Supplement Materials for “Online Forecast Evaluation with High-Frequency Proxies”

Jia Li[*], Yifan Li[†], Andrew J. Patton[‡], Tian Xie[§]

March 1, 2025

Abstract

This supplement contains: (1) extensions of the theoretical results in the paper to multi-period forecasting evaluation problem and multivariate forecast targets; (2) proofs to all theoretical results in the paper; (3) additional empirical results.

[*]School of Economics, Singapore Management University, Singapore.
[†]Accounting and Finance Division, University of Manchester, UK.
[‡]Department of Economics, Duke University, NC, U.S.
[§]College of Business, Shanghai University of Finance and Economics, Shanghai, China.

A Extensions

In this section, we extend the evaluation framework developed thus far in two key directions. The first extension addresses the multi-period evaluation problem. The second extension generalizes the methodology to handle multivariate forecast targets.

A.1 Confidence band and multi-period evaluation

The theory presented in Section 2 focused on the baseline inference problem for a single period t. In practice, however, it is often necessary to jointly evaluate the performance of a sequence of forecasts $(F_t)_{t\in\mathcal{T}}$, indexed by a finite set $\mathcal{T}$. This scenario corresponds to a multiple testing problem. In this subsection, we extend the one-period framework to a multi-period setting. For simplicity, we present the extension in the context of Proposition 2, noting that similar generalizations can be applied to other settings in a straightforward manner. The following generalization of Assumption 3 accommodates the multi-period context:

Assumption 5 *The following conditions hold: (i) $(\hat{R}_t/R_t)_{t\in\mathcal{T}} \xrightarrow{d} (\xi_t)_{t\in\mathcal{T}}$ where the joint distribution of the limit variables is known and continuous; (ii) the loss function $\mathcal{L}(u,\hat{u})$ satisfies the same condition as in Assumption 3.*

The joint convergence imposed by condition (i) of the assumption is not restrictive. Indeed, realized measures across different periods are often constructed from non-overlapping subsamples of high-frequency returns, which renders the standardized estimators asymptotically independent. As a result, joint convergence follows directly from marginal convergence. The limit variables may also be dependent under more general constructions. For instance, in the context of Examples 1 and 2, the joint convergence can be derived when the spot estimators are based on overlapping samples, using the coupling methods of Bollerslev et al. (2021) and Li et al. (2024).

Under the multi-period setting, we test the null hypothesis $H_0 : R_t = F_t$ for all $t \in \mathcal{T}$. A natural test statistic is the "sup" statistic, $\max_{t\in\mathcal{T}} \mathcal{L}(F_t, \hat{R}_t)$, which captures the worst-case proxy loss over $\mathcal{T}$. Mirroring the discussion in Section 2.1, we reject the null hypothesis if $\max_{t\in\mathcal{T}} \mathcal{L}(F_t, \hat{R}_t)$ exceeds the $1-\alpha$ quantile of the limit distribution of the "oracle" worst-case loss, $\max_{t\in\mathcal{T}} \mathcal{L}(R_t, \hat{R}_t)$. This test can be inverted to construct a confidence set for the entire path $(R_t)_{t\in\mathcal{T}}$. Proposition 4 details the formal result.

Proposition 4 *Let $\zeta = \max_{t\in\mathcal{T}} L(\xi_t)$. Under Assumption 5, the following holds:*

$$\mathbb{P}\left(\max_{t\in\mathcal{T}} \mathcal{L}(R_t, \hat{R}_t) \le Q_{\zeta,1-\alpha}\right) \to 1-\alpha.$$

Furthermore, the confidence sets, $(\mathrm{CS}_{1-\alpha,t})_{t\in\mathcal{T}}$, defined as

$$\mathrm{CS}_{1-\alpha,t} = \left[\frac{\hat{R}_t}{\bar{c}(L, Q_{\zeta,1-\alpha})}, \frac{\hat{R}_t}{\underline{c}(L, Q_{\zeta,1-\alpha})}\right], \tag{1}$$

form a $1-\alpha$ level uniform confidence band satisfying

$$\mathbb{P}(R_t \in \mathrm{CS}_{1-\alpha,t}, \forall t \in \mathcal{T}) \to 1-\alpha. \tag{2}$$

Compared to the single-period result in Proposition 2, the key distinction in the multi-period setting is that the critical value $Q_{\zeta,1-\alpha}$ is now determined as the quantile of the maximal limit loss $\zeta = \max_{t\in\mathcal{T}} L(\xi_t)$, which is larger than its single-period counterpart. This inflated critical value is necessary to account for the joint testing across periods, ensuring uniform coverage of the confidence band.

A.2 The case with multivariate forecasting targets

The discussion so far has focused on the univariate case with a scalar-valued forecasting target. In this subsection, we extend the framework to consider R_t as a d-dimensional vector. Prominent examples include the forecasting of (vectorized) integrated covariance matrices or integrated betas (Barndorff-Nielsen and Shephard, 2004), which are particularly relevant in portfolio optimization problems. This naturally aligns with the large-sample framework discussed in Section 2.2.[1] The following generalization of Assumption 1 accommodates the multivariate setting:

Assumption 6 *For some sequence $a_n \to \infty$ and estimator $\hat{\Sigma}_t$: (i) $a_n(\hat{R}_t - R_t) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma_t)$, for some positive definite random matrix $\Sigma_t \in \mathbb{R}^{d\times d}$; (ii) $\hat{\Sigma}_t \xrightarrow{\mathbb{P}} \Sigma_t$.*

Compared to the univariate setting, the multivariate case introduces a key subtlety related to the trade-offs across different components of the vector forecast. This complexity is encoded in the loss function. While many univariate loss functions are "locally equivalent," as shown in Section 2.2, this equivalence no longer holds in the multivariate setting because multivariate loss functions can differ significantly even in the $\hat{R}_t \approx R_t$ local neighborhood. Consequently, the formulation of the loss function in (5), which was deliberately designed to highlight the local equivalence result, is less relevant in the multivariate context. We instead consider a more general class of loss functions specified in the following assumption.

Assumption 7 *The loss function $\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ takes the following form:*

$$\mathcal{L}(u, \hat{u}) = \tilde{L}(\hat{u} - u; u, \hat{u}) + G(u, \hat{u}), \tag{3}$$

where (i) $\tilde{L} : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ is continuous and satisfies $\lambda^p \tilde{L}(x; u, \hat{u}) = \tilde{L}(\lambda x; u, \hat{u})$ for all $\lambda > 0$ and some constant $p > 0$; (ii) $G(R_t, \hat{R}_t) = o_p(a_n^{-p})$ with a_n defined in Assumption 6.

Assumption 7 readily accommodates power-type losses via the leading term $\tilde{L}(\cdot)$. An important special case where the residual term $G(u, \hat{u})$ also plays a role is the Bregman divergence, defined as:

$$\mathcal{L}(u, \hat{u}) = \phi(\hat{u}) - \phi(u) - \nabla\phi(u)^\top(\hat{u} - u), \tag{4}$$

for some three-times differentiable and strictly convex function $\phi : \mathbb{R}^d \to \mathbb{R}$, where $\nabla\phi(\cdot)$ denotes its gradient vector.[2] Indeed, expanding $\phi(\hat{u})$ around $\hat{u} \approx u$ via a Taylor expansion yields:

$$\mathcal{L}(u, \hat{u}) = \frac{1}{2}(u - \hat{u})^\top H_\phi(u)(u - \hat{u}) + G(u, \hat{u}), \tag{5}$$

[1] The existing literature on the small-sample framework has been almost exclusively focused on univariate problems, primarily due to the difficulty of achieving finite-sample pivotalization.

[2] This class of loss functions was introduced by Bregman (1967) and further studied by Laurent et al. (2013) for robust ranking of multivariate volatility forecasts. When $\phi(u) = u^\top W u$ for some positive-definite matrix W, $\mathcal{L}(u, \hat{u})$ simplifies to $(u - \hat{u})^\top W(u - \hat{u})$, representing a general quadratic loss function with weighting matrix W. This formulation encompasses distance metrics such as the Euclidean, Frobenius, and Mahalanobis distances. Additional examples can be found in Laurent et al. (2013).

where $H_\phi(\cdot)$ is the Hessian matrix of $\phi(\cdot)$. Consequently, the Bregman divergence satisfies Assumption 7 with $\tilde{L}(x; u, \hat{u}) = x^\top H_\phi(u)x/2$, $p = 2$, and the implicitly defined residual term satisfying $G(R_t, \hat{R}_t) = o_p(a_n^{-2})$ under Assumption 6.

Since presenting d-dimensional confidence sets is often impractical, we adopt the hypothesis testing perspective. Proposition 5, below, provides a valid critical value for testing the null hypothesis $F_t = R_t$, where $Q_{1-\alpha}(\Sigma, u, \hat{u})$ denotes the $1-\alpha$ quantile of $L(\xi_d; u, \hat{u})$, with $\xi_d \sim \mathcal{N}(0, \Sigma)$ representing a d-dimensional Gaussian random vector.

Proposition 5 *Under Assumptions 6 and 7, we have:*

$$\mathbb{P}\left(\mathcal{L}(R_t, \hat{R}_t) \leq \bar{L}_{1-\alpha}(R_t)\right) \to 1-\alpha,$$

where $\bar{L}_{1-\alpha}(R_t) = a_n^{-p} Q_{1-\alpha}(\hat{\Sigma}_t, R_t, \hat{R}_t)$. *Moreover, the result remains valid if* $\bar{L}_{1-\alpha}(R_t)$ *is replaced by* $\bar{L}_{1-\alpha}(\hat{R}_t)$.

Similar to Proposition 1, Proposition 5 provides a rigorous framework for testing the hypothesis $H_0 : R_t = F_t$, with both $\bar{L}_{1-\alpha}(F_t)$ and $\bar{L}_{1-\alpha}(\hat{R}_t)$ serving as asymptotically valid critical values. Unlike the univariate case, where many loss functions yield asymptotically equivalent inference, inference in the multivariate setting is more nuanced. It depends on how forecasting errors across different components are balanced and how their interdependencies are captured. In the special case where $\mathcal{L}(u, \hat{u}) = (u-\hat{u})^\top W(u-\hat{u})$, this effect is governed by the matrix W, which encapsulates the relative importance and interactions among the components. Consequently, distinct loss functions, reflecting different local trade-offs, generally lead to different inferential outcomes.

B Proofs

Proof of Proposition 1. We first prove the claim in part (a). Under Assumption 2,

$$S\left(R_t, \hat{R}_t\right)\left(\frac{a_n}{\sqrt{\hat{\Sigma}_t}}\right)^p \mathcal{L}(R_t, \hat{R}_t) = L\left(\frac{a_n\left(\hat{R}_t - R_t\right)}{\hat{\Sigma}_t^{1/2}}\right) + \frac{S\left(R_t, \hat{R}_t\right)}{\hat{\Sigma}_t^{p/2}} a_n^p G(R_t, \hat{R}_t),$$

and recall that $a_n^p G(R_t, \hat{R}_t) = o_p(1)$ by Assumption 2. Also, the homogeneity of L implies continuity on $\mathbb{R}$. Then by Assumption 1 and the continuous mapping theorem, we further have $S(R_t, \hat{R}_t)\hat{\Sigma}_t^{-p/2} \xrightarrow{\mathbb{P}} S\left(R_t, R_t\right)\Sigma_t^{-p/2} = O_p(1)$, which implies that the last term of the above equation is of order $o_p(1)$. Meanwhile, $a_n(\hat{R}_t - R_t)/\hat{\Sigma}_t^{1/2} \xrightarrow{d} \xi \sim \mathcal{N}(0,1)$ by Assumption 1 and the properties of stable convergence. By the continuous mapping theorem,

$$S\left(R_t, \hat{R}_t\right)\left(\frac{a_n}{\hat{\Sigma}_t^{1/2}}\right)^p \mathcal{L}(R_t, \hat{R}_t) \xrightarrow{d} L\left(\xi\right).$$

Hence, with $\bar{L}_{1-\alpha}\left(R_t\right) = a_n^{-p} Q_{L(\xi),1-\alpha} \hat{\Sigma}_t^{p/2} / S(R_t, \hat{R}_t) + G(R_t, \hat{R}_t)$,

$$\mathbb{P}\left(\mathcal{L}(R_t, \hat{R}_t) \leq \bar{L}_{1-\alpha}\left(R_t\right)\right) = \mathbb{P}\left(S\left(R_t, \hat{R}_t\right)\left(\frac{a_n}{\sqrt{\hat{\Sigma}_t}}\right)^p \mathcal{L}(R_t, \hat{R}_t) \leq Q_{L(\xi),1-\alpha}\right) \to 1-\alpha.$$

The claim in part (b) is proved by observing

$$\begin{aligned} \mathrm{CS}_{1-\alpha} &\equiv \left\{ r : \mathcal{L}(r, \hat{R}_t) \leq \bar{L}_{1-\alpha}(r) \right\} \\ &= \left\{ r : \frac{L\left(\hat{R}_t - r\right)}{S\left(r, \hat{R}_t\right)} \leq \frac{a_n^{-p} Q_{L(\xi),1-\alpha} \hat{\Sigma}_t^{p/2}}{S\left(r, \hat{R}_t\right)} \right\} \\ &= \left\{ r : L\left(\frac{a_n\left(\hat{R}_t - r\right)}{\hat{\Sigma}_t^{1/2}} \right) \leq Q_{L(\xi),1-\alpha} \right\}. \end{aligned}$$

To see the claim in part (c), we note that the pth-order homogeneity assumption on L implies that $L(x) = k_-|x|^p 1_{\{x\leq 0\}} + k_+|x|^p 1_{\{x\geq 0\}}$ for some constants $k_-, k_+ > 0$. Therefore, if $L(x)$ is symmetric, then $L(x) \propto |x|^p$ which is strictly increasing in $|x|$, and hence part (c) follows from the invariance property of the confidence set. *Q.E.D.*

Proof of Proposition 2. The convexity of L implies continuity on $\mathbb{R}_+$. Therefore, by Assumption 3 and the continuous mapping theorem, $\mathcal{L}(R_t, \hat{R}_t) \xrightarrow{d} L(\xi)$, which leads to the first claim $\mathbb{P}\left(\mathcal{L}(R_t, \hat{R}_t) \leq Q_{L(\xi),1-\alpha}\right) \to 1 - \alpha$ by the assumed continuity of ξ. By the convexity of $L(\cdot)$ and the definitions of $\underline{c}(L, Q)$ and $\overline{c}(L, Q)$, we have

$$\left\{ r : \mathcal{L}(r, \hat{R}_t) \leq Q_{L(\xi),1-\alpha} \right\} = \left\{ r : \underline{c}\left(L, Q_{L(\xi),1-\alpha}\right) \leq \frac{\hat{R}_t}{r} \leq \overline{c}\left(L, Q_{L(\xi),1-\alpha}\right) \right\},$$

which implies the second claim. *Q.E.D.*

Proof of Proposition 3. By Assumption 4, L is pth order homogeneous hence continuous, and the continuous mapping theorem implies that

$$\frac{\mathcal{L}(R_t, \hat{R}_t)}{\hat{\Sigma}_t^{p/2}} = L\left(\frac{\hat{R}_t - R_t}{\hat{\Sigma}_t^{1/2}} \right) \xrightarrow{d} L(\xi).$$

Hence, by the assumed continuity of ξ,

$$\mathbb{P}\left(\frac{\mathcal{L}(R_t, \hat{R}_t)}{\hat{\Sigma}_t^{p/2}} \leq Q_{L(\xi),1-\alpha} \right) \to 1 - \alpha,$$

which implies the first claim of the proposition. By the pth-order homogeneity of L, the confidence set satisfy

$$\left\{ r : L\left(\frac{\hat{R}_t - r}{\hat{\Sigma}_t^{1/2}} \right) \leq Q_{L(\xi),1-\alpha} \right\} = \left\{ r : \underline{c}\left(L, Q_{L(\xi),1-\alpha}\right) \leq \frac{\hat{R}_t - r}{\hat{\Sigma}_t^{1/2}} \leq \overline{c}\left(L, Q_{L(\xi),1-\alpha}\right) \right\},$$

which implies the second claim of the proposition. The third claim follows from the same argument as the proof to part (c) of Proposition 1. *Q.E.D.*

Proof of Proposition 4. Assumption 5 implies that, for each $t \in \mathcal{T}$ we have:

$$\mathcal{L}(R_t, \hat{R}_t) \xrightarrow{d} L(\xi_t), \tag{6}$$

which further implies:

$$\max_{t \in \mathcal{T}} \mathcal{L}(R_t, \hat{R}_t) \xrightarrow{d} \zeta, \tag{7}$$

and hence proves the first claim by the assumed continuity of ξ_t. The second claim follows from the equality of events below and the definition of $\mathrm{CS}_{1-\alpha,t}$:

$$\left\{ r_{\mathcal{T}} : \max_{t \in \mathcal{T}} \mathcal{L}(r_t, \hat{R}_t) \leq Q_{\zeta,1-\alpha} \right\} = \left\{ r_{\mathcal{T}} : \underline{c}(L, Q_{\zeta,1-\alpha}) \leq \frac{\hat{R}_t}{r_t} \leq \bar{c}(L, Q_{\zeta,1-\alpha}), \forall t \in \mathcal{T} \right\}$$
$$= \left\{ r_{\mathcal{T}} : r_t \in \mathrm{CS}_{1-\alpha,t}, \forall t \in \mathcal{T} \right\},$$

as desired. *Q.E.D.*

Proof of Proposition 5. By Assumptions 6 and 7 and the continuous mapping theorem, we have:

$$a_n^p \mathcal{L}(R_t, \hat{R}_t) = \tilde{L}(a_n(\hat{R}_t - R_t); R_t, \hat{R}_t) + o_p(1) \xrightarrow{d} \tilde{L}(\xi_{d,t}; R_t, \hat{R}_t), \tag{8}$$

where $\xi_{d,t} \sim \mathcal{N}(0, \Sigma_t)$. Therefore, with $\bar{L}_{1-\alpha}(R_t) = a_n^{-p} Q_{1-\alpha}(\hat{\Sigma}_t, R_t, \hat{R}_t)$, the above convergence in distribution and the consistency of $\hat{\Sigma}_t$ imply

$$\mathbb{P}\left(\mathcal{L}(R_t, \hat{R}_t) \leq \bar{L}_{1-\alpha}(R_t) \right) \to 1 - \alpha,$$

and the result is unaffected by replacing R_t with its consistent estimator $\hat{R}_t$ in $\bar{L}_{1-\alpha}(R_t)$. *Q.E.D.*

C Additional Empirical Results

We provide additional empirical results to supplement the main analysis. First, we replicate the exercise from the main text using historical data from the SPY ETF as the source data. The empirical results are presented in Table C1.

Building on the analysis of ARM, we then examine five additional stocks that recently began trading and were later included in the S&P 500 index. Table C2 provides detailed information on these stocks. Following the forecasting exercises in the main text, we assess the acceptance rates of spot volatility forecasts for these stocks using the q-like loss function at a 95% confidence level, with SPY ETF data as the source. The empirical results are presented in Table C3.

Table C1: Acceptance Rates by Horizon with SPY Data as Source

Model	Forecast Horizon			
	5 min	1 hour	2 hour	4 hour
Panel A: Q-like Loss				
MCGARCH	0.6875	0.6357	0.6171	0.6033
HAR	0.7479	0.5777	0.5603	0.5430
MCHAR	0.7769	0.7187	0.6972	0.6743
Panel B: Quadratic Loss				
MCGARCH	0.7422	0.6883	0.6707	0.6566
HAR	0.7921	0.6414	0.6141	0.6023
MCHAR	0.8077	0.7577	0.7371	0.7100

Note: The table reports the acceptance rates of spot volatility forecasts for ARM stock using the MCGARCH, HAR, and MCHAR models. The acceptance rate is defined as the proportion of forecasts falling within the 95% evaluation confidence intervals, based on the q-like loss (Panel A) or the quadratic loss (Panel B), over a prediction sample spanning 100 trading days after the company's IPO on September 14, 2023. All models are trained using a 50-day rolling window scheme with 3,900 high-frequency observations. A transfer learning scheme is employed, using SPY ETF data as source data to augment the training sample prior to the IPO.

Table C2: Stock Information and Dates

Symbol	Company Name	Trading Start Date	S&P 500 Inclusion Date
GEV	GE Vernova Inc.	March 27, 2024	April 2, 2024
KVUE	Kenvue Inc.	May 4, 2023	August 9, 2023
SOLV	Solventum Corporation	March 26, 2024	April 1, 2024
SW	Smurfit WestRock PLC	July 8, 2024	July 8, 2024
VLTO	Veralto Corporation	September 27, 2023	October 2, 2023

Table C3: Acceptance Rates by Horizon for Different Target Data

Model	Forecast Horizon			
	5 min	1 hour	2 hour	4 hour
Panel A: GEV				
MCGARCH	0.6547	0.6204	0.6146	0.6056
HAR	0.7176	0.5674	0.5526	0.5470
MCHAR	0.7463	0.6963	0.6756	0.6585
Panel B: KVUE				
MCGARCH	0.5941	0.5557	0.5406	0.5299
HAR	0.6479	0.5598	0.5408	0.5257
MCHAR	0.6602	0.6035	0.5880	0.5721
Panel C: SOLV				
MCGARCH	0.5648	0.5361	0.5300	0.5297
HAR	0.6491	0.5371	0.5262	0.5278
MCHAR	0.6734	0.6325	0.6094	0.6071
Panel D: SW				
MCGARCH	0.6006	0.5582	0.5352	0.5230
HAR	0.6956	0.5910	0.5835	0.5766
MCHAR	0.7197	0.6709	0.6574	0.6458
Panel E: VLTO				
MCGARCH	0.6056	0.5764	0.5679	0.5576
HAR	0.6797	0.5694	0.5613	0.5573
MCHAR	0.7128	0.6745	0.6602	0.6535

Note: The table reports the acceptance rates of spot volatility forecasts for five recently listed stocks—GEV, KVUE, SOLV, SW, and VLTO—using the MCGARCH, HAR, and MCHAR models. Each panel corresponds to a specific stock, with forecasts evaluated using the q-like loss function at a 95% confidence level. The acceptance rate is defined as the proportion of forecasts falling within the confidence intervals over a prediction sample spanning 100 trading days after each stock's trading debut. A transfer learning scheme is employed, utilizing SPY ETF data as source data to augment the training sample prior to each stock's listing.

References

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):pp. 885–925.

Bollerslev, T., Li, J., and Liao, Z. (2021). Fixed-k inference for volatility. *Quantitative Economics*, 12(4):1053–1084.

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217.

Laurent, S., Rombouts, J. V. K., and Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, 173(1):1–10.

Li, J., Wang, D., and Zhang, Q. (2024). Reading the Candlesticks: An OK Estimator for Volatility. *Review of Economics and Statistics*, 106(4):1114–1128.