_t^{\text{factor}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{\, k^R_{i,t} = k^V_{i,t} \,\} \qquad \textit{and} \qquad p_t^{\text{cluster}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{\, c(k^R_{i,t}) = c(k^V_{i,t}) \,\}.$$

Across the sample, the median of p_t^{factor} is below 1%, and the median of p_t^{cluster} is below 10%, indicating that exact factor matches and even same-cluster matches are rare in the cross-section on any given day. Consequently, similarities between the left and right panels of Figure 6 reflect aggregate shifts in cluster composition across stocks, not synchronized selections for the same stocks. These findings are consistent across forecasting horizons. When the volatility factor is selected for weekly and monthly forecasts, the chosen clusters concentrate on a smaller subset, with the strongest concentration during high-volatility regimes. Results are qualitatively robust to alternative volatility estimators and to loss functions beyond QLIKE.

Table 4: Economic-value comparison between the HAR-1F model and the benchmarks across daily to monthly forecasting horizons, using the WV estimator of Li et al. (2025). For each benchmark, entries report the cross-sectional percentage of stocks for which HAR-1F delivers higher annualized utility than the benchmark (Out.perf.), the percentage for which the improvement is statistically significant at the 5% level using a two-sided Diebold–Mariano test (Sig.Out.perf.), and the percentage for which the utility difference exceeds 1 basis point per year (Diff $\geq$ 1 bp). Values in parentheses give the corresponding percentages in which the benchmark outperforms the HAR-1F model under the same criterion.

Benchmarks	Out.perf. (%)	Sig.Out.perf. (%)	Diff $\geq$ 1 bp (%)
Panel A: $h = 1$			
HAR	99.3 (0.7)	96.8 (0.0)	98.4 (0.5)
HARQ	97.5 (2.5)	88.8 (0.5)	94.1 (1.5)
SHAR	91.9 (8.1)	75.2 (0.8)	88.0 (3.6)
HAR-MKT	99.2 (0.8)	90.9 (0.0)	94.4 (0.5)
MFV-CRV	93.9 (6.1)	78.9 (0.2)	82.7 (0.4)
MFV-PC1	85.0 (15.0)	53.1 (2.3)	68.8 (1.7)
Panel B: $h = 5$			
HAR	99.9 (0.1)	99.2 (0.1)	99.8 (0.1)
HARQ	99.9 (0.1)	98.3 (0.1)	99.8 (0.1)
SHAR	99.3 (0.7)	96.7 (0.1)	99.3 (0.2)
HAR-MKT	99.9 (0.1)	97.7 (0.1)	99.6 (0.1)
MFV-CRV	98.9 (1.1)	93.1 (0.2)	98.0 (0.5)
MFV-PC1	95.9 (4.1)	85.1 (1.2)	94.6 (2.3)
Panel C: $h = 22$			
HAR	99.9 (0.1)	99.3 (0.0)	99.8 (0.1)
HARQ	99.9 (0.1)	99.3 (0.0)	99.8 (0.1)
SHAR	99.8 (0.2)	99.3 (0.0)	99.8 (0.2)
HAR-MKT	99.9 (0.1)	99.6 (0.0)	99.8 (0.2)
MFV-CRV	100.0 (0.0)	99.6 (0.0)	99.9 (0.0)
MFV-PC1	100.0 (0.0)	99.6 (0.0)	99.9 (0.0)

4.3 Economic significance

Beyond statistical accuracy, an important question is whether improvements in volatility forecasting translate into economically meaningful gains. To address this, we adopt the utility-based evaluation framework introduced by Bollerslev et al. (2018), which links variance forecasts to the performance of volatility-managed investment strategies. Holding stock i, the investor's utility expressed in annualized percentage terms is

$$U^i = \frac{1}{Q} \sum_{q=1}^{Q} \left(8\% \sqrt{\frac{RV_q^i}{\widehat{RV}_q^i}} - 4\% \frac{RV_q^i}{\widehat{RV}_q^i} \right), \tag{28}$$

where RV_q^i and $\widehat{RV}_q^i$ denote, respectively, the realized variance on day q and its forecasted value. The first term captures the expected return component, inversely scaled by conditional volatility, while the second term penalizes variance under a mean-variance investor framework with a risk aversion level implied by a 4% volatility penalty.

Table 4 summarizes the economic performance of our factor-switching model relative to the considered benchmarks across all forecasting horizons. For each comparison, we report three measures: (i) the percentage of stocks for which the factor model yields higher utility than the benchmark (Out.perf.), (ii) the percentage of stocks for which the utility improvement is statistically significant at the 5% level based on a DM test (Sig.Out.perf.), and (iii) the percentage of stocks for which the utility difference exceeds 1 basis point annually, a threshold used to capture economically significant improvements.

The results demonstrate that the predictive gains of our model translate into substantial economic value. Across all benchmark comparisons and forecast horizons, the factor-switching model consistently improves investor utility for the large majority of stocks, with improvements that are both statistically significant and economically meaningful. These findings reinforce the practical value of incorporating adaptive factor selection in the modeling of stock volatility, and remind that more accurate volatility forecasts yield tangible benefits in portfolio outcomes.

5 Conclusion

This paper introduces a novel volatility forecasting framework that integrates high-frequency factor information with a dynamic selection mechanism to improve the prediction of individual stock variances. By extending a multiplicative volatility model with heterogeneous autoregressive lags and leveraging a broad cross-section of 287 factor volatilities, our approach achieves substantial gains in forecast accuracy relative to standard benchmarks.

The key contribution lies in demonstrating that volatility is not uniformly driven by a fixed set of risk sources but is instead shaped by time-varying, asset-specific exposures to distinct factor volatilities. In this sense, our model attempts to contribute to the methodological gap between volatility modeling and asset pricing.

Beyond the econometric gains, our results underscore the feasibility and informativeness of replicating factor portfolios at ultra-high frequencies, opening a path toward more granular assessments of systematic volatility. Future research could extend this framework to multivariate volatility modeling and examine the implications for portfolio risk management and derivative pricing.

References

Aït-Sahalia, Y., Jacod, J., 2014. High Frequency Financial Econometrics. Princeton University Press.

Aït-Sahalia, Y., Kalnina, I., Xiu, D., 2020. High-frequency factor models and regressions. Journal of Econometrics 216, 86–105.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International economic review , 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. The review of economics and statistics 89, 701–720.

Andersen, T.G., Li, Y., Todorov, V., Zhou, B., 2023. Volatility measurement with pockets of extreme return persistence. Journal of Econometrics 237, 105048.

Andrews, D.W.K., 1987. Consistency in nonlinear econometric models: A generic uniform law of large numbers. Econometrica 55, 1465–1471.

Asai, M., Caporin, M., McAleer, M., 2015. Forecasting value-at-risk using block structure multivariate stochastic volatility models. International Review of Economics & Finance 40, 40–50.

Atak, A., Kapetanios, G., 2013. A factor approach to realized volatility forecasting in the presence of finite jumps and cross-sectional correlation in pricing errors. Economics Letters 120, 224–228.

Barigozzi, M., Hallin, M., 2017. Generalized dynamic factor models and volatilities: estimation and forecasting. Journal of Econometrics 201, 307–321.

Barigozzi, M., Hallin, M., 2020. Generalized dynamic factor models and volatilities: Consistency, rates, and prediction intervals. Journal of Econometrics 216, 4–34.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of financial econometrics 2, 1–37.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics 31, 307–327.

Bollerslev, T., 2022. Realized semi (co) variation: Signs that all volatilities are not created equal. Journal of Financial Econometrics 20, 219–252.

Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H., 2018. Risk everywhere: Modeling and managing volatility. The Review of Financial Studies 31, 2729–2773.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192, 1–18.

Calvet, L.E., Fisher, A.J., Thompson, S.B., 2006. Volatility comovement: a multifrequency approach. Journal of econometrics 131, 179–215.

Chen, A.Y., Zimmermann, T., 2022. Open source cross-sectional asset pricing. Critical Finance Review 11, 207–264.

Chinco, A., Clark-Joseph, A.D., Ye, M., 2019. Sparse signals in the cross-section of returns. The Journal of Finance 74, 449–492.

Christensen, K., Oomen, R.C., Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. Journal of Financial Economics 114, 576–599.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159, 276–288.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. Journal of Business & economic statistics 20, 134–144.

Ding, Y., Engle, R., Li, Y., Zheng, X., 2025. Multiplicative factor model for volatility. Journal of Econometrics 249, 105959.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the econometric society , 987–1007.

Engle, R.F., Campos-Martins, S., 2023. What are the events that shake our world? measuring and hedging global covol. Journal of Financial Economics 147, 221–242.

Engle, R.F., Ito, T., Lin, W.L., 1988. Meteor showers or heat waves? heteroskedastic intra-daily volatility in the foreign exchange market.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of financial economics 116, 1–22.

Fama, E.F., French, K.R., 2018. Choosing factors. Journal of financial economics 128, 234–252.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. The Review of Financial Studies 33, 2326–2377.

Gençay, R., Dacorogna, M., Muller, U.A., Pictet, O., Olsen, R., 2001. An introduction to high-frequency finance. Elsevier.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Harvey, C.R., Liu, Y., Zhu, H., 2016. ... and the cross-section of expected returns. The Review of Financial Studies 29, 5–68.

Herskovic, B., Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2016. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. Journal of Financial Economics 119, 249–283.

Herskovic, B., Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2020. Firm volatility in granular networks. Journal of Political Economy 128, 4097–4162.

Hizmeri, R., Izzeldin, M., Nolte, I., Pappas, V., 2022. A generalized heterogeneous autoregressive model using market information. Quantitative Finance 22, 1513–1534.

Holden, C.W., Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. The Journal of Finance 69, 1747–1785.

Hong, S.Y., Nolte, I., Taylor, S.J., Zhao, X., 2023. Volatility estimation and forecasts based on price durations. Journal of Financial Econometrics 21, 106–144.

Hou, K., Xue, C., Zhang, L., 2018. Replicating anomalies. The Review of Financial Studies 33, 2019–2133.

Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes. Springer.

Jensen, T.I., Kelly, B., Pedersen, L.H., 2023. Is there a replication crisis in finance? The Journal of Finance 78, 2465–2518.

Kapadia, N., Linn, M., Paye, B., 2024. One vol to rule them all: Common volatility dynamics in factor returns. Journal of Financial and Quantitative Analysis 59, 1185–1212.

Li, Y., Nolte, I., Nolte, S., Yu, S., 2025. Realized candlestick wicks. Journal of Econometrics 250, 106014.

Luciani, M., Veredas, D., 2015. Estimating and forecasting large panels of volatilities with approximate dynamic factor models. Journal of Forecasting 34, 163–176.

Mancini, C., 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scandinavian Journal of Statistics 36, 270–296.

Nolte, I., Voev, V., 2012. Least squares inference on integrated volatility and the relationship between efficient prices and noise. Journal of Business & Economic Statistics 30, 94–108.

Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Ross, S.A., 2013. The arbitrage theory of capital asset pricing, in: Handbook of the fundamentals of financial decision making: Part I. World Scientific, pp. 11–30.

Sharpe, W.F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The journal of finance 19, 425–442.

Taylor, S.J., 1982. Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. Time series analysis: theory and practice 1, 203–226.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Zheng, X., Li, Y., 2011. On the estimation of integrated covariance matrices of high dimensional diffusion processes. The Annals of Statistics 39, 3121–3151.

Appendix to

Volatility Forecasting Factors

A Proof of main results

Proof of Theorem 1. Recall that, for any candidate factor $k \in \mathcal{K}$, stock i and time t, we write the parameter vector for the corresponding model as $\theta := (\beta_0, \beta_{CRV}, \beta_\xi, \beta_k) \in \Theta_k$, and

$$L_{i,t}(k;\theta) = \frac{1}{S} \sum_{s=t-S}^{t-1} \ell(V_{i,s}, X_{i,s}^k, \theta). \tag{1}$$

The estimated parameter vector, $\widehat{\theta}_k$, is the M-estimator that minimizes this sample loss:

$$\widehat{\theta}_k = \arg\min_{\theta \in \Theta_k} L_{i,t}(k;\theta). \tag{2}$$

Showing that $\mathbb{P}(\widehat{k}_{i,t}^\star = k_{i,t}^\star) \to 1$ as $S \to \infty$, where $k_{i,t}^\star$ is the index of the true active factor is equivalent to showing that for any incorrect factor $k \neq k_{i,t}^\star$, the probability of the event $\{L_{i,t}(k_{i,t}^\star; \widehat{\theta}_{k_{i,t}^\star}) < L_{i,t}(k;\widehat{\theta}_k)\}$ converges to 1.

We will prove that the difference $L_{i,t}(k;\widehat{\theta}_k) - L_{i,t}(k_{i,t}^\star; \widehat{\theta}_{k_{i,t}^\star})$ is strictly positive with probability approaching one.

Given $\mathcal{L}_{i,t}(k) = \inf_{\theta \in \Theta_k} \mathbb{E}[\ell(V_{i,s}, X_{i,s}^k, \theta)]$, the minimized population loss for model k, by Assumptions A1-A3 there exists a unique population minimizer at the true factor $k_{i,t}^\star$. Hence, there exists a fixed gap $\Delta_f \equiv \mathcal{L}_{i,t}(k) - \mathcal{L}_{i,t}(k_{i,t}^\star) > 0$ for all $k \neq k_{i,t}^\star$. Let $\Delta_{\min} = \min_{k \neq k_{i,t}^\star} \Delta_f > 0$.

Now we have,

$$\begin{aligned}
& L_{i,t}(k;\widehat{\theta}_k) - L_{i,t}(k_{i,t}^\star; \widehat{\theta}_{k_{i,t}^\star}) \\
&\quad = \left(\mathcal{L}_{i,t}(k) - \mathcal{L}_{i,t}(k_{i,t}^\star)\right) + \Big(L_{i,t}(k;\widehat{\theta}_k) - \mathcal{L}_{i,t}(k)\Big) - \Big(L_{i,t}(k_{i,t}^\star; \widehat{\theta}_{k_{i,t}^\star}) - \mathcal{L}_{i,t}(k_{i,t}^\star)\Big) \\
&\quad \geq \Delta_{\min} + \Big(L_{i,t}(k;\widehat{\theta}_k) - \mathcal{L}_{i,t}(k)\Big) - \Big(L_{i,t}(k_{i,t}^\star; \widehat{\theta}_{k_{i,t}^\star}) - \mathcal{L}_{i,t}(k_{i,t}^\star)\Big) \\
&\quad \geq \Delta_{\min} - \Big|L_{i,t}(k;\widehat{\theta}_k) - \mathcal{L}_{i,t}(k)\Big| - \Big|L_{i,t}(k_{i,t}^\star; \widehat{\theta}_{k_{i,t}^\star}) - \mathcal{L}_{i,t}(k_{i,t}^\star)\Big| \\
&\quad \geq \Delta_{\min} - 2 \cdot \max_{k \in \mathcal{K}} \sup_{\theta \in \Theta_k} \Big|L_{i,t}(k;\theta) - \mathbb{E}\,\ell(V_{i,s}, X_{i,s}^k, \theta)\Big| .
\end{aligned}$$

Let $H_{i,t}(S) \equiv \max_{k \in \mathcal{K}} \sup_{\theta \in \Theta_k} \Big|L_{i,t}(k;\theta) - \mathbb{E}\,\ell(V_{i,s}, X_{i,s}^k, \theta)\Big|$. By Assumptions A4, the uniform law of large numbers applies, Andrews (1987), Davidson (1994), and this maximum deviation converges in probability to zero:

$$H_{i,t}(S) \xrightarrow{p} 0 \quad \text{as} \quad S \to \infty. \tag{3}$$

Therefore, for any $\epsilon > 0$ and any $\delta > 0$, there exists a window size W_0 such that for all $S > W_0$, we have $\mathbb{P}(H_{i,t}(S) < \delta) > 1 - \epsilon$ by definition.

Choose $\delta = \Delta_{\min}/2$. Then for $S > W_0$, with probability greater than $1 - \epsilon$, we have:

$$L_{i,t}(k; \widehat{\theta}_k) - L_{i,t}(k^\star_{i,t}; \widehat{\theta}_{k^\star_{i,t}}) \;\geq\; \Delta_{\min} - 2 \cdot H_{i,t}(S) \;>\; \Delta_{\min} - 2 \cdot (\Delta_{\min}/2) \;=\; 0. \tag{4}$$

As a result, for any $k \neq k^\star_{i,t}$, the sample loss for model k is strictly greater than the sample loss for the true model $k^\star_{i,t}$ with probability approaching 1. Therefore, the minimizer of the sample loss must be $k^\star_{i,t}$:

$$\mathbb{P}\left(\operatorname*{argmin}_{k \in \mathcal{K}} L_{i,t}(k; \widehat{\theta}_k) = k^\star_{i,t}\right) \to 1 \quad \text{as} \quad S \to \infty. \tag{5}$$

This completes the proof. □

Proof of Theorem 2. Write $R^0_{i,t}(k, \theta)$ for the population risk with latent factor inputs and $R^\eta_{i,t}(k, \theta)$ for the same risk with *noisy* inputs. Let

$$\theta^\circ_{k^\star_{i,t}} \in \arg\min_\theta R^0_{i,t}(k^\star_{i,t}, \theta)$$

and

$$\theta^\eta_{k^\star_{i,t}} \in \arg\min_\theta R^\eta_{i,t}(k^\star_{i,t}, \theta).$$

Define $\mathcal{L}^0_{i,t}(k) := \inf_\theta R^0_{i,t}(k, \theta)$ and $\mathcal{L}^\eta_{i,t}(k) := \inf_\theta R^\eta_{i,t}(k, \theta)$. By A1–A3 there is a unique minimizer at $k^\star_{i,t}$ in the latent case with gap $\Delta_{\min} > 0$.

Under B1 and B4, perturbing the factor regressor of the $k^\star_{i,t}$-model by $e_{k^\star_{i,s}}$ changes the linear predictor by $\gamma_{k^\star_{i,s}} e_{k^\star_{i,s}}$. A second-order expansion of $R^\eta_{i,t}(k^\star_{i,t}, \theta)$ around $\theta^\circ_{k^\star_{i,t}}$ and the curvature bound in B4 yield

$$\mathcal{L}^\eta_{i,t}(k^\star_{i,t}) - \mathcal{L}^0_{i,t}(k^\star_{i,t}) \;\geq\; c_0\, \mathbb{E}\big[\gamma^2_{k^\star_{i,s}} e^2_{k^\star_{i,s}}\big] \;\geq\; c_0\, C^2\, \bar{\sigma}^2_{k^\star_{i,t}}\, (1 + o(1)), \tag{6}$$

uniformly on the window, using the bound $|\gamma_{k^\star_{i,s}}| \geq C > 0$ from Assumption A2 and $\mathbb{E}[e^2_{k^\star_{i,s}}] = \sigma^2_{k^\star_{i,s}}$ from Assumption B1.

For any $k \neq k^\star_{i,t}$, the j-model's factor input does not contain $e_{k^\star_{i,s}}$; under Assumptions B1, B4 and bounded moments, the perturbation of its best attainable risk is at most of order $\bar{\sigma}^2_{k^\star_{i,t}}$ due to cross-effects via $\tilde{Z}_{i,s}$, i.e.,

$$\big|\mathcal{L}^\eta_{i,t}(k) - \mathcal{L}^0_{i,t}(k)\big| \;\leq\; C_2\, \bar{\sigma}^2_{k^\star_{i,t}}\, (1 + o(1)) \tag{7}$$

for some constant $C_2 > 0$ independent of S.

Combining the two steps,

$$\mathcal{L}^{\eta}_{i,t}(k) - \mathcal{L}^{\eta}_{i,t}(k^\star_{i,t}) \;\geq\; \underbrace{\left(\mathcal{L}^{0}_{i,t}(k) - \mathcal{L}^{0}_{i,t}(k^\star_{i,t})\right)}_{\geq\, \Delta_{\min}} - (C_2 + c_0 C^2)\, \bar{\sigma}^2_{k^\star_{i,t}} \;+\; o(\bar{\sigma}^2_{k^\star_{i,t}}), \tag{8}$$

which is (22) with $C_1 := C_2 + c_0 C^2$.

Let $H_{i,t}(S) := \max_{k\in\mathcal{K}} \sup_{\theta\in\Theta_k} \left|L_{i,t}(k;\theta) - R^{\eta}_{i,t}(k,\theta)\right|$. By Assumption A4, $H_{i,t}(S) \xrightarrow{p} 0$ as $S \to \infty$. Therefore, with probability at least $\mathbb{P}\big(H_{i,t}(S) < \frac{1}{2}[\Delta_{\min} - C_1 \bar{\sigma}^2_{k^\star_{i,t}}]\big)$, we have $L_{i,t}(k;\widehat{\theta}_k) > L_{i,t}(k^\star_{i,t};\widehat{\theta}_{k^\star_{i,t}})$ for all $k \neq k^\star_{i,t}$. Finally, B2 and B3 imply $\bar{\sigma}^2_{k^\star_{i,t}}$ is (up to constants) the window-average of $FIQ_{k^\star_{i,s}}$ and can be proxied by the window-average of $FRQ_{k^\star_{i,s}}$, yielding the rest of the statement in Theorem 2. □

B Data

This appendix documents the datasets and summary properties used throughout the paper. We report descriptive statistics for 1-second stock returns and their time-variation and report cross-sectional quantiles of annualized realized volatility for both stocks and factors. We then show how resampling creates aggregation bias for high-minus-low style portfolios and provide an exact aggregation rule. The section concludes with the catalog of the high-frequency factors included in the sample.

B.1 Descriptive statistics

This section provides descriptive statistics for the time series used in the empirical analysis. Figure 7 summarizes cross-sectional features of 1-second stock returns. The distribution of average returns is tightly concentrated near zero, with a slight positive median and mild right skew, indicating negligible drift at the one-second horizon. Standard deviations cluster at a few basis points and display a pronounced right tail, consistent with substantial heterogeneity in intraday volatility across firms. The share of nonzero one-second returns is low for most stocks, reflecting price discreteness and more sparse updates, but exhibits a wide upper tail associated with more actively traded names.

Figure 8 tracks the daily evolution of the cross-section of one-second stock returns using the 5th, 25th, 50th, 75th, and 95th percentiles. The median remains essentially at zero throughout, while dispersion fluctuates over time, widening sharply during market stress. The pattern highlights time-variation in both scale and tail thickness, with negative skew in drawdowns.

Table 5 reports, by calendar year, cross-sectional quantiles of annualized daily realized volatility for the stock universe and for the long–short factor portfolios. By construction, stocks are substantially more volatile than factors: pre-2020 medians lie in the 16–20% range versus 3–5% for factors, and both panels exhibit a pronounced surge in 2020 (medians 29.7% for

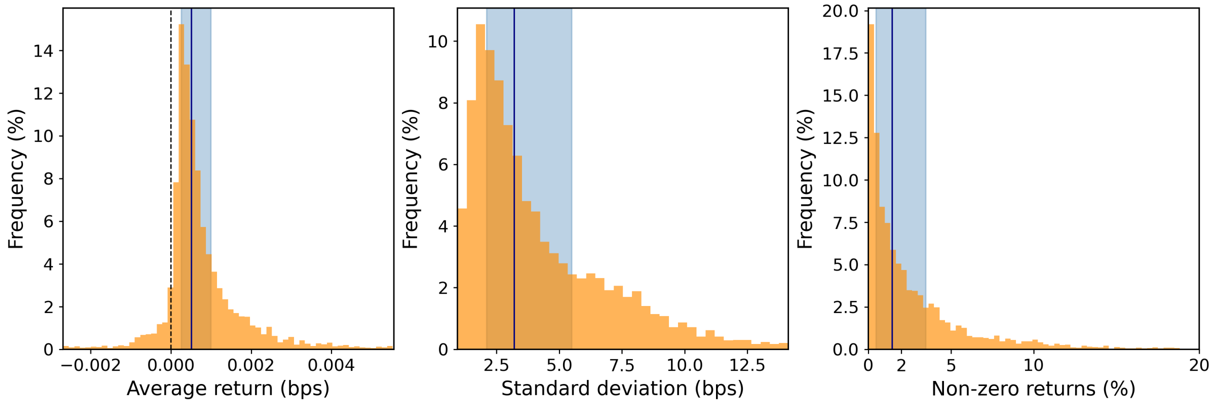

Figure 7: Stock returns at the one-second horizon: cross-sectional histograms of the average returns (in basis points), their standard deviation (bps), and the percentage share of nonzero returns. For readability, the top and bottom 1% of observations are trimmed. The solid vertical line marks the cross-sectional median and the shaded band spans the interquartile range (25th to 75th percentiles).

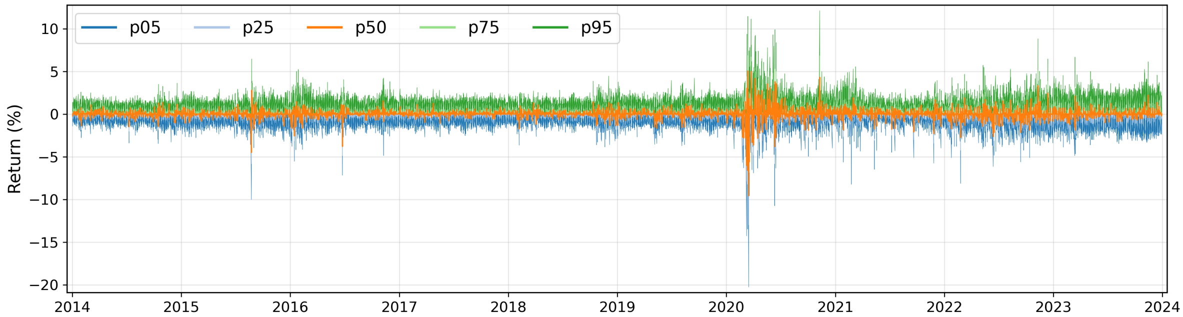

Figure 8: Stock returns at the one-second horizon: time series of cross-sectional percentiles (p05, p25, p50, p75, p95), expressed in percent. Percentiles are computed at 1-second frequency across the stock universe.

Table 5: Cross-sectional distribution of annualized realized volatility across, respectively, sample stocks and factors. Values are reported in percentage by year for median and 10^{th}, 25^{th}, 75^{th} and 90^{th} quantiles.

	Stocks					Factors				
	p_{10}	p_{25}	p_{50}	p_{75}	p_{90}	p_{10}	p_{25}	p_{50}	p_{75}	p_{90}
2014	9.64	12.21	16.26	22.49	31.10	2.44	2.84	3.44	4.30	5.56
2015	10.86	13.58	17.78	24.29	34.11	2.82	3.33	4.08	5.07	6.34
2016	10.69	13.67	18.52	26.22	37.52	2.79	3.37	4.28	5.63	7.37
2017	9.50	11.97	16.03	22.10	30.64	2.53	2.97	3.60	4.40	5.37
2018	11.93	14.99	20.03	27.56	37.25	3.25	3.93	4.97	6.63	8.92
2019	12.05	14.92	19.36	25.76	34.43	3.08	3.64	4.52	5.62	6.87
2020	16.28	21.32	29.69	44.12	67.40	4.66	6.04	8.26	11.58	15.77
2021	14.05	17.72	23.53	32.06	43.03	4.73	5.94	7.70	9.89	12.36
2022	17.05	21.07	27.22	35.91	46.23	5.71	7.07	9.20	12.18	15.60
2023	13.24	16.24	20.91	27.65	36.54	4.37	5.22	6.55	8.36	10.30

stocks and 8.3% for factors). The dispersion widens similarly, with the interquartile range reaching about 23 percentage points for stocks and 5 percentage points for factors during the pandemic, before narrowing in the subsequent years. The parallel movements underscore common volatility conditions across underlying equities and factor portfolios.

B.2 Aggregation bias for HML-style factors

High-frequency factor returns are recorded at one-second resolution, yet many empirical tasks (e.g., computing 5-minute realized variance or evaluating the factor replication over daily and monthly horizons) require resampling to coarser grids. For high-minus-low (and low-minus-high) style portfolios, compounding the factor return itself is mathematically incorrect because cross-product terms appear when simple returns of different portfolios are multiplied.

The problem is generic: it arises whenever a combination of portfolio returns is cumulated from a fine grid to any lower frequency. Figure 9 illustrates the bias with the Fama–French six factors. We obtain the daily and monthly versions of each factor from the Kenneth R. French data library, cumulate the daily series up to monthly frequency, and compare the result with the monthly returns available in the library. We also include the replicated 1-second returns aggregated to the same horizon. The market factor lines up perfectly, whereas high-minus-low style factors diverge.

The algebra is straightforward for a single portfolio. Let $R_1 = r_{0,1}$ and $R_2 = r_{1,2}$ denote two successive simple returns given $t \in \{0, 1, 2\}$; compounding yields

$$(1 + R_1)(1 + R_2) \; = \; 1 + R_1 + R_2 + R_1 R_2 \quad \Longrightarrow \quad R_{12} = R_1 + R_2 + R_1 R_2. \tag{9}$$

Consider a high-minus-low factor formed each sub-period as $R_t^{HML} = R_t^H - R_t^L$. Cumulating the factor returns after differencing at a higher frequency, we obtain

$$\begin{aligned} 1 + R_{12}^{HML} &= (1 + R_1^H - R_1^L)(1 + R_2^H - R_2^L) \\ &= 1 + R_1^{HML} + R_2^{HML} + R_1^{HML} R_2^{HML} - \big(R_1^H R_2^L + R_1^L R_2^H - 2R_1^L R_2^L\big). \end{aligned} \tag{10}$$

The additional terms $-R_1^H R_2^L$, $-R_1^L R_2^H$, and $2R_1^L R_2^L$ have no counterpart in the single-portfolio identity (9). Each is second-order in intraday returns, yet their effect cumulates with both sampling frequency and portfolios number, so aggregating a pre-differenced HML series will generally diverge from the true low-frequency spread. To eliminate this distortion, we compound the high- and low-leg portfolios independently and take their difference only at the target horizon. This rule preserves exact aggregation from one-second data to any lower frequency used in the analysis.

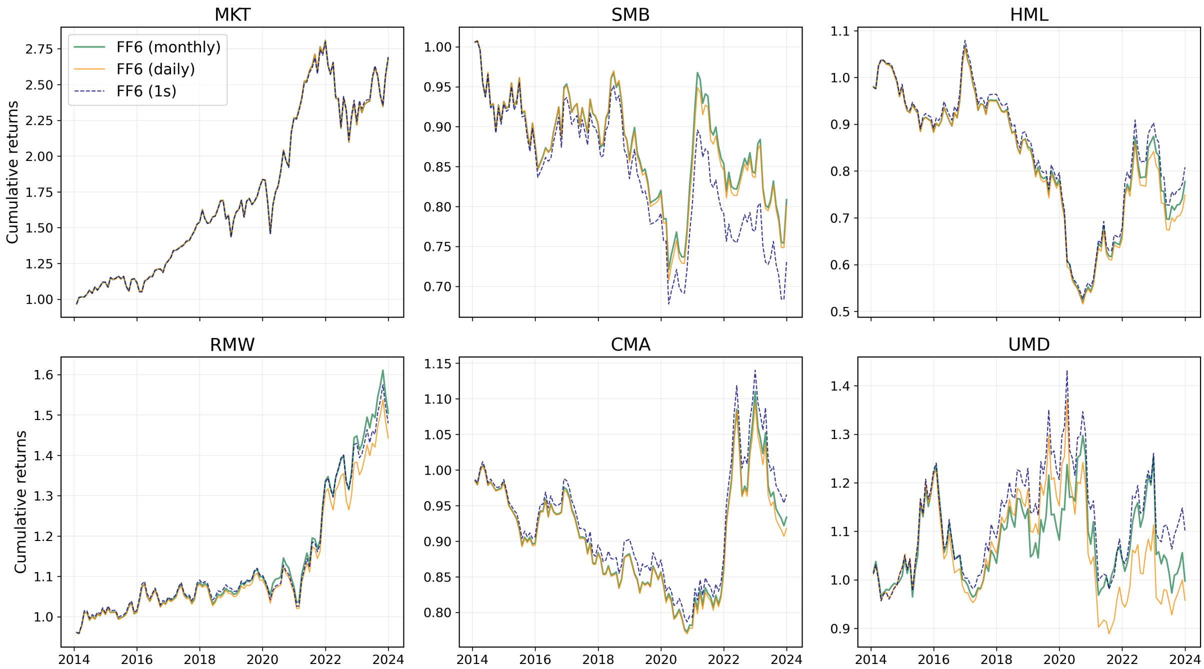

Figure 9: Cumulative gross monthly returns for the Fama–French six factors between 2014 and 2023. Green thick lines are the benchmark series downloaded at monthly frequency from the Kenneth R. French data library, orange solid lines compound the library's daily returns to the monthly horizon and blue dashed lines aggregate the 1-second replication. The market factor (MKT) aggregates exactly, while the five high-minus-low (or low-minus-high) spreads diverge, with the discrepancy widening as the starting sampling interval shortens.

B.3 Sample factors

Table 6: Catalog of the asset-pricing anomalies replicated at high frequency for this study. Columns list the original study and mnemonic, a brief description of the signal, its economic theme, and the data library sourced for replication. The table continues across pages.

Reference	Factor	Description	Cluster	Universe
Abarbanell and Bushee (1998)	dgp_dsale	Change gross margin minus change sales	Quality	JKP
Abarbanell and Bushee (1998)	dsale_dinv	Change sales minus change Inventory	Profit Growth	JKP
Abarbanell and Bushee (1998)	dsale_drec	Change sales minus change receivables	Profit Growth	JKP
Abarbanell and Bushee (1998)	dsale_dsga	Change sales minus change SG&A	Profit Growth	JKP
Abarbanell and Bushee (1998)	sale_emp_gr1	Labor force efficiency	Profit Growth	JKP
Abarbanell and Bushee (1998)	ChInvIA	Change in capital inv (ind adj)	Low Leverage	CZ
Abarbanell and Bushee (1998)	GrSaleToGrInv	Sales growth over inventory growth	Quality	CZ
Abarbanell and Bushee (1998)	GrSaleToGrOverhead	Sales growth over overhead growth	Value	CZ
Ali et al. (2003)	ivol_capm_252d	Idiosyncratic volatility from the CAPM (252 days)	Low Risk	JKP
Alwathainani (2009)	EarningsConsistency	Earnings consistency	Quality	CZ
Amihud (2002)	ami_126d	Amihud Measure	Size	JKP
Amihud and Mendelson (1986)	BidAskSpread	Bid-ask spread	Low Leverage	CZ
Anderson and Garcia-Feijoo (2006)	capx_gr2	CAPEX growth (2 years)	Investment	JKP
Anderson and Garcia-Feijoo (2006)	capx_gr3	CAPEX growth (3 years)	Investment	JKP
Anderson and Garcia-Feijoo (2006)	grcapx	Change in capex (two years)	Investment	CZ
Anderson and Garcia-Feijoo (2006)	grcapx3y	Change in capex (three years)	Investment	CZ
Ang et al. (2006)	ivol_ff3_21d	Idiosyncratic volatility from the Fama-French 3-factor model	Low Risk	JKP
Ang et al. (2006)	rvol_21d	Return volatility	Low Risk	JKP
Ang et al. (2006)	betadown_252d	Downside beta	Low Risk	JKP
Ang et al. (2006)	CoskewACX	Coskewness using daily returns	Profitability	CZ
Asness et al. (2019)	qmj	Quality minus Junk: Composite	Quality	JKP
Asness et al. (2019)	qmj_growth	Quality minus Junk: Growth	Quality	JKP
Asness et al. (2019)	qmj_prof	Quality minus Junk: Profitability	Quality	JKP

Reference	Factor	Description	Cluster	Universe
Asness et al. (2019)	qmj_safety	Quality minus Junk: Safety	Quality	JKP
Asness et al. (2020)	corr_1260d	Market correlation	Seasonality	JKP
Asness et al. (2020)	rmax5_rvol_21d	Highest 5 days of return scaled by volatility	Short-Term Reversal	JKP
Baik and Ahn (2007)	OrderBacklogChg	Change in order backlog	Profitability	CZ
Balakrishnan et al. (2010)	niq_at	Quarterly return on assets	Quality	JKP
Balakrishnan et al. (2010)	roaq	Return on assets (qtrly)	Profitability	CZ
Bali et al. (2017)	rmax5_21d	Highest 5 days of return	Low Risk	JKP
Bali et al. (2011)	rmax1_21d	Maximum daily return	Low Risk	JKP
Bali et al. (2016)	iskew_ff3_21d	Idiosyncratic skewness from the Fama-French 3-factor model	Short-Term Reversal	JKP
Bali et al. (2016)	rskew_21d	Total skewness	Short-Term Reversal	JKP
Bali et al. (2016)	ReturnSkew3F	Idiosyncratic skewness (3F model)	Short-Term Reversal	CZ
Ball et al. (2016)	cop_atl1	Cash-based operating profits-to-lagged book assets	Quality	JKP
Ball et al. (2016)	op_atl1	Operating profits-to-lagged book assets	Quality	JKP
Ball et al. (2016)	CBOperProf	Cash-based operating profitability	Profitability	CZ
Ball et al. (2016)	OperProfRD	Operating profitability R&D adjusted	Profitability	CZ
Banz (1981)	market_equity	Market Equity	Size	JKP
Barbee Jr et al. (1996)	sale_me	Sales-to-market	Value	JKP
Barry and Brown (1984)	FirmAge	Firm age based on CRSP	Low Risk	CZ
Barth et al. (1999)	ni_inc8q	Number of consecutive quarters with earnings increases	Quality	JKP
Basu (1983)	ni_me	Earnings-to-price	Value	JKP
Basu (1977)	EP	Earnings-to-Price Ratio	Value	CZ
Belo and Lin (2012)	inv_gr1	Inventory growth	Investment	JKP
Belo and Lin (2012)	InvGrowth	Inventory Growth	Investment	CZ
Belo et al. (2014)	emp_gr1	Hiring rate	Investment	JKP
Belo et al. (2014)	BrandInvest	Brand capital investment	Profitability	CZ
Bhandari (1988)	debt_me	Debt-to-market	Value	JKP
Blitz et al. (2011)	resff3_12_1	Residual momentum t-12 to t-1	Momentum	JKP

Reference	Factor	Description	Cluster	Universe
Blitz et al. (2011)	resff3_6_1	Residual momentum t-6 to t-1	Momentum	JKP
Blitz et al. (2011)	ResidualMomentum	Momentum based on FF3 residuals	Momentum	CZ
Blume and Husic (1973)	Price	Price	Size	CZ
Bouchaud et al. (2019)	ocf_at	Operating cash flow to assets	Profitability	JKP
Bouchaud et al. (2019)	ocf_at_chg1	Change in operating cash flow to assets	Profit Growth	JKP
Boudoukh et al. (2007)	eqnpo_me	Net payout yield	Value	JKP
Boudoukh et al. (2007)	eqpo_me	Payout yield	Value	JKP
Boudoukh et al. (2007)	NetPayoutYield	Net Payout Yield	Value	CZ
Boudoukh et al. (2007)	PayoutYield	Payout Yield	Value	CZ
Bradshaw et al. (2006)	dbnetis_at	Net debt issuance	Seasonality	JKP
Bradshaw et al. (2006)	eqnetis_at	Net equity issuance	Value	JKP
Bradshaw et al. (2006)	netis_at	Net total issuance	Value	JKP
Bradshaw et al. (2006)	NetDebtFinance	Net debt financing	Debt Issuance	CZ
Bradshaw et al. (2006)	NetEquityFinance	Net equity financing	Low Risk	CZ
Bradshaw et al. (2006)	XFIN	Net external financing	Low Risk	CZ
Brennan et al. (1998)	dolvol_126d	Dollar trading volume	Size	JKP
Ang et al. (2006)	FEPS	Analyst earnings per share	Profitability	CZ
Chan et al. (1996)	AnnouncementReturn	Earnings announcement return	Momentum	CZ
Chan et al. (1996)	REV6	Earnings forecast revisions	Momentum	CZ
Chan et al. (2001)	rd_me	R&D-to-market	Size	JKP
Chan et al. (2001)	rd_sale	R&D-to-sales	Low Leverage	JKP
Chan et al. (2001)	AdExp	Advertising Expense	Value	CZ
Chan et al. (2001)	RD	R&D over market cap	Low Leverage	CZ
Chandrashekar et al. (2009)	CashProd	Cash Productivity	Quality	CZ
Chen et al. (2002)	DelBreadth	Breadth of ownership	Momentum	CZ
Chordia et al. (2001)	VolSD	Volume Variance	Size	CZ
Chordia et al. (2001)	dolvol_var_126d	Coefficient of variation for dollar trading volume	Profitability	JKP

Reference	Factor	Description	Cluster	Universe
Chordia et al. (2001)	turnover_var_126d	Coefficient of variation for share turnover	Profitability	JKP
Cohen and Lou (2012)	retConglomerate	Conglomerate return	Investment	CZ
Cohen et al. (2013)	RDAbility	R&D ability	Value	CZ
Cooper et al. (2008)	at_gr1	Asset Growth	Investment	JKP
Cooper et al. (2008)	AssetGrowth	Asset growth	Investment	CZ
Corwin and Schultz (2012)	bidaskhl_21d	The high-low bid-ask spread	Low Leverage	JKP
Da and Warachka (2011)	EarningsForecastDisparity	Long-vs-short EPS forecasts	Quality	CZ
Daniel and Titman (2006)	eqnpo_12m	Equity net payout	Value	JKP
Daniel and Titman (2006)	CompEquIss	Composite equity issuance	Value	CZ
Daniel and Titman (2006)	IntanBM	Intangible return using BM	Quality	CZ
Daniel and Titman (2006)	IntanCFP	Intangible return using CFtoP	Quality	CZ
Daniel and Titman (2006)	IntanEP	Intangible return using EP	Quality	CZ
Daniel and Titman (2006)	IntanSP	Intangible return using Sale2P	Quality	CZ
Daniel and Titman (2006)	ShareIss5Y	Share issuance (5 year)	Investment	CZ
Datar et al. (1998)	turnover_126d	Share turnover	Low Risk	JKP
De Bondt and Thaler (1985)	ret_60_12	Long-term reversal	Investment	JKP
De Bondt and Thaler (1985)	LRreversal	Long-run reversal	Investment	CZ
De Bondt and Thaler (1985)	MRreversal	Medium-run reversal	Quality	CZ
Dechow et al. (2001)	ShortInterest	Short Interest	Size	CZ
Dechow et al. (2004)	eq_dur	Equity duration	Value	JKP
Dechow et al. (2004)	EquityDuration	Equity Duration	Value	CZ
Desai et al. (2004)	ocf_me	Operating cash flow-to-market	Value	JKP
Dichev (1998)	o_score	Ohlson O-score	Profitability	JKP
Dichev (1998)	z_score	Altman Z-score	Low Leverage	JKP
Diether et al. (2002)	ForecastDispersion	EPS Forecast Dispersion	Low Risk	CZ
Dimson (1979)	beta_dimson_21d	Dimson beta	Low Risk	JKP
Doyle et al. (2003)	ExclExp	Excluded Expenses	Profitability	CZ

Reference	Factor	Description	Cluster	Universe
Eisfeldt and Papanikolaou (2013)	OrgCap	Organizational capital	Low Risk	CZ
Elgers et al. (2001)	sfe	Earnings Forecast to price	Value	CZ
Fairfield et al. (2003)	lnoa_gr1a	Change in long-term net operating assets	Investment	JKP
Fairfield et al. (2003)	GrLTNOA	Growth in long term operating assets	Quality	CZ
Fama and French (2018)	cma	Asset growth	Investment	FF6
Fama and French (2018)	hml	Book assets-to-market value	Value	FF6
Fama and French (2018)	mkt	Value-weighted excess market return	Market	FF6
Fama and French (2018)	rmw	Operating profits-to-book equity	Profitability	FF6
Fama and French (2018)	smb	Market capitalization	Size	FF6
Fama and French (2018)	umd	Price momentum t-12 to t-1	Momentum	FF6
Eugene and French (1992)	at_be	Book leverage	Low Leverage	JKP
Eugene and French (1992)	at_me	Assets-to-market	Value	JKP
Fama and French (2015)	ope_be	Operating profits-to-book equity	Profitability	JKP
Eugene and French (1992)	BMdec	Book to market using December ME	Value	CZ
Eugene and French (1992)	BookLeverage	Book leverage (annual)	Low Leverage	CZ
Fama and French (2006)	OperProf	operating profits / book equity	Profitability	CZ
Fama and MacBeth (1973)	beta_60m	Market Beta	Low Risk	JKP
Fama and MacBeth (1973)	Beta	CAPM beta	Size	CZ
Foster et al. (1984)	niq_su	Standardized earnings surprise	Profit Growth	JKP
Foster et al. (1984)	EarningsSurprise	Earnings Surprise	Momentum	CZ
Francis et al. (2004)	earnings_variability	Earnings variability	Low Risk	JKP
Francis et al. (2004)	ni_ar1	Earnings persistence	Debt Issuance	JKP
Francis et al. (2004)	ni_ivol	Earnings volatility	Low Leverage	JKP
Frankel and Lee (1998)	ival_me	Intrinsic value-to-market	Value	JKP
Frankel and Lee (1998)	AOP	Analyst Optimism	Value	CZ
Frankel and Lee (1998)	AnalystValue	Analyst Value	Value	CZ
Frankel and Lee (1998)	PredictedFE	Predicted Analyst forecast error	Value	CZ

Reference	Factor	Description	Cluster	Universe
Franzoni and Marin (2006)	FR	Pension Funding Status	Profitability	CZ
Frazzini and Pedersen (2014)	betabab_1260d	Frazzini-Pedersen market beta	Low Risk	JKP
George and Hwang (2004)	prc_highprc_252d	Current price to high price over last year	Momentum	JKP
Moskowitz and Grinblatt (1999)	IndMom	Industry Momentum	Momentum	CZ
Hafzalla et al. (2011)	oaccruals_ni	Percent operating accruals	Accruals	JKP
Hafzalla et al. (2011)	taccruals_ni	Percent total accruals	Accruals	JKP
Hafzalla et al. (2011)	PctAcc	Percent Operating Accruals	Accruals	CZ
Hafzalla et al. (2011)	PctTotAcc	Percent Total Accruals	Investment	CZ
Hahn and Lee (2009)	tangibility	Asset tangibility	Low Leverage	JKP
Hahn and Lee (2009)	tang	Tangibility	Size	CZ
Harvey and Siddique (2000)	coskew_21d	Coskewness	Seasonality	JKP
Harvey and Siddique (2000)	Coskewness	Coskewness	Value	CZ
Haugen and Baker (1996)	at_turnover	Capital turnover	Quality	JKP
Haugen and Baker (1996)	ni_be	Return on equity	Profitability	JKP
Haugen and Baker (1996)	RoE	net income / book equity	Profitability	CZ
Haugen and Baker (1996)	VarCF	Cash-flow to price variance	Size	CZ
Haugen and Baker (1996)	VolMkt	Volume to market equity	Low Risk	CZ
Haugen and Baker (1996)	VolumeTrend	Volume Trend	Low Risk	CZ
Hawkins et al. (1984)	AnalystRevision	EPS forecast revision	Momentum	CZ
Heston and Sadka (2008)	seas_11_15an	Years 11-15 lagged returns, annual	Seasonality	JKP
Heston and Sadka (2008)	seas_11_15na	Years 11-15 lagged returns, nonannual	Seasonality	JKP
Heston and Sadka (2008)	seas_16_20an	Years 16-20 lagged returns, annual	Seasonality	JKP
Heston and Sadka (2008)	seas_16_20na	Years 16-20 lagged returns, nonannual	Accruals	JKP
Heston and Sadka (2008)	seas_1_1an	Year 1-lagged return, annual	Profit Growth	JKP
Heston and Sadka (2008)	seas_1_1na	Year 1-lagged return, nonannual	Momentum	JKP
Heston and Sadka (2008)	seas_2_5an	Years 2-5 lagged returns, annual	Seasonality	JKP
Heston and Sadka (2008)	seas_2_5na	Years 2-5 lagged returns, nonannual	Investment	JKP

Reference	Factor	Description	Cluster	Universe
Heston and Sadka (2008)	seas_6_10an	Years 6-10 lagged returns, annual	Seasonality	JKP
Heston and Sadka (2008)	seas_6_10na	Years 6-10 lagged returns, nonannual	Low Risk	JKP
Heston and Sadka (2008)	Mom12mOffSeason	Momentum without the seasonal part	Momentum	CZ
Heston and Sadka (2008)	MomSeason11YrPlus	Return seasonality years 11 to 15	Quality	CZ
Hirshleifer et al. (2004)	noa_at	Net operating assets	Debt Issuance	JKP
Hirshleifer et al. (2004)	noa_gr1a	Change in net operating assets	Investment	JKP
Hirshleifer et al. (2004)	NOA	Net Operating Assets	Profitability	CZ
Hirshleifer et al. (2004)	dNoa	change in net operating assets	Investment	CZ
Hou (2007)	EarnSupBig	Earnings surprise of big firms	Profitability	CZ
Hou (2007)	IndRetBig	Industry return of big firms	Momentum	CZ
Hou and Robinson (2006)	Herf	Industry concentration (sales)	Low Risk	CZ
Hou and Robinson (2006)	HerfAsset	Industry concentration (assets)	Low Risk	CZ
Hou and Robinson (2006)	HerfBE	Industry concentration (equity)	Low Risk	CZ
Hou et al. (2015)	niq_be	Quarterly return on equity	Profitability	JKP
Huang (2009)	ocfq_saleq_std	Cash flow volatility	Low Risk	JKP
Jegadeesh (1990)	ret_1_0	Short-term reversal	Short-Term Reversal	JKP
Jegadeesh and Livnat (2006)	saleq_su	Standardized Revenue surprise	Profit Growth	JKP
Jegadeesh and Livnat (2006)	RevenueSurprise	Revenue Surprise	Momentum	CZ
Jegadeesh and Titman (1993)	ret_12_1	Price momentum t-12 to t-1	Momentum	JKP
Jegadeesh and Titman (1993)	ret_6_1	Price momentum t-6 to t-1	Momentum	JKP
Jegadeesh and Titman (1993)	Mom12m	Momentum (12 month)	Momentum	CZ
Jegadeesh et al. (2004)	ChangeInRecommendation	Change in recommendation	Size	CZ
Jegadeesh and Titman (1993)	ret_3_1	Price momentum t-3 to t-1	Momentum	JKP
Jegadeesh and Titman (1993)	ret_9_1	Price momentum t-9 to t-1	Momentum	JKP
Jensen et al. (2023)	cop_at	Cash-based operating profits-to-book assets	Quality	JKP
Jensen et al. (2023)	gp_atl1	Gross profits-to-lagged assets	Quality	JKP
Jensen et al. (2023)	iskew_capm_21d	Idiosyncratic skewness from the CAPM	Short-Term Reversal	JKP

Reference	Factor	Description	Cluster	Universe
Jensen et al. (2023)	iskew_hxz4_21d	Idiosyncratic skewness from the q-factor model	Short-Term Reversal	JKP
Jensen et al. (2023)	ivol_capm_21d	Idiosyncratic volatility from the CAPM (21 days)	Low Risk	JKP
Jensen et al. (2023)	ivol_hxz4_21d	Idiosyncratic volatility from the q-factor model	Low Risk	JKP
Jensen et al. (2023)	niq_at_chg1	Change in quarterly return on assets	Profit Growth	JKP
Jensen et al. (2023)	niq_be_chg1	Change in quarterly return on equity	Profit Growth	JKP
Jensen et al. (2023)	op_at	Operating profits-to-book assets	Quality	JKP
Jensen et al. (2023)	ope_bel1	Operating profits-to-lagged book equity	Profitability	JKP
Jensen et al. (2023)	saleq_gr1	Sales growth (1 quarter)	Investment	JKP
Jiang et al. (2005)	age	Firm age	Low Leverage	JKP
Kelly and Jiang (2014)	BetaTailRisk	Tail risk beta	Size	CZ
La Porta (1996)	fgr5yrLag	Long-term EPS forecast	Investment	CZ
Lakonishok et al. (1994)	fcf_me	Free cash flow-to-price	Value	JKP
Lakonishok et al. (1994)	sale_gr1	Sales Growth (1 year)	Investment	JKP
Lakonishok et al. (1994)	sale_gr3	Sales Growth (3 years)	Investment	JKP
Lakonishok et al. (1994)	CF	Cash flow to market	Value	CZ
Lakonishok et al. (1994)	MeanRankRevGrowth	Revenue Growth Rank	Value	CZ
Lamont et al. (2001)	kz_index	Kaplan-Zingales index	Seasonality	JKP
Landsman et al. (2011)	RDS	Real dirty surplus	Investment	CZ
Lev and Nissim (2004)	pi_nix	Taxable income-to-book income	Seasonality	JKP
Lev and Nissim (2004)	Tax	Taxable income to income	Investment	CZ
Li (2011)	rd5_at	R&D capital-to-book assets	Low Leverage	JKP
Litzenberger and Ramaswamy (1979)	div12m_me	Dividend yield	Value	JKP
Liu (2006)	zero_trades_126d	Number of zero trades with turnover as tiebreaker (6 months)	Low Risk	JKP
Liu (2006)	zero_trades_21d	Number of zero trades with turnover as tiebreaker (1 month)	Low Risk	JKP
Liu (2006)	zero_trades_252d	Number of zero trades with turnover as tiebreaker (12 months)	Low Risk	JKP
Liu (2006)	zerotradeAlt1	Days with zero trades	Low Risk	CZ
Liu (2006)	zerotradeAlt12	Days with zero trades	Low Risk	CZ

Reference	Factor	Description	Cluster	Universe
Lockwood and Prombutr (2010)	ChEQ	Growth in book equity	Investment	CZ
Loh and Warachka (2012)	EarningsStreak	Earnings surprise streak	Quality	CZ
Loh and Warachka (2012)	NumEarnIncrease	Earnings streak length	Profitability	CZ
Lou (2014)	GrAdExp	Growth in advertising expenses	Investment	CZ
Loughran and Wellman (2011)	ebitda_mev	Ebitda-to-market enterprise value	Value	JKP
Loughran and Wellman (2011)	EntMult	Enterprise Multiple	Value	CZ
Lyandres et al. (2008)	debt_gr3	Growth in book debt (3 years)	Debt Issuance	JKP
Lyandres et al. (2008)	ppeinv_gr1a	Change PPE and Inventory	Investment	JKP
Lyandres et al. (2008)	CompositeDebtIssuance	Composite debt issuance	Value	CZ
Lyandres et al. (2008)	InvestPPEInv	change in ppe and inv/assets	Investment	CZ
Menzly and Ozbas (2010)	iomom_cust	Customers momentum	Investment	CZ
Menzly and Ozbas (2010)	iomom_supp	Suppliers momentum	Quality	CZ
Miller and Scholes (1982)	prc	Price per share	Size	JKP
Nguyen and Swanson (2009)	Frontier	Efficient frontier index	Value	CZ
Novy-Marx (2013)	gp_at	Gross profits-to-assets	Quality	JKP
Novy-Marx (2011)	opex_at	Operating leverage	Quality	JKP
Novy-Marx (2012)	ret_12_7	Price momentum t-12 to t-7	Profit Growth	JKP
Novy-Marx (2013)	GP	gross profits / total assets	Quality	CZ
Novy-Marx (2012)	IntMom	Intermediate Momentum	Profit Growth	CZ
Ortiz-Molina and Phillips (2014)	aliq_at	Liquidity of book assets	Investment	JKP
Ortiz-Molina and Phillips (2014)	aliq_mat	Liquidity of market assets	Low Leverage	JKP
Palazzo (2012)	cash_at	Cash-to-assets	Low Leverage	JKP
Pástor and Stambaugh (2003)	BetaLiquidityPS	Pastor-Stambaugh liquidity beta	Accruals	CZ
Penman et al. (2007)	bev_mev	Book-to-market enterprise value	Value	JKP
Penman et al. (2007)	netdebt_me	Net debt-to-price	Low Leverage	JKP
Penman et al. (2007)	BPEBM	Leverage component of BM	Low Risk	CZ
Penman et al. (2007)	EBM	Enterprise component of BM	Value	CZ

Reference	Factor	Description	Cluster	Universe
Penman et al. (2007)	NetDebtPrice	Net debt to price	Low Leverage	CZ
Piotroski (2000)	f_score	Pitroski F-score	Profitability	JKP
Piotroski (2000)	PS	Piotroski F-score	Profitability	CZ
Pontiff and Woodgate (2008)	chcsho_12m	Net stock issues	Value	JKP
Pontiff and Woodgate (2008)	ShareIss1Y	Share issuance (1 year)	Value	CZ
Prakash and Sinha (2013)	DelDRC	Deferred Revenue	Low Leverage	CZ
Rajgopal et al. (2003)	OrderBacklog	Order backlog	Value	CZ
Richardson et al. (2005)	be_gr1a	Change in common equity	Investment	JKP
Richardson et al. (2005)	coa_gr1a	Change in current operating assets	Investment	JKP
Richardson et al. (2005)	col_gr1a	Change in current operating liabilities	Investment	JKP
Richardson et al. (2005)	cowc_gr1a	Change in current operating working capital	Accruals	JKP
Richardson et al. (2005)	fnl_gr1a	Change in financial liabilities	Debt Issuance	JKP
Richardson et al. (2005)	lti_gr1a	Change in long-term investments	Seasonality	JKP
Richardson et al. (2005)	ncoa_gr1a	Change in noncurrent operating assets	Investment	JKP
Richardson et al. (2005)	ncol_gr1a	Change in noncurrent operating liabilities	Debt Issuance	JKP
Richardson et al. (2005)	nfna_gr1a	Change in net financial assets	Debt Issuance	JKP
Richardson et al. (2005)	nncoa_gr1a	Change in net noncurrent operating assets	Investment	JKP
Richardson et al. (2005)	sti_gr1a	Change in short-term investments	Seasonality	JKP
Richardson et al. (2005)	taccruals_at	Total accruals	Accruals	JKP
Richardson et al. (2005)	DelCOA	Change in current operating assets	Investment	CZ
Richardson et al. (2005)	DelCOL	Change in current operating liabilities	Investment	CZ
Richardson et al. (2005)	DelEqu	Change in equity to assets	Investment	CZ
Richardson et al. (2005)	DelFINL	Change in financial liabilities	Investment	CZ
Richardson et al. (2005)	DelLTI	Change in long-term investment	Investment	CZ
Richardson et al. (2005)	DelNetFin	Change in net financial assets	Value	CZ
Richardson et al. (2005)	TotalAccruals	Total accruals	Investment	CZ
Ritter (1991)	AgeIPO	IPO and age	Value	CZ

Reference	Factor	Description	Cluster	Universe
Rosenberg et al. (1985)	be_me	Book-to-market equity	Value	JKP
Sloan (1996)	oaccruals_at	Operating accruals	Accruals	JKP
Sloan (1996)	Accruals	Accruals	Quality	CZ
Soliman (2008)	ebit_bev	Return on net operating assets	Profitability	JKP
Soliman (2008)	ebit_sale	Profit margin	Profitability	JKP
Soliman (2008)	sale_bev	Assets turnover	Quality	JKP
Soliman (2008)	ChAssetTurnover	Change in Asset Turnover	Value	CZ
Soliman (2008)	ChNNCOA	Change in Net Noncurrent Op Assets	Value	CZ
Soliman (2008)	ChNWC	Change in Net Working Capital	Low Risk	CZ
Stambaugh and Yuan (2017)	mispricing_mgmt	Mispricing factor: Management	Investment	JKP
Stambaugh and Yuan (2017)	mispricing_perf	Mispricing factor: Performance	Quality	JKP
Stattman (1980)	BM	Book to market (original definition)	Value	CZ
Thomas and Zhang (2002)	inv_gr1a	Inventory change	Investment	JKP
Thomas and Zhang (2011)	tax_gr1a	Tax expense surprise	Profit Growth	JKP
Thomas and Zhang (2002)	ChInv	Inventory Growth	Investment	CZ
Thomas and Zhang (2011)	ChTax	Change in Taxes	Quality	CZ
Titman et al. (2004)	capex_abn	Abnormal corporate investment	Debt Issuance	JKP
Titman et al. (2004)	Investment	Investment to revenue	Investment	CZ
Tuzel (2010)	realestate	Real estate holdings	Low Leverage	CZ
Xie (2001)	capx_gr1	CAPEX growth (1 year)	Investment	JKP
Xie (2001)	AbnormalAccruals	Abnormal Accruals	Value	CZ

C Empirical results

This appendix evaluates the robustness of the main empirical findings along four complementary dimensions: (i) the choice of the realized volatility estimator, (ii) the length of the estimation window $L \in \{756, 1008, 1260\}$, corresponding to 3, 4 and 5 trading years, (iii) the forecasting horizon $h \in \{1, 5, 22\}$ and (iv) the scoring rule (QLIKE and RMSE). Across all configurations, the model rankings established in the main text persist. Changes in the volatility estimator mainly shift the level of the loss but not the ordering, while varying L has a negligible effect and conclusions are consistent under both loss functions. Overall, the robustness checks indicate that adaptive multi-factor modeling delivers stable gains in statistical accuracy and economic value across horizons and measurement choices.

Across all panels in Figure 10, the ordering of models is stable. The class of proposed models dominates, with the three-factor specification closest to the origin (lowest QLIKE). Traditional benchmarks (HAR, HARQ, SHAR) and simple factor models (HAR-MKT, MFV-CRV, MFV-PC1) lie farther from the center. The relative distances between models are large compared with the differences induced by the choice of L, indicating that the estimation window length plays a second-order role for ranking. Absolute QLIKE levels vary across estimators, as expected, but the model ordering is invariant.

The formal comparisons in Table 7 corroborate these graphical patterns. Measured by annualized utility, our model specification outperforms the benchmarks for the vast majority of stocks across estimators (Panel A), and these gains are frequently statistically significant at the 5% level (Panel B). The outperformance rates typically exceed 90%. Panel C shows that improvements are economically sizable: in almost every combination of estimator and benchmark, the utility difference exceeds 1 bp per year for more than 99% of stocks.

Table 8 shows that these conclusions are not specific to the QLIKE loss metric. Under RMSE, the multi-factor models continue to deliver the lowest errors across quartiles and horizons. Relative to the baseline HAR, mean RMSE falls by about 8.6% for $h = 1$ (0.5654 vs. 0.5166), 9.8% for $h = 5$ (0.6827 vs. 0.6159), and 7.7% for $h = 22$ (0.7712 vs. 0.7116). The reductions are mirrored at the median and the quartiles. Performance improves monotonically with the number of selected factors, and our specifications remain uniformly ahead of the alternatives. Overall, the evidence indicates that the ranking of models is robust to the volatility estimator, the estimation window, and the scoring rule.

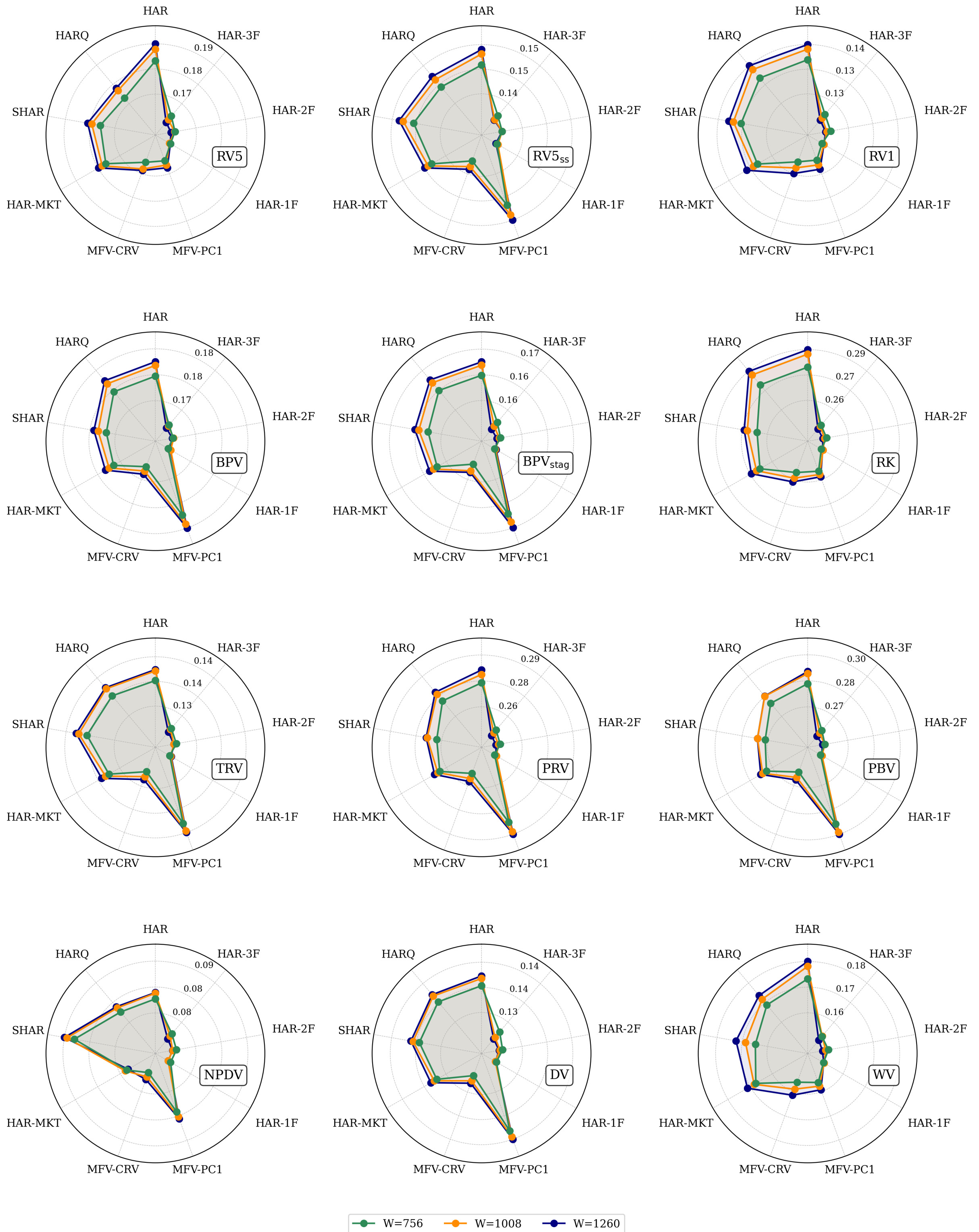

Figure 10: Radar plots of cross-sectional average QLIKE losses by volatility estimator (panels) and forecasting model (vertices). Lower values are closer to the origin. Different colors correspond to estimation window lengths $L \in 756, 1008, 1260$, as indicated in the legend. Within each panel the radial axis is scaled to that estimator. The multi-factor models dominate across all realized specification and estimation windows. Differences across L are small relative to model gaps, indicating that the ranking is less sensitive to the estimation window length.

Table 7: Utility-based outperformance of the HAR-1F model against benchmark models, by realized volatility estimator. Panel A reports Out.perf.,(% of stocks for which the HAR-1F model attains higher annualized utility); Panel B reports Sig.Out.perf.,(% with the improvement significant at the 5% level via a Diebold-Mariano test); Panel C reports Diff $\geq$ 1 bp (% with a utility gain of at least 1 basis point per year). Each entry shows the baseline percentage, with the updated specification in parentheses.

	Benchmarks					
Volatility estimators	HAR	HARQ	SHAR	HAR-MKT	MFV-CRV	MFV-PC1
Panel A: Out.perf. (%)						
RV5	98.9 (1.1)	92.3 (7.7)	95.8 (4.2)	98.2 (1.8)	90.4 (9.6)	82.4 (17.6)
$RV5_{ss}$	98.6 (1.4)	98.1 (1.9)	98.4 (1.6)	98.7 (1.3)	88.8 (11.2)	98.2 (1.8)
RV1	98.8 (1.2)	98.6 (1.4)	96.3 (3.7)	97.7 (2.3)	90.9 (9.1)	81.6 (18.4)
BPV	98.9 (1.1)	99.1 (0.9)	96.4 (3.6)	98.8 (1.2)	91.2 (8.8)	98.7 (1.3)
BPV_{stag}	99.0 (1.0)	99.0 (1.0)	96.3 (3.7)	98.9 (1.1)	89.0 (11.0)	98.8 (1.2)
RK	99.2 (0.8)	99.3 (0.7)	96.4 (3.6)	99.0 (1.0)	94.9 (5.1)	86.2 (13.8)
TRV	99.0 (1.0)	99.0 (1.0)	98.6 (1.4)	99.4 (0.6)	88.7 (11.3)	98.8 (1.2)
PRV	99.2 (0.8)	98.8 (1.2)	94.9 (5.1)	99.2 (0.8)	94.3 (5.7)	99.0 (1.0)
PBV	99.2 (0.8)	98.6 (1.4)	94.7 (5.3)	99.3 (0.7)	93.9 (6.1)	98.9 (1.1)
NPDV	99.3 (0.7)	99.2 (0.8)	99.2 (0.8)	95.6 (4.4)	89.4 (10.6)	99.2 (0.8)
DV	96.6 (0.5)	95.5 (0.6)	84.6 (1.5)	89.5 (0.5)	56.4 (0.9)	96.0 (1.0)
WV	96.8 (0.9)	88.8 (1.2)	75.2 (3.1)	90.9 (0.9)	78.9 (1.1)	53.1 (2.3)
Panel B: Sig.Out.perf. (%)						
RV5	98.2 (0.0)	97.0 (0.1)	97.9 (0.3)	95.1 (0.0)	84.9 (0.9)	77.0 (0.9)
$RV5_{ss}$	97.8 (0.0)	97.8 (0.1)	97.7 (0.3)	95.7 (0.0)	83.0 (0.9)	92.6 (0.9)
RV1	97.7 (0.0)	97.6 (0.1)	97.0 (0.3)	96.7 (0.0)	87.4 (1.0)	79.0 (0.9)
BPV	98.1 (0.0)	98.2 (0.1)	97.3 (0.3)	95.3 (0.0)	85.8 (0.9)	93.3 (0.9)
BPV_{stag}	98.0 (0.0)	98.2 (0.1)	97.1 (0.3)	95.5 (0.0)	84.8 (0.9)	93.0 (0.9)
RK	97.9 (0.0)	98.0 (0.1)	97.3 (0.3)	95.1 (0.0)	87.1 (0.9)	78.2 (0.9)
TRV	97.6 (0.0)	97.6 (0.1)	96.9 (0.3)	96.2 (0.0)	86.7 (0.9)	92.0 (0.9)
PRV	97.9 (0.0)	97.4 (0.1)	96.0 (0.3)	93.7 (0.0)	85.3 (0.9)	94.0 (0.9)
PBV	97.6 (0.0)	97.0 (0.1)	95.1 (0.3)	93.3 (0.0)	85.1 (0.9)	93.7 (0.9)
NPDV	99.1 (0.0)	99.0 (0.1)	98.4 (0.3)	96.6 (0.0)	93.3 (0.9)	96.8 (0.9)
DV	98.5 (0.0)	98.0 (0.0)	97.0 (0.3)	96.9 (0.0)	87.9 (0.9)	93.2 (0.9)
WV	99.2 (0.1)	98.3 (0.1)	96.7 (0.3)	97.7 (0.0)	93.1 (0.9)	85.1 (0.9)
Panel C: Diff $\geq$ 1 bp (%)						
RV5	99.3 (0.6)	99.9 (1.0)	99.8 (2.0)	99.8 (0.5)	99.9 (0.8)	99.9 (1.4)
$RV5_{ss}$	99.2 (0.6)	99.9 (1.0)	99.8 (2.0)	99.6 (0.5)	99.1 (0.8)	99.5 (1.4)
RV1	99.2 (0.6)	99.8 (1.0)	99.8 (2.0)	99.8 (0.5)	99.4 (0.8)	99.7 (1.4)
BPV	99.5 (0.6)	99.8 (1.0)	99.4 (2.0)	99.5 (0.5)	99.4 (0.8)	99.2 (1.4)
BPV_{stag}	99.4 (0.6)	99.8 (1.0)	99.5 (2.0)	99.7 (0.5)	99.6 (0.8)	99.4 (1.4)
RK	99.2 (0.6)	99.2 (1.0)	99.2 (2.0)	99.3 (0.5)	99.3 (0.8)	99.4 (1.4)
TRV	99.8 (0.6)	99.8 (1.0)	99.8 (2.0)	99.8 (0.5)	99.8 (0.8)	99.7 (1.4)
PRV	99.2 (0.6)	99.2 (1.0)	99.3 (2.0)	99.3 (0.5)	99.4 (0.8)	98.9 (1.4)
PBV	99.3 (0.6)	99.0 (1.0)	99.3 (2.0)	99.4 (0.5)	99.4 (0.8)	99.1 (1.4)
NPDV	99.7 (0.6)	99.6 (1.0)	99.5 (2.0)	99.6 (0.5)	99.8 (0.8)	99.5 (1.4)
DV	99.8 (0.6)	99.8 (1.0)	99.8 (2.0)	99.8 (0.5)	99.6 (0.8)	99.5 (1.4)
WV	99.3 (0.6)	98.3 (1.0)	99.3 (2.0)	99.6 (0.5)	99.6 (0.8)	99.6 (1.4)

Table 8: Cross-sectional RMSE of volatility forecasts by model and horizon. Entries are the 25th percentile (Q1), mean, median, and 75th percentile (Q3) across 1041 stocks. Within each row, the lowest value is in bold.

	Forecasting models								
	HAR	SHAR	HARQ	HAR-MKT	MFV-CRV	MFV-PC1	HAR-1F	HAR-2F	HAR-3F
Panel A: $h = 1$									
Q1	0.2868	0.2945	0.2825	0.3024	0.2709	0.2629	0.2589	0.2538	**0.2521**
Median	0.4210	0.4158	0.4107	0.4318	0.3898	0.3811	0.3769	0.3722	**0.3705**
Mean	0.5654	0.5624	0.5564	0.5795	0.5396	0.5311	0.5243	0.5188	**0.5166**
Q3	0.6218	0.6172	0.6143	0.6407	0.5793	0.5719	0.5709	0.5642	**0.5659**
Panel B: $h = 5$									
Q1	0.3736	0.3759	0.3723	0.3758	0.3577	0.3556	0.3414	0.3313	**0.3224**
Median	0.5111	0.5104	0.5057	0.5153	0.4854	0.4793	0.4717	0.4578	**0.4472**
Mean	0.6827	0.6801	0.6777	0.6821	0.6615	0.6578	0.6394	0.6262	**0.6159**
Q3	0.7546	0.7505	0.7507	0.7525	0.7202	0.7136	0.6968	0.6822	**0.6658**
Panel C: $h = 22$									
Q1	0.4362	0.4362	0.4365	0.4338	0.4349	0.4344	0.4209	0.4088	**0.3937**
Median	0.5849	0.5833	0.5848	0.5820	0.5816	0.5783	0.5700	0.5538	**0.5340**
Mean	0.7712	0.7703	0.7707	0.7695	0.7709	0.7699	0.7496	0.7322	**0.7116**
Q3	0.8596	0.8577	0.8597	0.8541	0.8549	0.8543	0.8329	0.8124	**0.7887**

References

Abarbanell, J.S., Bushee, B.J., 1998. Abnormal returns to a fundamental analysis strategy. Accounting Review , 19–45.

Ali, A., Hwang, L.S., Trombley, M.A., 2003. Arbitrage risk and the book-to-market anomaly. Journal of Financial Economics 69, 355–373.

Alwathainani, A.M., 2009. Consistency of firms' past financial performance measures and future returns. The British Accounting Review 41, 184–196.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets 5, 31–56.

Amihud, Y., Mendelson, H., 1986. Liquidity and stock returns. Financial Analysts Journal 42, 43–48.

Anderson, C.W., Garcia-Feijoo, L., 2006. Empirical evidence on capital investment, growth options, and security returns. The journal of finance 61, 171–194.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. The journal of finance 61, 259–299.

Asness, C., Frazzini, A., Gormsen, N.J., Pedersen, L.H., 2020. Betting against correlation: Testing theories of the low-risk effect. Journal of Financial Economics 135, 629–652.

Asness, C.S., Frazzini, A., Pedersen, L.H., 2019. Quality minus junk. Review of Accounting studies 24, 34–112.

Baik, B., Ahn, T., 2007. Changes in order backlog and future returns. Seoul Journal of Business 13, 105–126.

Balakrishnan, K., Bartov, E., Faurel, L., 2010. Post loss/profit announcement drift. Journal of Accounting and Economics 50, 20–41.

Bali, T.G., Brown, S.J., Murray, S., Tang, Y., 2017. A lottery-demand-based explanation of the beta anomaly. Journal of Financial and Quantitative Analysis 52, 2369–2397.

Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. Journal of financial economics 99, 427–446.

Bali, T.G., Engle, R.F., Murray, S., 2016. Empirical asset pricing: The cross section of stock returns. John Wiley & Sons.

Ball, R., Gerakos, J., Linnainmaa, J.T., Nikolaev, V., 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. Journal of Financial Economics 121, 28–45.

Banz, R.W., 1981. The relationship between return and market value of common stocks. Journal of financial economics 9, 3–18.

Barbee Jr, W.C., Mukherji, S., Raines, G.A., 1996. Do sales–price and debt–equity explain stock returns better than book–market and firm size? Financial Analysts Journal 52, 56–60.

Barry, C.B., Brown, S.J., 1984. Differential information and the small firm effect. Journal of financial economics 13, 283–294.

Barth, M.E., Elliott, J.A., Finn, M.W., 1999. Market rewards associated with patterns of increasing earnings. Journal of accounting research 37, 387–413.

Basu, S., 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. The journal of Finance 32, 663–682.

Basu, S., 1983. The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence. Journal of financial economics 12, 129–156.

Belo, F., Lin, X., 2012. The inventory growth spread. The Review of Financial Studies 25, 278–313.

Belo, F., Lin, X., Bazdresch, S., 2014. Labor hiring, investment, and stock return predictability in the cross section. Journal of Political Economy 122, 129–177.

Bhandari, L.C., 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. The journal of finance 43, 507–528.

Blitz, D., Huij, J., Martens, M., 2011. Residual momentum. Journal of Empirical Finance 18, 506–521.

Blume, M.E., Husic, F., 1973. Price, beta, and exchange listing. The Journal of Finance 28, 283–299.

Bouchaud, J., Krüger, P., Landier, A., Thesmar, D., 2019. Sticky Expectations and the Profitability Anomaly. Journal of Finance 74, 639–674.

Boudoukh, J., Michaely, R., Richardson, M., Roberts, M.R., 2007. On the importance of measuring payout yield: Implications for empirical asset pricing. The Journal of Finance 62, 877–915.

Bradshaw, M.T., Richardson, S.A., Sloan, R.G., 2006. The relation between corporate financing activities, analysts' forecasts and stock returns. Journal of accounting and economics 42, 53–85.

Brennan, M.J., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of financial Economics 49, 345–373.

Chan, L.K., Jegadeesh, N., Lakonishok, J., 1996. Momentum strategies. The journal of Finance 51, 1681–1713.

Chan, L.K., Lakonishok, J., Sougiannis, T., 2001. The stock market valuation of research and development expenditures. The Journal of finance 56, 2431–2456.

Chandrashekar, T., Muralidhara, M., Kashyap, K., Rao, P.R., 2009. Effect of growth restricting factor on grain refinement of aluminum alloys. The International Journal of Advanced Manufacturing Technology 40, 234–241.

Chen, J., Hong, H., Stein, J.C., 2002. Breadth of ownership and stock returns. Journal of financial Economics 66, 171–205.

Chordia, T., Subrahmanyam, A., Anshuman, V.R., 2001. Trading activity and expected stock returns. Journal of financial Economics 59, 3–32.

Cohen, L., Diether, K., Malloy, C., 2013. Misvaluing innovation. The Review of Financial Studies 26, 635–666.

Cohen, L., Lou, D., 2012. Complicated firms. Journal of financial economics 104, 383–400.

Cooper, M.J., Gulen, H., Schill, M.J., 2008. Asset growth and the cross-section of stock returns. the Journal of Finance 63, 1609–1651.

Corwin, S.A., Schultz, P., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. The journal of finance 67, 719–760.

Da, Z., Warachka, M., 2011. The disparity between long-term and short-term forecasted earnings growth. Journal of Financial Economics 100, 424–442.

Daniel, K., Titman, S., 2006. Market reactions to tangible and intangible information. The Journal of Finance 61, 1605–1643.

Datar, V.T., Naik, N.Y., Radcliffe, R., 1998. Liquidity and stock returns: An alternative test. Journal of financial markets 1, 203–219.

De Bondt, W.F., Thaler, R., 1985. Does the stock market overreact? The Journal of finance 40, 793–805.

Dechow, P.M., Hutton, A.P., Meulbroek, L., Sloan, R.G., 2001. Short-sellers, fundamental analysis, and stock returns. Journal of financial Economics 61, 77–106.

Dechow, P.M., Sloan, R.G., Soliman, M.T., 2004. Implied equity duration: A new measure of equity risk. Review of Accounting Studies 9, 197–228.

Desai, H., Rajgopal, S., Venkatachalam, M., 2004. Value-glamour and accruals mispricing: One anomaly or two? The Accounting Review 79, 355–385.

Dichev, I.D., 1998. Is the risk of bankruptcy a systematic risk? the Journal of Finance 53, 1131–1147.

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. The journal of finance 57, 2113–2141.

Dimson, E., 1979. Risk measurement when shares are subject to infrequent trading. Journal of financial economics 7, 197–226.

Doyle, J.T., Lundholm, R.J., Soliman, M.T., 2003. The predictive value of expenses excluded from pro forma earnings. Review of Accounting Studies 8, 145–174.

Eisfeldt, A.L., Papanikolaou, D., 2013. Organization capital and the cross-section of expected returns. The Journal of Finance 68, 1365–1406.

Elgers, P.T., Lo, M.H., Pfeiffer Jr, R.J., 2001. Delayed security price adjustments to financial analysts' forecasts of annual earnings. The Accounting Review 76, 613–632.

Eugene, F., French, K., 1992. The cross-section of expected stock returns. Journal of finance 47, 427–465.

Fairfield, P.M., Whisenant, J.S., Yohn, T.L., 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. The accounting review 78, 353–371.

Fama, E.F., French, K.R., 2006. The value premium and the capm. The Journal of Finance 61, 2163–2185.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of financial economics 116, 1–22.

Fama, E.F., French, K.R., 2018. Choosing factors. Journal of financial economics 128, 234–252.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607–636.

Foster, G., Olsen, C., Shevlin, T., 1984. Earnings releases, anomalies, and the behavior of security returns. Accounting Review , 574–603.

Francis, J., LaFond, R., Olsson, P.M., Schipper, K., 2004. Costs of equity and earnings attributes. The accounting review 79, 967–1010.

Frankel, R., Lee, C.M., 1998. Accounting valuation, market expectation, and cross-sectional stock returns. Journal of Accounting and economics 25, 283–319.

Franzoni, F., Marin, J.M., 2006. Pension plan funding and stock market efficiency. the Journal of Finance 61, 921–956.

Frazzini, A., Pedersen, L.H., 2014. Betting against beta. Journal of financial economics 111, 1–25.

George, T.J., Hwang, C.Y., 2004. The 52-week high and momentum investing. The Journal of Finance 59, 2145–2176.

Hafzalla, N., Lundholm, R., Matthew Van Winkle, E., 2011. Percent accruals. The Accounting Review 86, 209–236.

Hahn, J., Lee, H., 2009. Financial constraints, debt capacity, and the cross-section of stock returns. The Journal of Finance 64, 891–921.

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. The Journal of finance 55, 1263–1295.

Haugen, R.A., Baker, N.L., 1996. Commonality in the determinants of expected stock returns. Journal of financial economics 41, 401–439.

Hawkins, E.H., Chamberlin, S.C., Daniel, W.E., 1984. Earnings expectations and security prices. Financial Analysts Journal 40, 24–38.

Heston, S.L., Sadka, R., 2008. Seasonality in the cross-section of stock returns. Journal of Financial Economics 87, 418–445.

Hirshleifer, D., Hou, K., Teoh, S.H., Zhang, Y., 2004. Do investors overvalue firms with bloated balance sheets? Journal of accounting and economics 38, 297–331.

Hou, K., 2007. Industry information diffusion and the lead-lag effect in stock returns. The review of financial studies 20, 1113–1138.

Hou, K., Robinson, D.T., 2006. Industry concentration and average stock returns. The journal of finance 61, 1927–1956.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. The Review of Financial Studies 28, 650–705.

Huang, A.G., 2009. The cross section of cashflow volatility and expected stock returns. Journal of Empirical Finance 16, 409–429.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. The Journal of finance 45, 881–898.

Jegadeesh, N., Kim, J., Krische, S.D., Lee, C.M., 2004. Analyzing the analysts: When do recommendations add value? The journal of finance 59, 1083–1124.

Jegadeesh, N., Livnat, J., 2006. Revenue surprises and stock returns. Journal of Accounting and Economics 41, 147–171.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of finance 48, 65–91.

Jensen, T.I., Kelly, B., Pedersen, L.H., 2023. Is there a replication crisis in finance? The Journal of Finance 78, 2465–2518.

Jiang, G., Lee, C.M., Zhang, Y., 2005. Information uncertainty and expected returns. Review of Accounting Studies 10, 185–221.

Kelly, B., Jiang, H., 2014. Tail risk and asset prices. The Review of Financial Studies 27, 2841–2871.

La Porta, R., 1996. Expectations and the cross-section of stock returns. The Journal of Finance 51, 1715–1742.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. The journal of finance 49, 1541–1578.

Lamont, O., Polk, C., Saaá-Requejo, J., 2001. Financial constraints and stock returns. The review of financial studies 14, 529–554.

Landsman, W.R., Miller, B.L., Peasnell, K., Yeh, S., 2011. Do investors understand really dirty surplus? The Accounting Review 86, 237–258.

Lev, B., Nissim, D., 2004. Taxable income, future earnings, and equity values. The accounting review 79, 1039–1074.

Li, D., 2011. Financial constraints, r&d investment, and stock returns. The Review of Financial Studies 24, 2974–3007.

Litzenberger, R.H., Ramaswamy, K., 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. Journal of financial economics 7, 163–195.

Liu, W., 2006. A liquidity-augmented capital asset pricing model. Journal of financial Economics 82, 631–671.

Lockwood, L., Prombutr, W., 2010. Sustainable growth and stock returns. Journal of Financial Research 33, 519–538.

Loh, R.K., Warachka, M., 2012. Streaks in earnings surprises and the cross-section of stock returns. Management Science 58, 1305–1321.

Lou, D., 2014. Attracting investor attention through advertising. The Review of Financial Studies 27, 1797–1829.

Loughran, T., Wellman, J.W., 2011. New evidence on the relation between the enterprise multiple and average stock returns. Journal of Financial and Quantitative analysis 46, 1629–1650.

Lyandres, E., Sun, L., Zhang, L., 2008. The new issues puzzle: Testing the investment-based explanation. The review of financial studies 21, 2825–2855.

Menzly, L., Ozbas, O., 2010. Market segmentation and cross-predictability of returns. The Journal of Finance 65, 1555–1580.

Miller, M.H., Scholes, M.S., 1982. Dividends and taxes: Some empirical evidence. Journal of Political Economy 90, 1118–1141.

Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? The Journal of finance 54, 1249–1290.

Nguyen, G.X., Swanson, P.E., 2009. Firm characteristics, relative efficiency, and equity returns. Journal of Financial and Quantitative Analysis 44, 213–236.

Novy-Marx, R., 2011. Operating leverage. Review of Finance 15, 103–134.

Novy-Marx, R., 2012. Is momentum really momentum? Journal of financial economics 103, 429–453.

Novy-Marx, R., 2013. The other side of value: The gross profitability premium. Journal of financial economics 108, 1–28.

Ortiz-Molina, H., Phillips, G.M., 2014. Real asset illiquidity and the cost of capital. Journal of Financial and Quantitative Analysis 49, 1–32.

Palazzo, B., 2012. Cash holdings, risk, and expected returns. Journal of Financial Economics 104, 162–185.

Pástor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political economy 111, 642–685.

Penman, S.H., Richardson, S.A., Tuna, I., 2007. The book-to-price effect in stock returns: accounting for leverage. Journal of accounting research 45, 427–467.

Piotroski, J.D., 2000. Value investing: The use of historical financial statement information to separate winners from losers. Journal of accounting research , 1–41.

Pontiff, J., Woodgate, A., 2008. Share issuance and cross-sectional returns. The Journal of Finance 63, 921–945.

Prakash, R., Sinha, N., 2013. Deferred revenues and the matching of revenues and expenses. Contemporary Accounting Research 30, 517–548.

Rajgopal, S., Shevlin, T., Venkatachalam, M., 2003. Does the stock market fully appreciate the implications of leading indicators for future earnings? evidence from order backlog. Review of Accounting Studies 8, 461–492.

Richardson, S.A., Sloan, R.G., Soliman, M.T., Tuna, I., 2005. Accrual reliability, earnings persistence and stock prices. Journal of accounting and economics 39, 437–485.

Ritter, J.R., 1991. The long-run performance of initial public offerings. The journal of finance 46, 3–27.

Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of portfolio management 11, 9–16.

Sloan, R.G., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? Accounting review , 289–315.

Soliman, M.T., 2008. The use of dupont analysis by market participants. The accounting review 83, 823–853.

Stambaugh, R.F., Yuan, Y., 2017. Mispricing factors. The review of financial studies 30, 1270–1315.

Stattman, D., 1980. Book values and stock returns. The Chicago MBA: A journal of selected papers 4, 25–45.

Thomas, J., Zhang, F.X., 2011. Tax expense momentum. Journal of Accounting Research 49, 791–821.

Thomas, J.K., Zhang, H., 2002. Inventory changes and future returns. Review of Accounting Studies 7, 163–187.

Titman, S., Wei, K.J., Xie, F., 2004. Capital investments and stock returns. Journal of financial and Quantitative Analysis 39, 677–700.

Tuzel, S., 2010. Corporate real estate holdings and the cross-section of stock returns. The Review of Financial Studies 23, 2268–2302.

Xie, H., 2001. The mispricing of abnormal accruals. The accounting review 76, 357–373.