Dynamic Covariance Matrix Estimation and Portfolio Analysis with High-Frequency Data

Binyan Jiang[†], Cheng Liu[‡], Cheng Yong Tang[§]

[†]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. Phone: (852) 2766 6349. E-mail: by.jiang@polyu.edu.hk

[‡]School of Economics and Management, Wuhan University, Wuhan, Hubei 430072, China. Phone: (86) 6875 3034. E-mail: chengliu_eco@whu.edu.cn. Liu gratefully acknowledges research support from Humanities and Social Science Foundation of China Education Ministry, 20YJC790074

[§]Department of Statistical Science, Temple University, Philadelphia, PA 19122-6083, USA. Phone: (215) 204 3191. E-mail: yongtang@temple.edu

Abstract

Variances and covariances of multiple financial returns play foundational roles in many financial applications. Rather than being homogeneous, covariance matrices are well known to be dynamic. To address such heterogeneity in a setting with high-frequency noisy data contaminated by measurement errors, we propose a new approach for estimating the dynamic covariances with high-dimensional series of data. By utilizing an appropriate localization, our approach attempts to incorporate generic accompanying variables in order to account for the heterogeneity. We then analyze the high-dimensional global minimal-variance sparse portfolio construction that incorporates the proposed dynamic covariance estimator. Our theory establishes the validity of our methods in handling high-dimensional, high-frequency noisy data. The promising performance of our methods is illustrated by extensive simulations and empirical studies. In particular, we demonstrate that the accuracy of the covariance estimator and the portfolio performance are substantially improved by accounting for the dynamics.

Keywords: Dynamic covariance estimation; Global minimal-variance sparse portfolio; High-frequency data analysis; High-dimensional data analysis; Measurement errors.

1 Introduction

Variances and covariances play elementary roles in finance; see Markowitz (1952), Merton (1990), and Back (2010) for a recent overview. Modern practical problems usually involve series of high-dimensional data. Extensive studies have focused on handling large covariance matrices; see Bickel and Levina (2008a,b) and the reviews of Fan et al. (2016) and Cai (2017). Recently, dealing with the covariance matrix of multivariate and high-dimensional time-series data has received increasing attention; see the monographs Tsay (2013) and Tsay and Chen (2019) and references therein.

Use of the sample covariance matrix is standard for investigating a large covariance matrix. Clearly, a foundational assumption for the validity of a study based on the sample covariance matrix is that replications of the random vectors share the same distribution, that is, there is an assumption of a homogeneous distribution. However, in practice, such homogeneity may be questionable; see the recent work by Chen and Leng (2016) addressing this issue. Furthermore, the covariations between the random variables may be explained by connections with other variables. In finance, this kind of observation has also been documented; see Brandt (1999) and Aït-Sahalia and Brandt (2001) for approaches that handle this issue with conditioning variables.

For handling of heterogeneity of the covariance matrix with conditioning variables, Chen and Leng (2016) recently proposed a kernel-smoothing approach to estimate the dynamic covariance matrices; see also Jiang et al. (2020) and Wang et al. (2020). More recently, Chen et al. (2019) developed an approach for handling of many conditioning variables whose number may diverge. We observe that in the existing literature, dynamic covariance estimation is developed in the context of so-called low frequency (e.g., daily or weekly) data with a relatively long time span that covers months and years.

Besides low-frequency data, high-frequency data such as within-day tick-by-tick trading/quote data are becoming increasingly available in practice. Covariance estimation with high-frequency data is an actively developing topic; see Aït-Sahalia et al. (2010), Tao et al. (2013), Aït-Sahalia and Xiu (2017), Aït-Sahalia and Xiu (2018), and references therein. Applications in portfolio construction with the covariance estimated from high-frequency data have been investigated in Fan et al. (2012) and Cai et al. (2020). Nonetheless, this area remains less explored in the context of high-frequency data where dynamic covariance estimation is used and the aforementioned possible heterogeneity is addressed.

In the context of covariance estimations, features and challenges of high-frequency data are unique. On one hand, thanks to high-frequency data, recovering the realized value of the latent stochastic volatility and the covariances becomes possible; see Hansen and Lunde (2006) and Aït-Sahalia et al. (2005). On the other hand, high-frequency data are contaminated by measurement errors from sources such as the market microstructure noises. For an overview of methods and theory for handling of noisy high-frequency data, we refer to the monographs Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014) and reference therein. For recovering the stochastic realized variances and covariances from high-frequency data, numerous approaches are available, for example the two-scale realized variance (TSRV) approach (Zhang et al., 2005), the multiple-scale realized variance (MSRV) approach (Zhang, 2006), the quasi-maximum likelihood approach (QMLE) (Xiu, 2010; Aït-Sahalia et al., 2010; Liu and Tang, 2014), and the realized kernel approach (Barndorff-Nielsen et al., 2011).

Both the classic and the current literature support a well-known discovery that the dynamics of the financial returns in the market are driven by some common systematic sources; see the classical studies of Markowitz (1952) and Fama and French (1993). Current investigations with high-frequency data are revealing analogous phenomena; see Aït-Sahalia and Xiu (2017), Aït-Sahalia and Xiu (2018), and Dai et al. (2019). Such common factors may account for the aforementioned heterogeneity in the dynamic variances and covariances. Also, abundant information is available in addition to the high-frequency data alone. Such extra information may provide powerful support in attempting to identify the common sources driving the dynamics. For example, the CBOE volatility index (VIX) and prices of its tradable contingents such as futures and options are available. Since they can be viewed as market forward-looking volatility measures, incorporating such information is expected to improve the accuracy of the variance and covariance estimations.

We are thus motivated to conduct this investigation, which considers dynamic covariance estimations with high-frequency high-dimensional data. Our approach is built upon first recovering the realized stochastic variances and covariances of the high-dimensional variables from noisy high-frequency data over a short period of time. Then we propose a localized covariance function estimation based on realized high-dimensional covariance matrices over multiple short periods of time. The localization is developed by conditioning on information that accompanies the realized high-dimensional covariance matrices. Our approach provides a new method for covariance estimation with high-frequency data which accounts for the heterogeneity. We attempt to effectively combine information from the realized covariance matrices over multiple periods of time. Compared with an approach that uses simple averaging, our localizing approach is capable of incorporating heterogeneity in the data via appropriate conditioning. Since the conditioning information does not have to be high frequency, our approach also provides a new method for combining both high-frequency and low-frequency information; for studies that combines multiple frequency information, see Tao et al. (2011). Our theoretical analysis reveals that dynamic covariance estimation is valid in a setting with measurement errors that are contaminating the price data. From a technical standpoint, establishing the theory of the proposed approach is challenging because of the additional dimension of uncertainty which is due to the recovery of the stochastic covariance matrices from high-frequency data contaminated by measurement errors. We show that in a high-dimensional setting the elementwise maximum discrepancy diminishes to zero in a setting in which the number of assets in the portfolio grows exponentially with the number of observations.

Based on the dynamic covariance estimated from high-frequency data, we pursue further analysis of portfolio allocation problems. Starting from the pioneering work of Markowitz (1952), extensive studies have focused on the mean-variance (MV) efficient portfolio and the global minimal-variance (GMV) portfolio. DeMiguel et al. (2009b) compared the naive $1/N$ portfolio with sample-based sophisticated MV portfolios and found that the naive $1/N$ portfolio could be more efficient. Tu and Zhou (2011) proposed to improve the performances of sophisticated mean-variance rules by combining them with the $1/N$ rule. Ledoit and Wolf (2017) improved the out-of-sample performance of sampled-based portfolios by developing a better covariance matrix estimator called the nonlinear shrinkage of the covariance matrix. For the GMV portfolio, Jagannathan and Ma (2003) showed that imposing no-short-sale constraints could reduce the risk of the portfolio. DeMiguel et al. (2009a) tried to improve the performance of sample-based GMV portfolios by adding a constraint on the norm of the portfolio-weight vector. In this study, we set out to improve the sample-based GMV portfolio by adding a 1-norm constraint and considering the dynamics of the covariance matrix. We propose a new approach to covariance estimation using high-frequency data. By using conditioning variables, our new estimator achieves a balance between the heterogeneity of the covariance matrix and the level of variation in the estimators. In addition to providing a theoretical justification of the validity of our approach, we show that its performance is promising, by providing extensive numerical examples of both simulations and real-data analysis.

The rest of this paper is structured as follows. The methodology for dynamic covariance estimation with high-frequency high-dimensional data is outlined in Section 2. The global minimum-variance sparse portfolio construction that incorporates the proposed dynamic covariance estimation is presented in Section 3. Section 4 contains all the theoretical results. The results of simulation and empirical studies are given in Sections 5 and 6, respectively. All proofs are provided in the Appendix.

2 Dynamic integrated covariance modeling

2.1 Overview

Following the convention in the literature and practice, we consider a multivariate stochastic process $\mathbf{X}_t = (X_{1,t}, \cdots, X_{p,t})'$ for the log-prices of equities or other tradable instruments. We work with a general model which assumes that $\mathbf{X}_t$ follows

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Sigma}_t^{1/2} d\mathbf{B}_t, \tag{1}$$

where $\mathbf{B}_t$ is a p-dimensional continuous-time standard stochastic process with stationary and independent increments (e.g., standard Brownian motion), $\boldsymbol{\mu}_t = (\mu_{1,t}, \cdots, \mu_{p,t})'$ is the drift, and $\boldsymbol{\Sigma}_t$ is the instantaneous variance-covariance matrix, which is allowed to be stochastic. We are interested in investigating problems that involve $\boldsymbol{\Sigma}_t$ in a setting with high-frequency data.

From known results in the literature based on high-frequency observations of $\mathbf{X}_t$, a realization of the integrated $\boldsymbol{\Sigma}_t$ over a short period of time—say one trading day—can be consistently recovered, even though the observations may be contaminated by measurement errors; see Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014). In our study, we consider multiple periods of time with each period of length δ, where δ is typically equivalent to one trading day. Then we refer to the integrated covariance matrix (ICM) over δ as

$$\overline{\boldsymbol{\Sigma}}_{k,\delta} := \int_{(k-1)\delta}^{k\delta} \boldsymbol{\Sigma}_t dt, \quad k = 1, \ldots, N. \tag{2}$$

Here, the use of multiple k's in (2) recognizes the fact that such an evaluation may occur over multiple periods of time. If $\boldsymbol{\Sigma}_t$ is stochastic, $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}\}_{k=1}^N$ forms a sequence of random matrices that are unobservable in practice but consistently recoverable with high-frequency data.

The series $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}\}_{k=1}^N$ is informative in a few aspects. Foremost among them is that δ is typically small, so that it precisely reflects the short-term and even spot volatility and covariations between the prices. Therefore, compared with a return series calculated with only the daily opening and closing prices, $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}\}_{k=1}^N$ is ideally a more accurate account. Second, the series $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}\}_{k=1}^N$ over multiple periods of time comprehensively reflects the behavior of the variances and covariances between the equity prices. For example, under reasonable assumptions, the increments of $\mathbf{X}_t$ are stationary, so that the series $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}\}_{k=1}^N$ is expected to be informative for recovering the stationary variance-covariance matrix. In the context of the classical capital asset pricing model (CAPM), such variance-covariance is the foundation for a portfolio theory; see Markowitz (1952).

Nevertheless, the stationary assumption of a series is questionable, especially when the market is experiencing huge between-day changes in volatility. For example, in the last two weeks of March 2020, the between-day volatility level changed drastically as a results of the huge market turmoil. In such a period of time, the assumption of between-day stationarity seems shaky. Even worse, the assumption of local stationarity—stationarity within a short period of time—also becomes questionable when between-day volatility changes are so drastic. Therefore, the previous day's ICM is less informative, and even counter-productive, in predicting the next day's volatility and covariations.

Market information is broad and provides opportunities for investments. Concretely, we denote the random quantities by $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}, \mathbf{U}_k\}$, where $\mathbf{U}_k$ accompanies the ICM and is allowed to be stochastic too. For instance, the CBOE VIX index is instantaneously available, which is an ideal candidate for $\mathbf{U}_k$. Since dependence between $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ and $\mathbf{U}_k$ is anticipated, an estimation of $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ given $\mathbf{U}_k$ should more adequately capture the dynamics in the random ICMs. Specifically, we propose to consider the ICM as a dynamic function:

$$\boldsymbol{\Sigma}_\delta(\mathbf{u}) = \mathbb{E}\{\overline{\boldsymbol{\Sigma}}_{k,\delta} | \mathbf{U}_k = \mathbf{u}\}. \tag{3}$$

Clearly, $\boldsymbol{\Sigma}_\delta(\mathbf{u})$ is dynamic with $\mathbf{u}$—an attempt to incorporate information from $\mathbf{U}_k$. The more informative $\mathbf{U}_k$ is, the more effective $\boldsymbol{\Sigma}_\delta(\mathbf{u})$ is in reflecting the variance-covariance of the equity prices. We note that our focus in (3) does not require stationarity or local stationarity of $\overline{\boldsymbol{\Sigma}}_{k,\delta}$; in fact, it can actually be viewed as a generalization of local stationarity, more broadly localizing via $\mathbf{U}_k$. In practice, the choices of $\mathbf{U}_k$ in our framework are broad and include localizing at a given time. As an example, the CBOE VIX and its contingency instruments—tradable futures and options—are natural candidates for investigating the dynamic covariance matrix.

In the context of the conventional CAPM, conditional modeling has received attentions; see Jagannathan and Wang (1996), Lewellen and Nagel (2006), and Petkova and Zhang (2005). The conditional CAPM recognizes the fact that betas and risk premiums may be time dependent, and it attempts approaches for incorporating such heterogeneity. In contrast, our framework concerns the variance-covariance between the log-prices, focusing on the possible heterogeneity that may depend on broadly defined $\mathbf{U}_k$.

We denote the conditioning variable that accompanies the time interval $((k-1)\delta, k\delta]$ by $\mathbf{U}_k$. The true underlying process in our setting is thus $(\overline{\boldsymbol{\Sigma}}_{k,\delta}, \mathbf{U}_k)$ $(k = 1, \ldots, N)$. In practice, $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}\}$ is not directly observable. Thanks to the opportunities enabled by high-frequency data, consistently recovering $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ is feasible. Our practical objective is thus to estimate $\boldsymbol{\Sigma}_\delta(\mathbf{u})$ with the sequence $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}\}$ that is recovered from high-frequency data.

2.2 Recovering $\{\overline{\mathbf{\Sigma}}_{k,\delta}\}$

We denote by $\overline{\sigma}_{k,\delta}^{(i,j)}$ the (i,j)th element of $\overline{\mathbf{\Sigma}}_{k,\delta}$ $(i,j = 1,\dots,p)$. We consider a setting with high-frequency observations over each period of length δ, for a total of N periods. Thus the entire range of time we consider is $T = N\delta$. Over each time interval $((k-1)\delta, k\delta]$ $(k = 1,\dots,N)$, we denote by $\{t_{k,1},\dots,t_{k,n_k}\}$ the times at which the prices are observed. Under model (1) and as $n_k \to \infty$, $\overline{\sigma}_{k,\delta}^{(i,j)}$ can be consistently approximated by

$$[X_i, X_j]_{(k)} := \sum_{l=1}^{n_k-1} [X_{i,t_{k,l+1}} - X_{i,t_{k,l}}][X_{j,t_{k,l+1}} - X_{j,t_{k,l}}], \tag{4}$$

provided that $\max_j\{(t_{k,j} - t_{k,j-1})\} \to 0$, thanks to the benefit of high-frequency data. For ease of presentation, we start with the ideal case where the observations of all processes are synchronous. Our method and theory are also applicable to asynchronous data—data-synchronization and pairwise-estimation approaches are applicable for handling of asynchronous data.

High-frequency data are commonly contaminated by measurement errors; see Hansen and Lunde (2006), Aït-Sahalia et al. (2005), and Zhang et al. (2005), among others. To incorporate the measurement errors, we consider the additive error model for the observations:

$$Y_{i,t} = X_{i,t} + \epsilon_{i,t}, \quad i = 1,\dots,p, \tag{5}$$

where t represents a time at which the price is observed and the $\epsilon_{i,t}$'s are random errors.

Because of the contamination by the measurement errors, the realized covariance (4) becomes seriously biased when $Y_{i,t}$ is observed instead of $X_{i,t}$. Recovering the ICM from the observed $Y_{i,t}$ with model (5) has been extensively studied. Numerous approaches are available, for example the two scales covariance estimator (Zhang, 2011), the quasi-maximum likelihood estimator (QMLE) (Aït-Sahalia et al., 2010; Liu and Tang, 2014), and the realized kernel approach (Barndorff-Nielsen et al., 2011).

Concretely in our development, to handle the measurement errors in the high-frequency observations, we adopt the pair-wise two scales covariance estimator of Zhang (2011) to recover $\overline{\mathbf{\Sigma}}_{k,\delta}$ for the time interval $((k-1)\delta, k\delta]$. We note that our framework is equally applicable with the sequence $\{\overline{\mathbf{\Sigma}}_{k,\delta}\}$ consistently recovered from contaminated high-frequency data using other approaches.

To accommodate asynchronous data, we denote by $T_k = \{t_{k,1},\dots,t_{k,n_k}\}$ the set of all times in the interval $((k-1)\delta, k\delta]$ at which at least one price process is observable. Let $T_k^{(i)} \subset T_k$ be the set of times at which the ith process is observed in the period k, $(i = 1,\dots,p; k = 1,\dots,N)$.

The dependence of $T_k^{(i)}$ on both k and i allows for asynchronous data and for unequal numbers of observations over multiple days. We denote the observations in this interval by $Y_{i,t}$ ($t \in T_k^{(i)}, i = 1, \ldots, p, k = 1, \ldots, N$).

To address data asynchronicity, we first synchronize the data using the pairwise-refresh time scheme of Barndorff-Nielsen et al. (2011). For the ith and jth assets, the set of pairwise-refresh times is given by $V_k = \{v_{k,0}, v_{k,1}, \ldots, v_{k,n_{ij,k}}\}$, where $v_{k,0} = (k-1)\delta$ and for all $l > 0$,

$$v_{k,l} = \max\left[\min\left\{t \in T_k^{(i)} : \ t > v_{k,(l-1)}\right\}, \min\left\{t \in T_k^{(j)} : \ t > v_{k,(l-1)}\right\}\right].$$

Hence $n_{ij,k}$ is the number of observations after synchronizing the observations of the ith and jth assets by the pairwise-refresh time scheme.

As a result of the refresh times, the actual sampling times of the ith and jth assets are respectively

$$\tilde{t}_{k,l}^{(i)} = \max\left\{t \in T_k^{(i)}, t \le v_{k,l}\right\}, \quad \tilde{t}_{k,l}^{(j)} = \max\left\{t \in T_k^{(j)}, t \le v_{k,l}\right\}.$$

The two-scale method estimates $\overline{\sigma}_{k,\delta}^{(i,j)}$ by

$$\widetilde{\sigma}_{k,\delta}^{(i,j)} = [Y_i, Y_j]_{(k)}^{(K)} - \frac{J(n_{ij,k} - K + 1)}{K(n_{ij,k} - J + 1)}[Y_i, Y_j]_{(k)}^{(J)}, \tag{6}$$

where

$$[Y_i, Y_j]_{(k)}^{(K)} = \frac{1}{K}\sum_{l=K}^{n_{ij,k}}\left(Y_{i,\tilde{t}_{k,l}^{(i)}} - Y_{i,\tilde{t}_{k,l-K}^{(i)}}\right)\left(Y_{j,\tilde{t}_{k,l}^{(j)}} - Y_{j,\tilde{t}_{k,l-K}^{(j)}}\right).$$

Zhang (2011) suggested choosing $K = O(n_{ij,k}^{2/3})$ and setting J to a constant. In this paper, we simply set J to 1.

2.3 Estimating $\boldsymbol{\Sigma}_\delta(\mathbf{u})$

From (6), we get $\widetilde{\boldsymbol{\Sigma}}_{k,\delta} = (\widetilde{\sigma}_{k,\delta}^{(i,j)})_{1\le i,j\le p}$ ($k = 1, \ldots, N$). Combining this with other dynamic variables, we have $\{\widetilde{\boldsymbol{\Sigma}}_{k,\delta}, \mathbf{U}_k\}_{k=1}^N$, which we use to estimate $\boldsymbol{\Sigma}_\delta(\mathbf{u})$ in (3). We propose to estimate $\boldsymbol{\Sigma}_\delta(\mathbf{u})$ by localizing via kernel smoothing:

$$\widehat{\boldsymbol{\Sigma}}_\delta(\mathbf{u}) = \frac{\sum_{k=1}^N K_h(\mathbf{U}_k - \mathbf{u})\widetilde{\boldsymbol{\Sigma}}_{k,\delta}}{\sum_{k=1}^N K_h(\mathbf{U}_k - \mathbf{u})}, \tag{7}$$

where $K_h(\cdot)$ is a kernel function with bandwidth h.

Clearly, if the $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ in (2) replaces the $\widetilde{\boldsymbol{\Sigma}}_{k,\delta}$ in (7), this becomes the conventional kernel smoothing, which is an unrealistic version. Since $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ itself is stochastic, replacing it with an estimator $\widetilde{\boldsymbol{\Sigma}}_{k,\delta}$ from high-frequency data introduces new features of our study. In particular, in the context of kernel smoothing, our estimator (7) is new and handles the response variables that are recovered with the estimation errors. Additionally, if p is large, (7) is a high-dimensional matrix-valued function estimations—another challenge of the estimation problem. As demonstrated by our theory, (7) converges uniformly to our target (3) if technical challenges due to the estimation procedure (6) with high-frequency data and high data dimensionality are handled appropriately.

Our covariance estimator (7) is dynamic, incorporating information via $\mathbf{U}_k$. Dynamic covariance estimation has received attention recently; see, for example, Chen and Leng (2016), Chen et al. (2019), and Jiang et al. (2020). Despite the fact that our method also exploits the dynamics via the localization, we broaden the scope of dynamic covariance estimation by incorporating a new dimension induced by the estimation errors. As shown in the next section, we propose a new approach using (7) in constructing a high-dimensional portfolio. Without requiring further regularization in (7), we show that a sparse optimal portfolio can be delivered with our dynamic covariance estimation approach. Such a scope is different from that in Chen and Leng (2016), who proposed to further regularize the high-dimensional dynamic covariance estimator.

3 Portfolio analysis with high-dimensional dynamic ICM

We are interested in constructing a portfolio based on tradable $\mathbf{X}_t \in \mathbb{R}^p$. A portfolio at time T is seen to be equivalent to a vector $\mathbf{w}_T \in \mathbb{R}^p$ that satisfies $\mathbf{w}_T'\mathbf{1} = 1$, where $\mathbf{1}$ is a p-dimensional vector all of whose components are equal to 1. We consider a generic re-balancing horizon τ, where the portfolio will be held for a period of time τ after T. Here, τ is finite, and for simplicity we assume that both $T/\delta = N$ and $\tau/\delta = N^*$ are integers. Thus over the period $[T, T+\tau]$ the portfolio has a return $\mathbf{w}_T' \left(\int_T^{T+\tau} d\mathbf{X}_t\right)$, and its expected risk is $R_{T,\tau}(\mathbf{w}_T) = \mathbf{w}_T'\mathbf{\Sigma}^*_{T,\tau}\mathbf{w}_T$, where

$$\mathbf{\Sigma}^*_{T,\tau} = \mathbb{E}_T\left(\int_T^{T+\tau} \mathbf{\Sigma}_t dt\right) = \mathbb{E}_T\left(\sum_{k=1}^{N^*} \overline{\mathbf{\Sigma}}_{(N+k),\delta}\right) \tag{8}$$

and $\mathbb{E}_T$ denotes the expectation which is taken by conditioning on information up to time T. Therefore, once $\mathbf{\Sigma}^*_{T,\tau}$ is obtained, the portfolio allocation strategy can be applied upon getting the weights $\mathbf{w}_T$. We observe that

$$\mathbb{E}_T\left\{\sum_{k=1}^{N^*} \overline{\mathbf{\Sigma}}_{(N+k),\delta}\right\} = \mathbb{E}_T\left\{\sum_{k=1}^{N^*} \mathbb{E}\left(\overline{\mathbf{\Sigma}}_{(N+k),\delta}\middle|\, \mathbf{U}_{N+k}\right)\right\} = \mathbb{E}_T\left\{\sum_{k=1}^{N^*} \mathbf{\Sigma}_\delta(\mathbf{U}_{N+k})\right\}.$$

Hence by choosing $\mathbf{U}_{N+k}$, the variable associated with the kth period after T, as some $\mathcal{F}_T$-measurable quantity, we propose to estimate $\mathbf{\Sigma}^*_{T,\tau}$ by

$$\widehat{\mathbf{\Sigma}}^*_{T,\tau} = \sum_{k=1}^{N^*} \widehat{\mathbf{\Sigma}}_\delta(\mathbf{U}_{N+k}), \tag{9}$$

where $\widehat{\mathbf{\Sigma}}_\delta(\mathbf{u})$ is given by (7). Here, we note that $\mathbf{U}_{N+k}$ should be taken to be $\mathcal{F}_T$-measurement, and constructing such a quantity is conceptually feasible. For example, in some settings where $\mathbf{U}_{N+k}$ itself is not observable, a convenient choice in our study is $\widehat{\mathbf{U}}_{N+k} = \mathbb{E}_T(\mathbf{U}_{N+k})$, which is the predicted value of $\mathbf{U}_{N+k}$.

We consider a high-dimensional portfolio construction where p is large. A realistic strategy is to construct a sparse portfolio, one that targets at a $\mathbf{w}$ many of whose components are 0; see Fan et al. (2012) and references therein. In particular, for a given tuning parameter c, the global minimum-variance (GMV) sparse portfolio is constructed by

$$\min \mathbf{w}'\mathbf{\Sigma}^*_{T,\tau}\mathbf{w} \quad \text{s. t.} \quad \mathbf{w}'\mathbf{1} = 1 \quad \text{and} \quad \|\mathbf{w}\|_1 \le c. \tag{10}$$

We note that this is equivalent to identifying the optimal direction of $\mathbf{w}$, which we denote by $\boldsymbol{\beta}$ and then normalizing it via $\mathbf{w} = \boldsymbol{\beta}/(\boldsymbol{\beta}'\mathbf{1})$. To appreciate this, in a case without the constraint $\|\mathbf{w}\|_1 \le c$ the solution of (10) is $\mathbf{w}_{T,\tau} := \frac{\mathbf{\Sigma}^{*-1}_{T,\tau}\mathbf{1}}{\mathbf{1}'\mathbf{\Sigma}^{*-1}_{T,\tau}\mathbf{1}}$. We consider $\boldsymbol{\beta}_{T,\tau} := \mathbf{\Sigma}^{*-1}_{T,\tau}\mathbf{1}$, and $\mathbf{w}_{T,\tau}$ can be obtained by normalizing $\boldsymbol{\beta}_{T,\tau}$.

Note that $\boldsymbol{\beta}_{T,\tau}$ is the solution of $\min \ell(\boldsymbol{\beta}) := \boldsymbol{\beta}'\mathbf{\Sigma}^*_{T,\tau}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{1}$. By the convexity of the objective function and the Kuhn-Tucker Theorem, there exists a constant $\lambda > 0$ such that the GMV sparse portfolio formulation (10) is equivalent to solving the following l_1 penalized formulation up to a normalization:

$$\boldsymbol{\beta}^\lambda_{T,\tau} := \operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p}\{\boldsymbol{\beta}'\mathbf{\Sigma}^*_{T,\tau}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{1} + \lambda\|\boldsymbol{\beta}\|_{\mathbf{1}}\}. \tag{11}$$

Specifically, ignoring the normalization constraint $\mathbf{w}'\mathbf{1} = 1$: if $c \ge \|\mathbf{w}_{T,\tau}\|_1$, then (10) is equivalent to (11) with $\lambda = 0$; and if $c \to 0$, then (10) is equivalent to (11) with $\lambda \to +\infty$. Otherwise, we have $c = \frac{\|\boldsymbol{\beta}^\lambda_{T,\tau}\|_1}{|\mathbf{1}'\boldsymbol{\beta}^\lambda_{T,\tau}|}$. Similar to c, λ controls the sparsity of the portfolio allocation.

For a given sparsity parameter λ, we then propose to estimate the sparse portfolio direction by

$$\widehat{\boldsymbol{\beta}}^\lambda_{T,\tau} := \operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p}\boldsymbol{\beta}'\widehat{\mathbf{\Sigma}}^*_{T,\tau}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{1} + \lambda\|\boldsymbol{\beta}\|_1. \tag{12}$$

We remark that since $\widehat{\mathbf{\Sigma}}_\delta(\mathbf{u})$ is constructed elementwise based on the two scales covariance estimator (Zhang, 2011), it may not be positive semidefinite. To ensure that $\widehat{\mathbf{\Sigma}}^*_{T,\tau}$ is positive semidefinite in practice, inspired by the approach of Fan et al. (2012), we apply the following projection procedure to $\widehat{\boldsymbol{\Sigma}}_\delta(\mathbf{u}_{N+k})$, $k = 1, \ldots, N^*$. Here, $\mathbf{u}_{N+k}$ is the observation or an estimator of $\mathbf{U}_{N+k}$. Specifically, for any $p \times p$ symmetric matrix $\mathbf{A}$, we denote by $\mathbf{D}$ the diagonal matrix with the same diagonal elements as $\mathbf{A}$, and we define $\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. Suppose the eigen-decomposition of $\mathbf{R}$ is given by $\mathbf{R} = \boldsymbol{\Gamma}'\text{diag}(\lambda_1, ...\lambda_p)\boldsymbol{\Gamma}$, where $\lambda_1, ..., \lambda_p$ are the eigenvalues and $\boldsymbol{\Gamma}$ is an orthogonal matrix with the p eigenvectors of $\mathbf{R}$ as its columns. Define

$$\mathbf{A}^+ = \mathbf{D}^{1/2}\left\{(\mathbf{R} + \lambda_{\min}^-\mathbf{I}_p)/(1 + \lambda_{\min}^-)\right\}\mathbf{D}^{1/2},$$

where $\lambda_{\min}^-$ is the absolute value of the minimum eigenvalue of $\mathbf{R}$ if that eigenvalue is less than 0, and $\lambda_{\min}^- = 0$ otherwise, and $\mathbf{I}_p$ denotes the $p \times p$ identity matrix.

Note that (12) is an l_1-penalized quadratic function of $\boldsymbol{\beta}$. After obtaining $\widehat{\boldsymbol{\Sigma}}^*_{T,\tau}$, (12) can be efficiently solved using, for example, coordinate descent algorithms. We can then estimate the sparse portfolio $\mathbf{w}^\lambda_{T,\tau} := \frac{\boldsymbol{\beta}^\lambda_{T,\tau}}{\mathbf{1}'\boldsymbol{\beta}^\lambda_{T,\tau}}$ by $\widehat{\mathbf{w}}^\lambda_{T,\tau} = \frac{\widehat{\boldsymbol{\beta}}^\lambda_{T,\tau}}{\mathbf{1}'\widehat{\boldsymbol{\beta}}^\lambda_{T,\tau}}$. Denote the conditional minimum variance given $\mathbf{u}_{N+1}, \ldots, \mathbf{u}_{N+N^*}$ by $R_{T,\tau} = (\mathbf{w}^\lambda_{T,\tau})'\boldsymbol{\Sigma}^*_{T,\tau}\mathbf{w}^\lambda_{T,\tau}$. We can then estimate $R_{T,\tau}$ by $\widehat{R}^\lambda_{T,\tau} = (\widehat{\mathbf{w}}^\lambda_{T,\tau})'\widehat{\boldsymbol{\Sigma}}^*_{T,\tau}\widehat{\mathbf{w}}^\lambda_{T,\tau}$. We shall see in Section 4 that, under mild conditions, uniform consistency results for $\widehat{\boldsymbol{\beta}}^\lambda_{T,\tau}$, $\widehat{\mathbf{w}}^\lambda_{T,\tau}$ and $\widehat{R}^\lambda_{T,\tau}$ are achievable.

We remark that the $\boldsymbol{\Sigma}^*_{T,\tau}$ in (8) is the key to the portfolio construction, hence the estimation of $\boldsymbol{\Sigma}^*_{T,\tau}$ is most important. In our sparse portfolio construction, we propose to apply (9). An alternative approximation in the high-frequency data context is by

$$\tau^{-1}\boldsymbol{\Sigma}^*_{T,\tau} \approx \delta^{-1}\int_{T-\delta}^{T}\boldsymbol{\Sigma}_t dt = \delta^{-1}\overline{\boldsymbol{\Sigma}}_{N,\delta} \tag{13}$$

for some δ; see Fan et al. (2012). By observing that $\overline{\boldsymbol{\Sigma}}_{N,\delta}$ is the most recent ICM at time T, such an approximation is seen as a localization by time. In our approach using (8), the localization is more general, with broad choices. In our empirical study, we demonstrate the promising performance of our approach using the VIX index as $\mathbf{U}_k$.

4 Theory

4.1 Conditions

Note that the impact of the drift $\boldsymbol{\mu}_t$ on the estimation of the ICM is asymptotically negligible when sampling interval lengths that shrink to zero in high-frequency financial data analysis. For simplicity, we assume hereinafter that $\boldsymbol{\mu}_t = 0$. In addition, we assume that the dynamic factor $\mathbf{U}_k$ is univariate. Generalization of our results to the case where $\mathbf{U}_k$ is multivariate is straightforward by using a productive kernel function as in Jiang et al. (2020), with a minor modification of the asymptotic bounds to take into account the dimension d.

Denote the support of $\mathbf{U}_k$ by Ω. For any matrix $\mathbf{A} = (a_{ij})_{p\times q}$, we denote the elementwise matrix max norm by $\|\mathbf{A}\|_\infty = \max_{1\le i\le p, 1\le j\le q} |a_{ij}|$, the Frobenius norm by $\|\mathbf{A}\|_F$, and the vector l_∞ norm induced matrix norm by $\|\mathbf{A}\|_L := \max_{1\le i\le p}\sum_{j=1}^{a}|a_{ij}|$. For any vector $\mathbf{b} = (b_1, \ldots, b_p)'$, we denote the vector l_1 norm by $\|\mathbf{b}\|_1 := \sum_{i=1}^{p}|b_i|$ and the l_0 norm by $\|\mathbf{b}\|_0 := \sum_{i=1}^{p} I\{b_i \neq 0\}$. Here, $I\{\cdot\}$ is the indicator function. We shall use $\lfloor\cdot\rfloor$ to denote the smallest integer function.

Before proceeding to the main theoretical results, we introduce some regularity conditions:

(C1) Denote the hypothetical conditional covariance by $\mathbf{\Sigma}_t(\mathbf{u}) = [\mathbf{\Sigma}_t|\mathbf{U}_{\lfloor t/\delta\rfloor} = \mathbf{u}]$, and its (i,j)th element by $\sigma_{X,t}^{(i,j)}(\mathbf{u})$. There exists a constant M such that $\sup_{\mathbf{u}\in\Omega}\|\mathbf{\Sigma}_t(\mathbf{u})\|_\infty \le M < \infty$. For all i, j with $1 \le i, j \le p$, let $\tilde{X}_s = X_{i,s},\ X_{i,s} + X_{j,s}$, or $X_{i,s} - X_{j,s}$, and let $\tilde{\sigma}_{X,t}^2(\mathbf{u}) = \sigma_{X,t}^{(i,i)}(\mathbf{u}),\ \sigma_{X,t}^{(i,i)}(\mathbf{u}) + 2\sigma_{X,t}^{(i,j)}(\mathbf{u}) + \sigma_{X,t}^{(j,j)}(\mathbf{u})$, or $\sigma_{X,t}^{(i,i)}(\mathbf{u}) - 2\sigma_{X,t}^{(i,j)}(\mathbf{u}) + \sigma_{X,t}^{(j,j)}(\mathbf{u})$, respectively. For all ν, s with $0 < \nu \le s \le \delta$, let $\Delta_\nu\tilde{X}_s := \tilde{X}_s - \tilde{X}_{s-\nu}$. There exist constants $C > 0$ and $\theta_\nu > 0$ such that if $\sqrt{\nu}\theta_\nu$ is sufficiently small we have $E\{\exp\{\theta_\nu\nu^{-1/2}[(\Delta_\nu\tilde{X}_s)^2 - \int_{s-\nu}^{s}\tilde{\sigma}_{X,t}^2(\mathbf{u})dt]\}|\tilde{\mathcal{F}}_{s-\nu}, \mathbf{U}_{\lfloor s/\delta\rfloor} = \mathbf{u}\} \le 1 + C\nu\theta_\nu^2$, where $\tilde{\mathcal{F}}_t$ is the filtration of σ-algebras generated by the path $\tilde{X}_{[0,t]}$.

(C2) For all i, t in the additive error model (5), the measurement errors $\epsilon_{i,t}$'s are independent of the $X_{i,t}$ process and the $\mathbf{U}_k$'s, and are independently distributed with $E\epsilon_{i,t} = 0$. Denote $\epsilon_{i,t}^2 - E\epsilon_{i,t}^2$ by $\bar{\eta}_{i,t}$. There exist constants σ_ϵ^2 and σ_η^2 such that $E\epsilon_{i,t}^2 \le \sigma_\epsilon^2$ and $E\bar{\eta}_{i,t}^2 \le \sigma_\eta^2$. In addition, there exists a positive constant L such that for all positive integers k, $E|\epsilon_{i,t}^k| \le \sigma_\epsilon^2 L^{k-2}k!/2$ and $E|\bar{\eta}_{i,t}^k| \le \sigma_\eta^2 L^{k-2}k!/2$.

(C3) The observation times are independent of the $\mathbf{X}_t$ process. For any pair of elements $X_{i,t}, X_{j,t}$ in $\mathbf{X}_t$, the set of refresh times $v_{k,1}, \ldots, v_{k,n_{ij,k}}$ satisfies: $\sup_{1\le l\le n_{ij,k}} n_{ij,k}(v_{k,l} - v_{k,l-1}) \le C_\delta$ for some constant C_δ. In addition, we assume that there exist a constant $C_n > 1$ such that $C_n^{-1}n \le n_{ij,k} \le C_n n$ for i, j with $1 \le i \le j \le p$. In addition, we assume that the scale parameter K in (6) satisfies $K \asymp O(n^{2/3})$, that is, there exist constants a_K, b_K with $0 < a_K < b_K < \infty$ such that $a_K n^{2/3} \le K \le b_K n^{2/3}$.

(C4) $\left(\frac{\log p}{N}\right)^{2/5}(\log N)(\log\log N) \to 0$ and $\frac{\log p}{n^{1/3}} \to 0$.

(C5) $\{\mathbf{U}_1, \mathbf{U}_2, \ldots\}$ is stationary, and $\{(\overline{\mathbf{\Sigma}}_{k,\delta}, \mathbf{U}_k); k = 1, 2, \ldots\}$ is strong mixing such that

$$\alpha(t) := \sup_{k\ge 1}\ \sup_{A\in\mathcal{A}_k, B\in\mathcal{B}_{k+t}} |P(A\cap B) - P(A)P(B)| \le \exp\{-2sn\}$$

for some constant $s > 0$. Here, $\mathcal{A}_k := \sigma(\overline{\mathbf{\Sigma}}_{i,\delta}, \mathbf{U}_i; i \le k)$ is the σ-field generated by $\{\overline{\mathbf{\Sigma}}_{i,\delta}, \mathbf{U}_i; i \le k\}$, and $\mathcal{B}_{k+t} := \sigma(\overline{\mathbf{\Sigma}}_{j,\delta}, \mathbf{U}_j; j \ge k+t)$ is the σ-field generated by $\{\overline{\mathbf{\Sigma}}_{j,\delta}, \mathbf{U}_j; j \ge k+t\}$. Denote the marginal probability density function of $\mathbf{U}_i$ by $f(\mathbf{u}), \mathbf{u} \in$

Ω. We assume that the support Ω is compact and that $f(\mathbf{u})$ is bounded away from zero and infinity on Ω. In addition, we assume that $f(\mathbf{u})$ has bounded continuous first and second derivatives on Ω.

(C6) The entries of $\boldsymbol{\Sigma}_\delta(\mathbf{u})$ and their first and second derivatives are uniformly bounded and continuous on $\mathbf{u} \in \Omega$. There exists a positive constant η such that $\eta^{-1} \leq \inf_{\mathbf{u}\in\Omega} \lambda_1(\boldsymbol{\Sigma}_\delta(\mathbf{u})) \leq \sup_{\mathbf{u}\in\Omega} \lambda_p(\boldsymbol{\Sigma}_\delta(\mathbf{u})) \leq \eta$, where $\lambda_1(\boldsymbol{\Sigma}_\delta(\mathbf{u}))$ and $\lambda_p(\boldsymbol{\Sigma}_\delta(\mathbf{u}))$ are the smallest and largest eigenvalues of $\boldsymbol{\Sigma}_\delta(\mathbf{u})$.

(C7) The kernel function is symmetric in that $K(x) = K(-x)$, and there exist constants K_1, K_2 such that $\sup_{x\in R} |K(x)| < K_1$ and $\sup_{x\in R} |K'(x)| < K_2$.

These conditions are standard for studying kernel smoothing and high-frequency data in a high-dimensional setting (Fan et al., 2012; Chen and Leng, 2016). Note that $1 + C\nu\theta_\nu^2 \leq \exp\{C\nu\theta_\nu^2\}$. Condition (C1) thus imposes a sub-Gaussian structure on the process $\Delta_\nu \tilde{X}_s$. Condition (C2) holds if the noise is normally distributed. The introduction of a same order n for the sample sizes of the covariates in Condition (C3) could greatly simplify our notation. Condition (C4) characterizes the relationships between N, n, and p for establishing estimation consistency. Condition (C5) allows the underlying factor sequence $\{(\overline{\boldsymbol{\Sigma}}_{k,\delta}, \mathbf{U}_k); k = 1, 2, \ldots\}$ to be weakly dependent. In particular, (C5) is true if $\{(\overline{\boldsymbol{\Sigma}}_{k,\delta}, \mathbf{U}_k); k = 1, 2, \ldots\}$ is locally dependent (Bradley, 2005). Conditions (C6) and (C7) are regular conditions in the kernel-smoothing literature.

4.2 Theoretical results for the estimation of the ICM function

Next, we establish uniform concentration inequalities for the estimation of the ICM function. Let $\overline{\sigma}_{k,\delta}^{(s,t)}$, $\widetilde{\sigma}_{k,\delta}^{(s,t)}$, $\sigma_{\delta}^{(s,t)}(\mathbf{u})$, and $\widehat{\sigma}_{\delta}^{(s,t)}(\mathbf{u})$ be the (s,t)-th elements of $\overline{\boldsymbol{\Sigma}}_{k,\delta}$, $\widetilde{\boldsymbol{\Sigma}}_{k,\delta}$, $\boldsymbol{\Sigma}_\delta(\mathbf{u})$, and $\widehat{\boldsymbol{\Sigma}}_\delta(\mathbf{u})$, respectively.

The following proposition provides a concentration inequality for the two-scale estimator $\overline{\boldsymbol{\Sigma}}_{k,\delta}$.

Proposition 1. *Under Conditions (C1)–(C5), there exist positive constants C_0, C_1, C_2 such that if N is sufficiently large, we have that for any ϵ with $0 \leq \epsilon \leq C_0$ and all s, t with $1 \leq s, t \leq p$,*

$$P\left(\max_{k=1,\ldots,N} \left|\widetilde{\sigma}_{k,\delta}^{(s,t)}(\mathbf{U}_k) - \overline{\sigma}_{k,\delta}^{(s,t)}(\mathbf{U}_k)\right| \geq \epsilon\right) \leq C_1 N \exp\left\{-C_2 n^{1/3}\epsilon^2\right\}.$$

The next proposition establishes a uniform concentration for local smoothing based on $\overline{\boldsymbol{\Sigma}}_{k,\delta}$.

Proposition 2. *Suppose* $\epsilon_N(\log N)(\log\log N) \to 0$, $Nh\epsilon_N^2 \to \infty$, *and* $\epsilon_N^2 > C_h h^4$ *for some sufficiently large constant* C_h. *Under Conditions (C6) and (C7), there exist positive constants* C_3, C_4, *and* C_5, C_6 *such that*

$$P\Big(\sup_{\mathbf{u}\in\Omega}\Big|\frac{1}{N}\sum_{k=1}^{N}K_h(\mathbf{U}_k-\mathbf{u})-f(\mathbf{u})\Big|\geq\epsilon_N\Big)\leq C_3h^{-4}\exp\left\{-C_4Nh\epsilon_N^2\right\} \tag{14}$$

and for all s,t *with* $1\leq s,t\leq p$,

$$P\Big(\sup_{\mathbf{u}\in\Omega}\Big|\frac{1}{N}\sum_{k=1}^{N}K_h(\mathbf{U}_k-\mathbf{u})\overline{\sigma}_{k,\delta}^{(s,t)}(\mathbf{U}_k)-\sigma_\delta^{(s,t)}(\mathbf{u})f(\mathbf{u})\Big|\geq\epsilon_N\Big)\leq C_5h^{-4}\exp\left\{-C_6Nh\epsilon_N^2\right\}. \tag{15}$$

Combining Propositions 1 and 2, we obtain the following uniform concentration inequality for $\widehat{\mathbf{\Sigma}}_\delta(\mathbf{u})$.

Theorem 1. *Suppose* $\epsilon_N(\log N)(\log\log N) \to 0$, $Nh\epsilon_N^2 \to \infty$, *and* $\epsilon_N^2 > C_h h^4$ *for some sufficiently large constant* $C_h > 0$. *Under Conditions (C1)–(C7), there exist positive constants* B_1, B_2, B_3, *and* B_4 *such that if* N *is sufficiently large, for all* s,t *with* $1\leq s,t\leq p$,

$$\begin{aligned}&P\left(\sup_{\mathbf{u}\in\Omega}\left|\widehat{\sigma}_\delta^{(s,t)}(\mathbf{u})-\sigma_\delta^{(s,t)}(\mathbf{u})\right|\geq\epsilon_N\right)\\ \leq\;& B_1N\exp\left\{-B_2n^{1/3}\epsilon_N^2\right\}+B_3h^{-4}\exp\left\{-B_4Nh\epsilon_N^2\right\}.\end{aligned} \tag{16}$$

From the proof of this theorem, we know that the first term on the right-hand side of (16) is mainly introduced by the two-scale estimation scheme, while the second term is introduced by local smoothing. The two terms are connected to each other via a common ϵ_N, which depends on h and N. If n is large enough so that $n^{1/3} \succ Nh + \epsilon_N^{-2}\log N$, the upper bound on the two-scale estimator will be dominated by the local smoothing error.

If we follow the traditional local smoothing scheme and set the bandwidth h so that $h \asymp \left(\frac{\log p}{N}\right)^{1/5}$, we have the following corollary.

Corollary 1. *Assume that the bandwidth* h *satisfies* $h \asymp \left(\frac{\log p}{N}\right)^{1/5}$. *Under conditions (C1)–(C7), there exist positive constants* B_5, B_6, B_7 *such that for any sufficiently large constant* $B_0 > \sqrt{2B_6^{-1}}$,

$$\begin{aligned}&P\left(\sup_{\mathbf{u}\in\Omega}\left\|\widehat{\mathbf{\Sigma}}_\delta(\mathbf{u})-\mathbf{\Sigma}_\delta(\mathbf{u})\right\|_\infty\geq B_0\left(\frac{\log p}{N}\right)^{2/5}\right)\\ \leq\;& B_5p^2N\exp\left\{-B_0^2B_6n^{1/3}\left(\frac{\log p}{N}\right)^{4/5}\right\}+B_7\left(\frac{N}{\log p}\right)^{4/5}p^{2-B_0^2B_6}.\end{aligned} \tag{17}$$

For a sufficiently large B_0, the first term on the right-hand side of (17) tends to zero if $(\log(pN))n^{-1/3}\left(\frac{N}{\log p}\right)^{4/5} \to 0$. While it is not surprising to see that the first term would vanish when n grows fast enough, it is interesting to see that the powers of N and n that ensure convergence are different in this case. The second term on the right-hand side of (17), which tends to zero at a geometric rate of p, is an error term introduced by local smoothing among all N observations.

On the other hand, to balance the two parts of right-hand side of (16) while controlling the estimation error ϵ_N to be as small as possible, we should consider letting h depend on both n and N. Note that ϵ_N would be as least $O(h^2)$. If $h \asymp \left(\frac{\log(pN)}{N}\right)^{1/5} \asymp \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/4}$, we have $n^{1/3}\epsilon_N^2 \asymp Nh\epsilon_N^2 \asymp \log(pN)$. The two terms on the right-hand side of (16) would then both vanish geometrically in p if the constants B_2 and B_4 are sufficiently large. Specifically, we have the following corollary:

Corollary 2. *Let* $h \asymp \left\{\left(\frac{\log(pN)}{N}\right)^{1/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/4}\right\}$, *and assume that* $h^2(\log N)(\log\log N) \to 0$. *Under the assumptions of Theorem 1, for any constant* $B_8 > 0$ *there exists a sufficiently large constant* $B_9 > 0$ *which is independent of* n, N, *and* p *and such that*

$$P\left(\sup_{\mathbf{u}\in\Omega} \|\widehat{\mathbf{\Sigma}}_\delta(\mathbf{u}) - \mathbf{\Sigma}_\delta(\mathbf{u})\|_\infty \geq B_9 \left\{\left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2}\right\}\right) \leq p^{-B_8}. \qquad (18)$$

4.3 Theoretical results on portfolio estimation

Before we introduce the theoretical results on portfolio estimation, we study the theoretical properties of the solution for the following generic penalized quadratic loss: For a given $\mathbf{u} \in \Omega$, define

$$\widehat{\mathbf{b}}(\mathbf{u}) = \text{argmin}_{\mathbf{b}\in\mathbb{R}^p} \mathbf{b}'\widehat{\mathbf{A}}(\mathbf{u})\mathbf{b}/2 - \mathbf{b}'\mathbf{1} + \lambda_{\mathbf{u}}\|\mathbf{b}\|_1, \qquad (19)$$

where $\lambda_{\mathbf{u}}$ is the tuning parameter and $\widehat{\mathbf{A}}(\mathbf{u})$ is a consistent estimator of the $p \times p$ parameter matrix $\mathbf{A}(\mathbf{u})$ for a given $\mathbf{u}$. Let

$$\mathbf{b}^*(\mathbf{u}) = \text{argmin}_{\mathbf{b}\in\mathbb{R}^p} \mathbf{b}'\mathbf{A}(\mathbf{u})\mathbf{b}/2 - \mathbf{b}'\mathbf{1} + \lambda_{\mathbf{u}}\|\mathbf{b}\|_1,$$

and let $\mathcal{S}_{\mathbf{u}} = \{i : b_i^*(\mathbf{u}) \neq 0\}$ be the support of $\mathbf{b}^*(\mathbf{u})$. When there is no ambiguity, we shall use $\mathcal{S}$ instead of $\mathcal{S}_{\mathbf{u}}$. For example, we shall use $\mathbf{b}_{\mathcal{S}}(\mathbf{u})$ instead of $\mathbf{b}_{\mathcal{S}_{\mathbf{u}}}(\mathbf{u})$ to denote the subset of $\mathbf{b}(\mathbf{u})$ whose elements are nonzero. In addition, we shall use $\mathcal{S}^c$ to denote the complement of $\mathcal{S}$.

Proposition 3. *Denote* $\|[\widehat{\mathbf{A}}(\mathbf{u}) - \mathbf{A}(\mathbf{u})]\mathbf{b}^*(\mathbf{u})\|_\infty$ *by* $\epsilon_{\mathbf{u}}$. *Suppose* $\sup_{\mathbf{u}\in\Omega} \|\mathbf{A}_{\mathcal{S}^c,\mathcal{S}}(\mathbf{u})\mathbf{A}_{\mathcal{S},\mathcal{S}}(\mathbf{u})^{-1}\|_L < 1 - \kappa$ *and* $\sup_{\mathbf{u}\in\Omega} \|\mathbf{b}^*(\mathbf{u})\|_0 \|\mathbf{A}_{\mathcal{S},\mathcal{S}}(\mathbf{u})^{-1}\|_L \|\widehat{\mathbf{A}}(\mathbf{u}) - \mathbf{A}(\mathbf{u})\|_\infty \leq 1 - \kappa$ *for some constant* κ *with* $0 < \kappa < 1$. *For* $\lambda_{\mathbf{u}} > \epsilon_{\mathbf{u}}$, *we have the following:*

(i) $\widehat{\mathbf{b}}_{\mathcal{S}^c}(\mathbf{u}) = \mathbf{0}$ *for any* $\mathbf{u} \in \Omega$*;*

(ii) $\sup_{\mathbf{u}\in\Omega} \|\widehat{\mathbf{b}}(\mathbf{u}) - \mathbf{b}^*(\mathbf{u})\|_\infty \leq \kappa^{-1} \sup_{\mathbf{u}\in\Omega} \epsilon_{\mathbf{u}} \|\mathbf{A}_{\mathcal{S},\mathcal{S}}(\mathbf{u})^{-1}\|_L$.

Proposition 3 provides sufficient conditions for establishing consistency results for the generic estimator defined in (19), and it captures the dependence of the estimation errors on the structure of $\mathbf{A}(\mathbf{u})$ and the estimation accuracy of $\widehat{\mathbf{A}}(\mathbf{u})$. From Proposition 3, we immediately have the following proposition.

Proposition 4. *Assume that* $\sup_{\mathbf{u}\in\Omega} \|\mathbf{A}_{\mathcal{S}^c,\mathcal{S}}(\mathbf{u})\mathbf{A}_{\mathcal{S},\mathcal{S}}(\mathbf{u})^{-1}\|_L < 1-\kappa$ *for some constant* κ *with* $0<\kappa<1$*, and that for any positive constant* B*, we have that, with probability larger than* $1 - O(p^{-B})$*,* $\sup_{\mathbf{u}\in\Omega} \|\widehat{A}(\mathbf{u}) - A(\mathbf{u})\|_\infty < C_1 \left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\}$ *for some sufficiently large constant* C_1*. In addition, assume that* $\left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}\in\Omega} \|\mathbf{b}^*(\mathbf{u})\|_0 \|\mathbf{A}_{\mathcal{S},\mathcal{S}}(\mathbf{u})^{-1}\|_L \to 0$*. For a given* $\lambda_{\mathbf{u}}$ *such that* $\lambda_{\mathbf{u}} \geq C_2 \left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}\in\Omega} \|\mathbf{b}^*(\mathbf{u})\|_1$ *for some sufficiently large positive constant* C_2*, we have that, with probability larger than* $1 - O(p^{-B})$*, the following hold:*

(i) $\widehat{\mathbf{b}}_{\mathcal{S}^c}(\mathbf{u}) = \mathbf{0}$ *for any* $\mathbf{u} \in \Omega$*;*

(ii) $\sup_{\mathbf{u}\in\Omega} \|\widehat{\mathbf{b}}(\mathbf{u}) - \mathbf{b}^*(\mathbf{u})\|_\infty \leq C \left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}\in\Omega} \|\mathbf{b}^*(\mathbf{u})\|_1 \|\mathbf{A}_{\mathcal{S},\mathcal{S}}(\mathbf{u})^{-1}\|_L$ *for some sufficiently large constant* $C > 0$.

Now we establish theoretical results for portfolio estimation. Denote the optimal sparse portfolio for a given λ by $\mathbf{w}^\lambda_{T,\tau} = (w^\lambda_{1,\tau}, \ldots, w^\lambda_{p,\tau})'$. With some abuse of notation, let $\mathcal{S} = \{i : w^\lambda_{i,\tau} \neq 0\}$ be the support of $\mathbf{w}^\lambda_{T,\tau}$ and let s_λ be the cardinality of $\mathcal{S}$. From Proposition 4 and Corollary 2, we obtain the following theorem.

Theorem 2. *Assume that* $\sup_{\mathbf{u}_i\in\Omega, N\leq i\leq N+N^*} \|(\boldsymbol{\Sigma}^*_{T,\tau})_{\mathcal{S}^c,\mathcal{S}} (\boldsymbol{\Sigma}^*_{T,\tau})^{-1}_{\mathcal{S},\mathcal{S}}\|_L < 1-\kappa$ *for some constant* κ *with* $0<\kappa<1$*, and that* $\left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}_i\in\Omega, N\leq i\leq N+N^*} \|\boldsymbol{\beta}^\lambda_{T,\tau}\|_0 \|(\boldsymbol{\Sigma}^*_{T,\tau})^{-1}_{\mathcal{S},\mathcal{S}}\|_L \to 0$*. For any* $B>0$*, by choosing* $\lambda_{\mathbf{u}} > C_2 \left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}_i\in\Omega, N\leq i\leq N+N^*} \|\boldsymbol{\beta}^\lambda_{T,\tau}\|_1$ *for some sufficiently large positive constant* C_2*, we have that, with probability larger than* $1 - O(p^{-B})$*, the following hold:*

(i) $(\widehat{\boldsymbol{\beta}}_{T,\tau})_{\mathcal{S}^c} = \mathbf{0}$ *for any* $\mathbf{u}_{N+1}, \ldots, \mathbf{u}_{N+N^*} \in \Omega$*;*

(ii) There exists a sufficiently large constant $C > 0$ *such that* $\sup_{\mathbf{u}_i\in\Omega, N\leq i\leq N+N^*} \|\widehat{\boldsymbol{\beta}}^\lambda_{T,\tau} - \boldsymbol{\beta}^\lambda_{T,\tau}\|_\infty \leq C \left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}_i\in\Omega, N\leq i\leq N+N^*} \|\boldsymbol{\beta}^\lambda_{T,\tau}\|_1 \|(\boldsymbol{\Sigma}^*_{T,\tau})^{-1}_{\mathcal{S},\mathcal{S}}\|_L$.

The next theorem provides consistency results for the weight estimator $\widehat{\mathbf{w}}^{\lambda}_{T,\tau}$ and risk estimator $\widehat{R}^{\lambda}_{T,\tau}$.

Theorem 3. *Under the assumptions of Theorem 2, and assuming that there exist constants* $D_1, D_2 > 0$ *such that* $D_1 \leq min(|\boldsymbol{\beta}^{\lambda}_{T,\tau}|) \leq max(|\boldsymbol{\beta}^{\lambda}_{T,\tau}|) \leq D_2$*, where* $min(|\boldsymbol{\beta}^{\lambda}_{T,\tau}|)$ *and* $max(|\boldsymbol{\beta}^{\lambda}_{T,\tau}|)$ *are the smallest and largest absolute values of elements of* $\boldsymbol{\beta}^{\lambda}_{T,\tau}$*, respectively. We have that, with probability larger than* $1 - O(p^{-B})$ *for a given positive constant* B*, the following hold:*

(i) $(\widehat{\mathbf{w}}_{T,\tau})_{\mathcal{S}^c} = \mathbf{0}$ *for any* $\mathbf{u}_{N+1}, \ldots, \mathbf{u}_{N+N^*} \in \Omega$*;*

(ii) There exists a sufficiently large constant $C > 0$ *such that* $\sup_{\mathbf{u}_i \in \Omega, N \leq i \leq N+N^*} \|\widehat{\mathbf{w}}^{\lambda}_{T,\tau} - \mathbf{w}^{\lambda}_{T,\tau}\|_\infty \leq C \left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}_i \in \Omega, N \leq i \leq N+N^*} \|(\boldsymbol{\Sigma}^*_{T,\tau})^{-1}_{\mathcal{S},\mathcal{S}}\|_L$.

(iii) There exists a large enough constant $C_1 > 0$ *such that* $\sup_{\mathbf{u}_i \in \Omega, N \leq i \leq N+N^*} |\widehat{R}^{\lambda}_{T,\tau} - R^{\lambda}_{T,\tau}| \leq s^2_\lambda C_1 \left\{ \left(\frac{\log(pN)}{N}\right)^{2/5} + \left(\frac{\log(pN)}{n^{1/3}}\right)^{1/2} \right\} \sup_{\mathbf{u}_i \in \Omega, N \leq i \leq N+N^*} \|(\boldsymbol{\Sigma}^*_{T,\tau})^{-1}_{\mathcal{S},\mathcal{S}}\|_L$.

5 Simulations

We conducted extensive simulations to show the advantages of our dynamic covariance estimation approach, and the merits of its use in constructing high-dimensional sparse portfolios. The tuning parameter h in (7) for estimating $\widehat{\boldsymbol{\Sigma}}_\delta(\mathbf{u})$ was chosen using the cross validation method. Specifically, given $\{\widetilde{\boldsymbol{\Sigma}}_{k,\delta}, \mathbf{U}_k\}_{k=1}^N$, h was chosen by minimizing

$$\ell(h) = \sum_{k=1}^{N} \left\| \left\{ \sum_{l \neq k}^{N} \frac{K_h(\mathbf{U}_l - \mathbf{U}_k)}{\sum_{m \neq k}^{N} K_h(\mathbf{U}_m - \mathbf{U}_k)} \widetilde{\boldsymbol{\Sigma}}_{l,\delta} \right\} - \widetilde{\boldsymbol{\Sigma}}_{k,\delta} \right\|_F .$$

In our simulations, we generated data with the following model:

$$\mathbf{Y}_t = \mathbf{X}_t + \boldsymbol{\epsilon}_t, \quad d\mathbf{X}_t = \boldsymbol{\Sigma}_t^{1/2} d\mathbf{B}_t, \tag{20}$$

where $\mathbf{B}_t$ is a p-dimensional standard Brownian motion and $\boldsymbol{\epsilon}_t$ is the measurement error. Similar to the settings in Aït-Sahalia et al. (2010) and Liu and Tang (2014), we generated independent and identically distributed $\boldsymbol{\epsilon}_t$'s from $N(0, 0.0005^2\mathbf{I}_p)$. In our simulations, we set $\mathbf{U}_k$ to be 1-dimension; thus we write it as U_k hereinafter.

We considered $\delta = 1/252$—corresponding to one business day—in the simulations. Together with $\mathbf{X}_t$ from (20) and the associated $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ defined by (2), we generated the U_k's from the CIR process (Cox et al., 1985):

$$dU_t = \kappa(\theta - U_t)dt + \eta\sqrt{U_t}dW_t,$$

where W_t is a Brownian motion independent of $\mathbf{B}_t$. By setting $(\kappa, \theta, \eta) = (6, 0.5^2, 1.2)$, we can ensure U_k's to be always positive. This mimics the fact that in our empirical studies we advocate using variance-related quantities such as the VIX index and its contingencies. The initial value of the $\{U_t\}$ process was generated from the stationary marginal distribution of the CIR process.

When generating $\{\overline{\boldsymbol{\Sigma}}_{k,\delta}, U_k\}$, we incorporated dependence between $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ and U_k. Correspondingly, we considered two models for $\boldsymbol{\Sigma}_t$.

- **Model 1:**

  We assumed that $\boldsymbol{\Sigma}_t$ is time invariant during each short time period $((k-1)\delta, k\delta]$, so that

$$\frac{1}{\delta}\overline{\boldsymbol{\Sigma}}_{k,\delta} = 2/3\boldsymbol{\beta}(U_k)\mathbf{R}(U_k)\boldsymbol{\beta}'(U_k) + 1/3\mathbf{F}(U_k)\mathbf{F}'(U_k),$$

  where $\boldsymbol{\beta}(U_k)$ is a diagonal matrix and $\mathbf{F}(U_k)$ is a vector; both are of appropriate sizes. We set $\boldsymbol{\beta}(U_k)$ to $\text{diag}^{1/2}(0.5e^{-U_k}\boldsymbol{\Gamma}_1', 2(1-U_k)^2\boldsymbol{\Gamma}_2')'$, and $\text{diag}\{\mathbf{F}(U_k)\}$ to $\text{diag}^{1/2}(6U_k^3\boldsymbol{\Gamma}_3', 0.5e^{-U_k}\boldsymbol{\Gamma}_4')$ with $\boldsymbol{\Gamma}_1, ..., \boldsymbol{\Gamma}_4$ being vectors of length $p/2$ whose components were independently generated from the uniform distribution U(0.2, 0.5). The (i, j)th element of $\mathbf{R}(U_k)$ was set to $\{\min(U_k, 0.8)\}^{|i-j|}$.

- **Model 2:**

  We considered a time-varying $\boldsymbol{\Sigma}_t$ where the spot correlation matrix of the log-returns is time invariant during each short time period $((k-1)\delta, k\delta]$ and the volatility of each asset is time varying. Specifically, for any $t \in ((k-1)\delta, k\delta]$, we let

$$\boldsymbol{\Sigma}_t = \mathbf{L}_t\mathbf{R}(U_k)\mathbf{L}_t, \quad \mathbf{R}(U_k) = \left(\text{diag}\left\{\mathbf{S}(U_k)\right\}\right)^{-1/2}\mathbf{S}(U_k)\left(\text{diag}\left\{\mathbf{S}(U_k)\right\}\right)^{-1/2},$$

  where $\mathbf{S}(U_k)$ was generated in the same way as $\frac{1}{\delta}\overline{\boldsymbol{\Sigma}}_{k,\delta}$ in Model 1 and $\mathbf{L}_t$ is a diagonal matrix whose jth diagonal element L_t^j follows a CIR process:

$$dL_t^j = \kappa_j(\theta_j - L_t^j)dt + \eta_j\sqrt{L_t^j}dW_t^j$$

  for $t \in (0, T+\tau]$, where W_t^j's are mutually independent Brownian motions. For all $j \in \{1, ..., p\}$, we let $(\kappa_j, \theta_j, \eta_j) = (5, 0.3^2, 0.3)$ and set the initial value of L_t^j to 0.3^2.

To generate data for each business day, we first generated n observations. To create an asynchronous scheme, we then used p independent Bernoulli trials with successful probability 0.6 to randomly select subsets of the sample.

We considered the global minimum-variance (GMV) portfolio problem with a norm constraint c for the weight in (10). We estimated $\overline{\boldsymbol{\Sigma}}_{k,\delta}$ for $k = 1, ..., N$ by the two scales covariance estimator (Zhang, 2011) and rebalanced the weights of each portfolio at the end of each day, hence we needed to estimate $\mathbf{\Sigma}^*_{T,\tau}$ in (8) for only $\tau = \delta$. We considered the investment period to be 252×2 business days, representing two years. For comparison, the following estimators for $\mathbf{\Sigma}^*_{T,\delta}$ were considered:

(1) The average of the estimators of the previous N daily ICMs, $\frac{1}{N}\sum_{k=1}^{N}\widetilde{\mathbf{\Sigma}}_{k,\delta}$, which does not use the information contained in the U_k's. This is an alternative estimator for (8) that was used in Fan et al. (2012). We refer to this estimator as Aver;

(2) $\widehat{\mathbf{\Sigma}}_\delta(\widehat{U}_{N+1})$ with $\widehat{U}_{N+1}$ predicted by the AR(1) model constructed based on $U_1, ..., U_N$. We refer to this estimator as KerAR1.

We first studied the performances of our proposed approach on predicting the future ICM. We considered $T = N\delta$. The averages of $\|\widehat{\mathbf{\Sigma}}_\delta(\widehat{U}_{N+1}) - \overline{\mathbf{\Sigma}}_{N+1,\delta}\|_F/|\overline{\mathbf{\Sigma}}_{N+1,\delta}\|_F \times 100$ and $\|\frac{1}{N}\sum_{k=1}^{N}\widetilde{\mathbf{\Sigma}}_{k,\delta} - \overline{\mathbf{\Sigma}}_{N+1,\delta}\|_F/|\overline{\mathbf{\Sigma}}_{N+1,\delta}\|_F \times 100$ over 504 investment days are reported in Table 1 under columns labeled by "FroNormDiff". The averages of $|\widehat{\mathbf{w}}^c_{\text{KerAR1}} - \mathbf{w}^c_{\text{Oracle}}|/p \times 100$ and $|\widehat{\mathbf{w}}^c_{\text{Aver}} - \mathbf{w}^c_{\text{Oracle}}|/p \times 100$ over different $c \in \{1, 1.2, 1.5, 2, 5\}$ and 504 investment days were reported in Table 1 under columns labeled by "WeightDiff". We considered the portfolio size $p = 30$ and 100, the original sample size of high-frequency data in each day $n = 1950$ and 23400, and the moving window size for predicting the future ICM $N = 10, 50, 100, 150, 200$, and 250.

In addition, we also reported the annualized risks (standard deviations) of daily returns of the portfolios constructed based on $\widehat{\mathbf{\Sigma}}_\delta(\widehat{U}_{N+1})$, $\frac{1}{N}\sum_{k=1}^{N}\widetilde{\mathbf{\Sigma}}_{k,\delta}$, and $\overline{\mathbf{\Sigma}}_{N+1,\delta}$ with $p = 30$ and 100 under Models 1 and 2 in Figures 1–4, along with the estimation errors, which are compared by the averages of $\|\frac{1}{N}\sum_{k=1}^{N}\widetilde{\mathbf{\Sigma}}_{k,\delta} - \overline{\mathbf{\Sigma}}_{N+1,\delta}\|_F$ and $\|\widehat{\mathbf{\Sigma}}_\delta(\widehat{U}_{N+1}) - \overline{\mathbf{\Sigma}}_{N+1,\delta}\|_F$. For comparison, we considered different lengths of historical intra-daily data with $N \in \{10, 20, 30, ..., 250\}$.

Some interesting findings can be concluded:

- From Table 1, we can find that for KerAR1, the ratios of prediction errors of the ICM and the difference between the weights constructed based on the oracle ICM—$\overline{\mathbf{\Sigma}}_{N+1,\delta}$ and $\widehat{\mathbf{\Sigma}}_\delta(\widehat{U}_{N+1})$ decrease as N and/or n increase. However, for Aver, the ratios of prediction errors of the ICM decrease as n increases only when $N = 10$, and the difference between the weights constructed based on the oracle ICM—$\overline{\mathbf{\Sigma}}_{N+1,\delta}$ and $\frac{1}{N}\sum_{k=1}^{N}\widetilde{\mathbf{\Sigma}}_{k,\delta}$ increases as N increases.

- The portfolio constructed based on KerAR1 performs much better than that based on Aver in terms of smaller risks. Moreover, Aver may perform worse and worse as N increases, while the performance of KerAR1 is much more stable.

- Comparing the results with different settings for $\mathbf{\Sigma}_t$, we find that the improvements caused by incorporating extra information from the index are more significant with Model 1. This is reasonable, as $\overline{\mathbf{\Sigma}}_{k,\delta}$ is a function of U_k in Model 1. However, for Model 2 the spot covariance matrix $\mathbf{\Sigma}_t$ is actually a sum of a function of U_t and some noises that stems from the randomness of volatilities, which does not depend on U_t.
- Comparing the results with $p = 30$ and $p = 100$, we find that the risks of all the portfolios decrease as p increases, as expected. In addition, the improvement from incorporating the index U_t does not seem to depend on the size of the portfolio, which is a very appealing property.

In conclusion, the simulation results show that extra information from the U_k is helpful for better estimating the ICM. With better covariance estimations, the portfolios constructed based on the dynamic ICM estimators perform better than those constructed by simple averaging.

6 Empirical data analysis

We conducted a real data analysis that constructed GMV portfolios by using high-frequency trading data from the stock market. The intra-day high-frequency data were downloaded from the TAQ database. We considered a pool of the components of the S&P 500 index whose high-frequency observations observations are available on all the trading days from 01/01/2008 to 06/30/2017. We considered three portfolio sizes: $p = 30, 60$, and 100. Beginning with a portfolio of $p = 30$ that randomly selected 30 stocks, we then added randomly selected additional 30 stocks to construct the portfolio of size $p = 60$, and then another 40 for $p = 100$. The period we considered starts from the first trading day of 2009 such that there are 2,139 trading days in total. The portfolios were rebalanced at the end of each day by updating the weights. By using the data in the periods of difference lengths ($N = 10, 20, 30, ..., 250$ days), we analyzed the resulting covariance estimations and the performances of the sparse portfolios. The CBOE volatility index, known as the VIX, was used as the index U_k in the proposed approach. Empirically, we find that the correlation between VIX's and the daily volatilities (the sum of volatilities of all 100 stocks during one day) is 0.8079. The correlation between the VIX's and the Frobenius norms of the daily ICMs is 0.8064.

Similar to that in the simulation study, we carried out comparisons between the portfolios constructed based on Oracle:= $\widetilde{\mathbf{\Sigma}}_{N+1,\delta}$ which is the two scales covariance estimator constructed based on the intra-daily data observed on the $(N+1)$st day, Aver:= $\frac{1}{N}\sum_{k=1}^{N}\widetilde{\mathbf{\Sigma}}_{k,\delta}$, and KerAR1 := $\widehat{\mathbf{\Sigma}}_\delta(\widehat{U}_{N+1})$, where $\widehat{U}_{N+1}$ was predicted by the AR(1) model built on $U_1, U_2, ..., U_N$. The annualized risks of the portfolios are reported in Figures 5 – 7. We have also compared the prediction abilities of Aver and KerAR1. Since the true realizations of the ICMs are unavailable, we computed the Frobenius norms of the differences between the predictors and the estimator $\widetilde{\mathbf{\Sigma}}_{N+1,\delta}$. Those results are also reported in Figures 5–7. The empirical results show the following:

- In general, Oracle performs the best and KerAR1 outperforms Aver almost for all cases in the sense of smaller portfolio risks. However, when the norm constraint for the weight $c = 5$ and the portfolio size $p = 60$ and 100, the performance of Oracle weight is worse than KerAR1 and Aver. A likely reason behind is the less than satisfactory performance of the non-sparse portfolio constructed with many stocks. That is, when c is large, the resulting weights tend to be dense, even the covariance estimations are accurate as the Oracle. In this real data example, since it is evidential that a sparse portfolio is preferred especially with many stocks, the less than satisfactory performance of a dense portfolio is not surprising.

- As N increases, in most cases the risks of the portfolios constructed based on KerAR1 seem to first decrease and then increase. This could be due to some structural changes in the dynamics over a longer period of time. Empirically, we find that it works quite well to use a period of 20 to 60 days' historical intra-daily data to estimate the covariances and to construct portfolios. While for Aver, their performances were generally better with smaller N, but worsened when N increased. This is likely because that using data from more days averaged out the between-periods heterogeneity in the covariances, so that the resulting portfolio lacked the adaptability to the market dynamics. Therefore, for applying approached with averaged covariance estimation, smaller N such as 30 or smaller is recommended as that in Fan et al. (2012) and Cai et al. (2020).

- Comparing the results with different portfolio sizes, we find that as p increases, the risks decrease. In addition, comparing the results with different constraints c, both Aver and KerAR1 seem to achieve minimal risks with $c = 2$. We observe that KerAR1 method showed some less stable performance when N is smaller. This is not surprising, and it is likely due to higher level of variation of this method involving localization when only utilizing fewer days of data. This also demonstrates that for more difficult high dimensional covariance estimations, the trade-off between the bias and variance in the resulting estimations is an important consideration.

- For the Frobenius norms of the differences between the predictors and $\widetilde{\mathbf{\Sigma}}_{N+1,\delta}$, we find that the norms of Aver are larger and increase much faster than those of KerAR1. This in some sense suggests that the information contained in the VIX index is useful for more accurately predicting the covariances.

Overall, we see the effectiveness of our dynamic covariance matrix estimation in this data analysis.

We further compare the performances of the portfolios constructed by using high-frequency data to those of the portfolios using only low-frequency daily data. The following methods using daily data were considered:

- $1/p$: The equal-weight portfolio, which can be viewed as the covariance matrix of returns is estimated by the $p \times p$ identity matrix.
- SP: $\mathbf{\Sigma}^*_{T,\tau}$ in (10) is estimated by the sample covariance matrix of daily returns.
- FF3: $\mathbf{\Sigma}^*_{T,\tau}$ in (10) is estimated based on the Fama–French three-factor model (Fama and French, 1993).
- LS: $\mathbf{\Sigma}^*_{T,\tau}$ in (10) is estimated by an optimal linear shrinkage of the sample covariance matrix of daily returns, which is developed in Ledoit and Wolf (2004).
- NL-SF: $\mathbf{\Sigma}^*_{T,\tau}$ in (10) is estimated by the single-factor preconditioned nonlinear shrinkage estimator, which is also based on daily returns and developed in Ledoit and Wolf (2017).
- DCC: $\mathbf{\Sigma}^*_{T,\tau}$ in (10) is predicted by the Dynamic Conditional Correlation (DCC) model developed in Engle (2002).

For the portfolios based on high-frequency intra-daily data, we use 30 days' historical data to construct the portfolios with the two methods Aver and KerAR1. For the estimators constructed based on daily returns, we follow the study in Ledoit and Wolf (2017) and use 250 days' daily returns. The c in (10) is set to be $1, 1.5$, or 2. Three measures are reported and used in comparing the out-of-sample performances of different portfolios. The results are reported in Table 2.

- AV: the annualized average of out-of-sample returns during the investment.
- SD: the annualized standard deviation of out-of-sample returns during the investment.
- SR: the sharp ratio, AV/SD.

Since we considered the GMV portfolio, we focused on the comparisons of the out-of-sample standard deviations firstly and then compared the sharp ratios of different portfolio strategies. We have found the followings:

- KerAR1 with $c = 2$ achieves the smallest risk among all portfolios with different portfolio sizes and different c. KerAR1 always performs the best in terms of the smallest risks and highest sharp ratios in all cases. This furthers confirms the benefits of our proposed approach from a). using high-frequency data in estimations the covariances, and b). incorporating some informative volatility measures such as the VIX index.

- NL-SF performs the best among the portfolios constructed based on daily returns, however, its performances are worse than the portfolio strategies constructed based on high-frequency data. Though the portfolios constructed using low-frequency data are considered as effective in risk management, more sophisticated investors may attempt to explore opportunities for better risk-reward balanced strategies that take advantages from the opportunities using high-frequency data.

7 Discussion

Thanks to the opportunities from using high-frequency data, we proposed a dynamic estimator for the covariance matrix of the returns of many stocks, based on which we investigated the performance of the sparse portfolio, both in theory and by empirical analysis. We show that our approach has some potential in providing an opportunity for better covariance matrix estimations, and better portfolio performances.

Clearly, to what extend our approach is effective depends on how informative the variable that is given. In our data analysis, the VIX index seems to be a good choice. Since VIX is based on the prices (implied volatilizes) of the options, such a market view of the future volatility level may not always be most effective, especially during the regime changing periods of time such that the investor's expectation and hedging strategy may not catch up soon enough. In light of this consideration, we advocate exploring other possible variables in applications. There are some promising possibilities. For example, besides the VIX, the volatility levels of some common factors that are driving the market are possible candidates, e.g., the predicted future volatilities of the Fama-French's factor models. Depending on the time-horizons of interests and the particular market sections, e.g., IT or industry, of interests when building a portfolio, the term structure of the bond returns may also be considered because not the all stocks are affected in the same way when there is a change in the term structure of the bond returns. These are all of valuable interests in our future investigations.

References

Aït-Sahalia, Y. and Brandt, M. W. (2001). Variable selection for portfolio choice. *The Journal of Finance*, 56(4):1297–1351.

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105:1504–1517.

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

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18:315–416.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Aït-Sahalia, Y. and Xiu, D. (2018). Principal component analysis of high-frequency data. *Journal of the American Statistical Association*, 114(525):287–303.

Back, K. (2010). *Asset Pricing and Portfolio Choice Theory*. Oxford University Press, USA, Oxford.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shepard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162:149–169.

Bickel, P. J. and Levina, E. (2008a). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604.

Bickel, P. J. and Levina, E. (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36:199–227.

Bradley, R. C. (2005). Basic properties of strong mixing conditions: A survey and some open questions. *Probability Surveys*, 2(0):107–144.

Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *The Journal of Finance*, 54(5):1609–1645.

Cai, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annual Review of Statistics and Its Application*, 4(1):423–446.

Cai, T. T., Hu, J., Li, Y., and Zheng, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics*, 214(2):482–494.

Chen, J., Li, D., and Linton, O. (2019). A new semiparametric estimation approach for large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics*, 212(1):155–176.

Chen, Z. and Leng, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association*, 111(515):1196–1207.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385.

Dai, C., Lu, K., and Xiu, D. (2019). Knowing factors or factor loadings, or neither? Evaluating estimators of large covariance matrices with noisy and asynchronous data. *Journal of Econometrics*, 208(1):43–79.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Fama, E. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.

Fan, J., Li, Y., and Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, 107(497):412–428.

Fan, J., Liao, Y., and Liu, H. (2016). An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1):C1–C32.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Jacod, J. and Protter, P. (2012). *Discretization of Processes*. New York: Springer.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

Jagannathan, R. and Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance*, 51(1):3–53.

Jiang, B., Chen, Z., and Leng, C. (2020). Dynamic linear discriminant analysis in high dimensional space. *Bernoulli*, 26(2):1234–1268.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.

Lewellen, J. and Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82(2):289–314.

Liu, C. and Tang, C. Y. (2014). A quasi-maximum likelihood approach for integrated covariance matrix estimation with high frequency data. *Journal of Econometrics*, 180:217–232.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77.

Merton, R. C. (1990). *Continuous-Time Finance*. Cambridge, Mass.

Petkova, R. and Zhang, L. (2005). Is value riskier than growth? *Journal of Financial Economics*, 78(1):187–202.

Tao, M., Wang, Y., Yao, Q., and Zou, J. (2011). Large volatility matrix inference via combing low-frequency and high-frequency approaches. *Journal of the American Statistical Association*, 106:1025–1040.

Tao, M., Wang, Y., and Zhou, H. (2013). Optimal sparse volatility matrix estimation for high dimensional Itô processes with measurement errors. *Annals of Statistics*, 41(4):1816–1864.

Tsay, R. (2013). *Multivariate Time Series Analysis*. John Wiley & Sons.

Tsay, R. and Chen, R. (2019). *Nonlinear Time Series Analysis*. John Wiley & Sons, Hoboken, NJ.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.

Wang, H., Peng, B., Li, D., and Leng, C. (2020). Nonparametric estimation of large covariance matrices with conditional sparsity. *SSRN manuscript*.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159:235–250.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12:1019–1043.

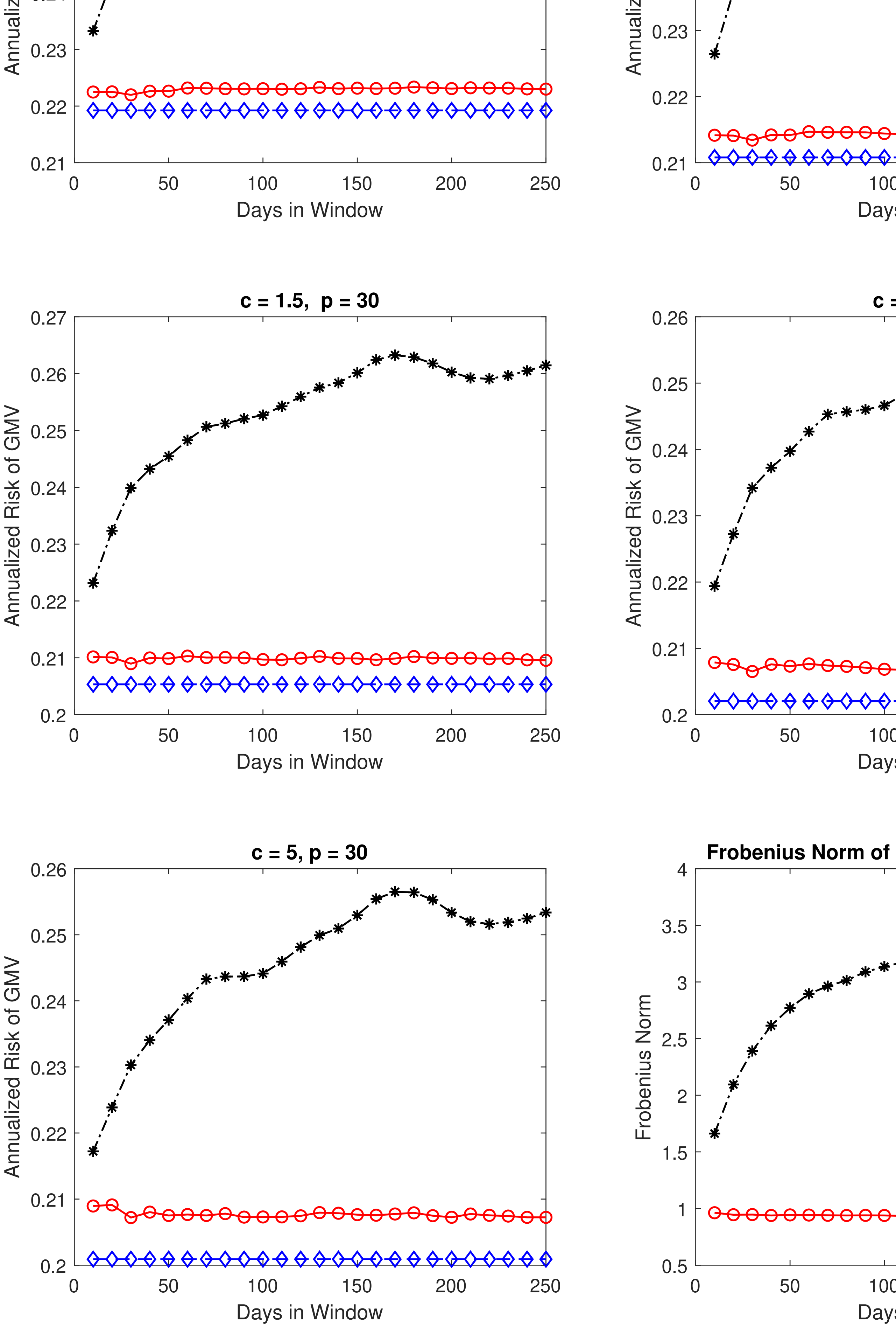
Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160:33–47.

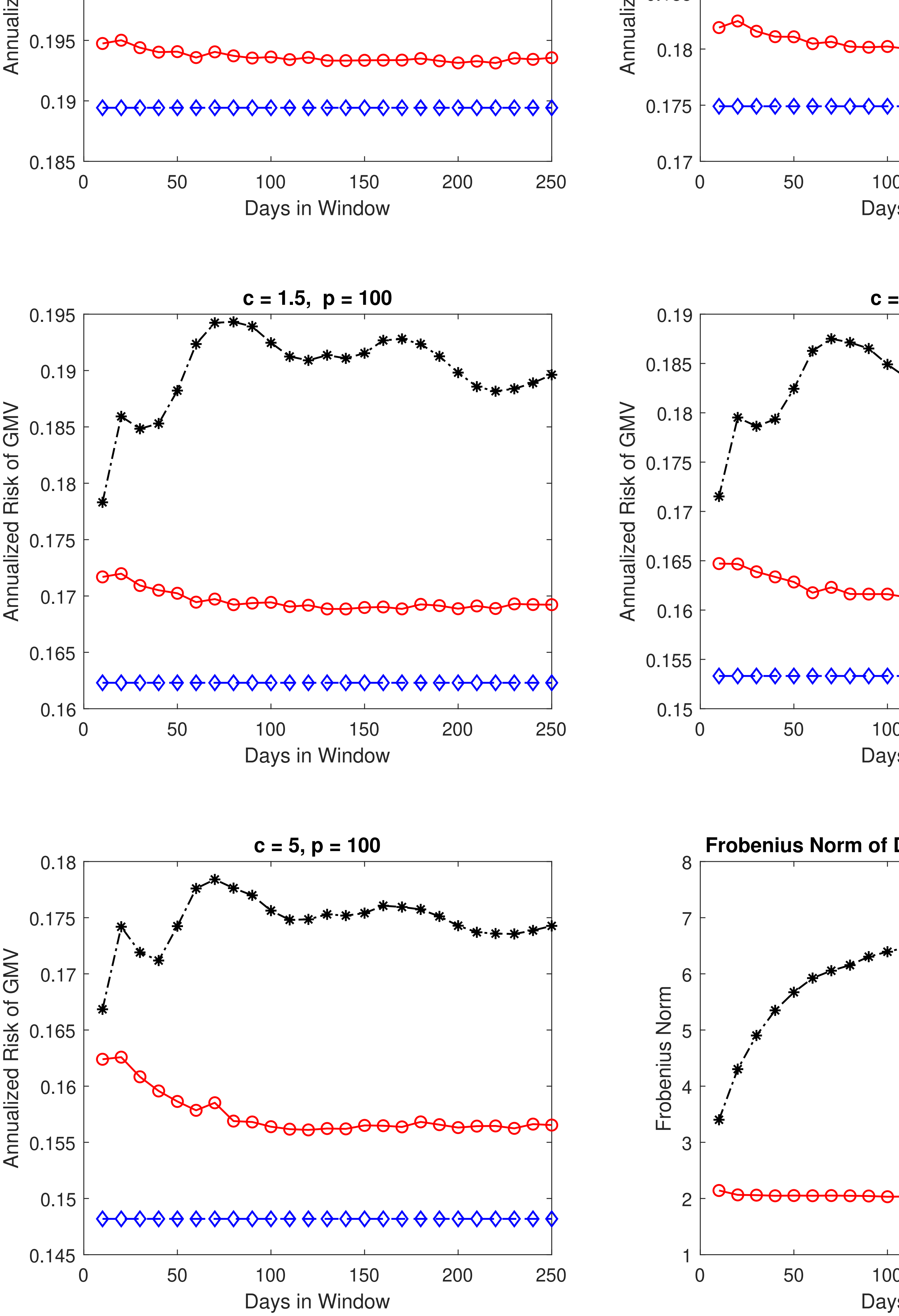
Zhang, L., Mykland, P. A., and Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, 100:1394–1411.

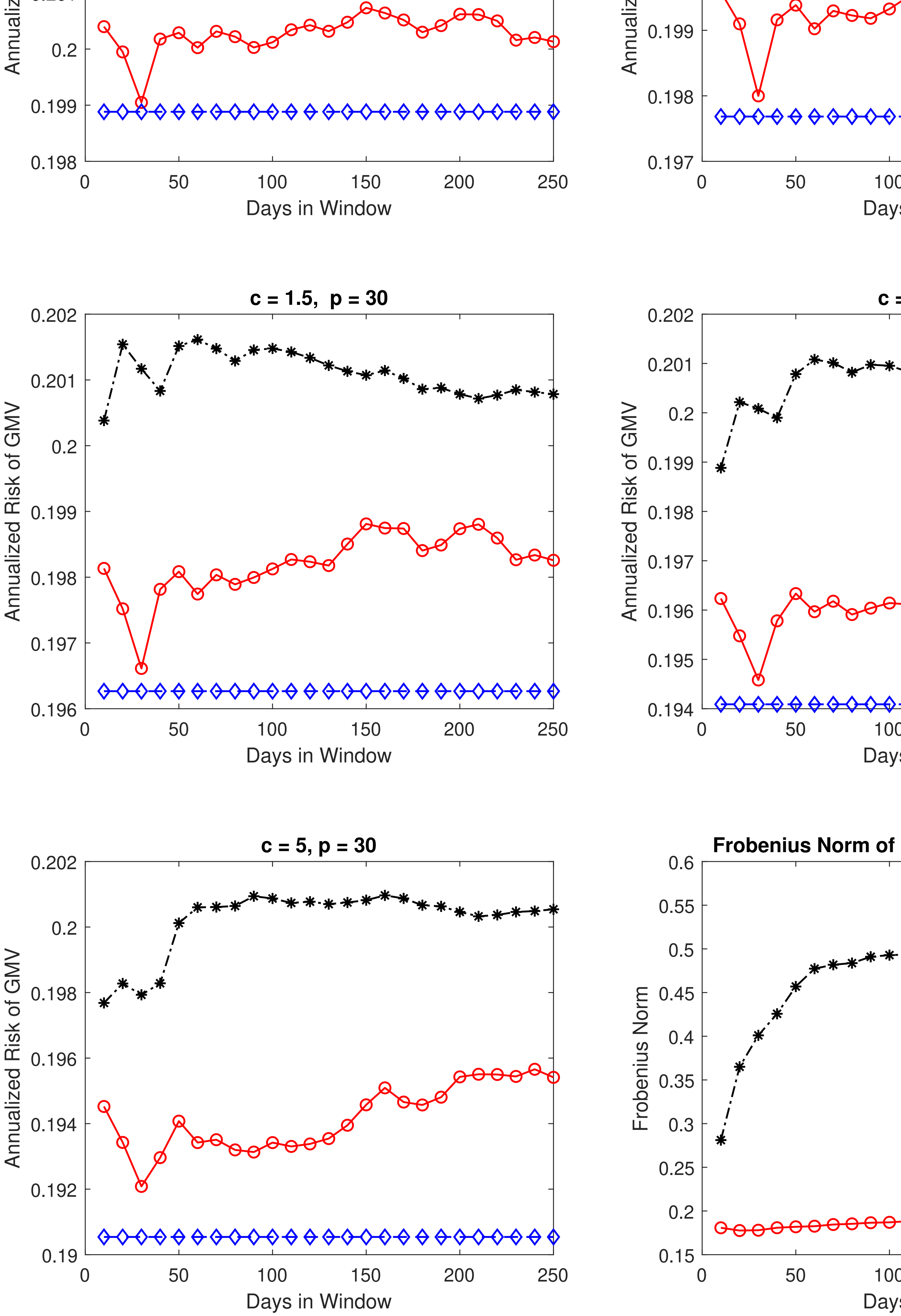
Table 1: Averages of prediction errors (ratios $\times 100$) and differences ($\times 100$) between KerAR1/Aver weight and the weight constructed based oracle estimator.

p	N	$n = 1950$				$n = 23400$			
		FroNormDiff		WeightDiff		FroNormDiff		WeightDiff	
		KerAR1	Aver	KerAR1	Aver	KerAR1	Aver	KerAR1	Aver
					Model 1				
30	10	28.89	41.93	3.72	3.26	22.58	41.47	2.16	2.29
	50	24.62	71.45	2.59	3.38	21.86	75.56	1.87	3.13
	100	23.02	90.98	2.16	3.51	21.51	96.60	1.72	3.36
	150	22.37	104.73	1.99	3.74	21.29	111.64	1.67	3.62
	200	21.66	115.70	1.89	3.85	21.03	123.57	1.61	3.77
	250	21.58	113.75	1.85	3.81	20.97	121.67	1.59	3.77
100	10	28.91	35.65	1.82	1.70	19.83	33.15	1.07	1.03
	50	23.03	54.04	1.35	1.46	18.74	56.38	0.87	1.27
	100	20.97	65.17	1.11	1.43	18.33	68.60	0.78	1.33
	150	20.09	73.74	1.00	1.41	18.04	78.05	0.75	1.35
	200	19.29	79.64	0.92	1.40	17.81	84.45	0.71	1.35
	250	19.17	78.27	0.89	1.38	17.74	83.09	0.70	1.35
					Model 2				
30	10	24.38	26.82	4.67	3.78	14.80	23.70	2.35	2.36
	50	19.90	40.41	3.56	3.59	14.90	39.56	2.38	3.01
	100	18.61	42.32	3.56	3.76	15.42	43.03	2.60	3.25
	150	18.61	46.24	3.59	3.81	15.70	47.23	2.64	3.34
	200	18.03	47.79	3.53	3.79	15.98	48.75	2.68	3.37
	250	17.92	47.66	3.51	3.78	16.09	48.29	2.72	3.40
100	10	28.16	29.61	1.96	1.70	16.29	25.24	1.08	1.00
	50	21.94	42.51	1.56	1.49	15.77	41.64	1.01	1.19
	100	20.19	45.54	1.50	1.51	16.01	46.25	1.05	1.28
	150	19.72	49.98	1.48	1.53	16.07	50.99	1.07	1.31
	200	19.16	52.07	1.46	1.52	16.16	52.98	1.08	1.32
	250	19.02	51.69	1.45	1.51	16.27	52.31	1.09	1.33

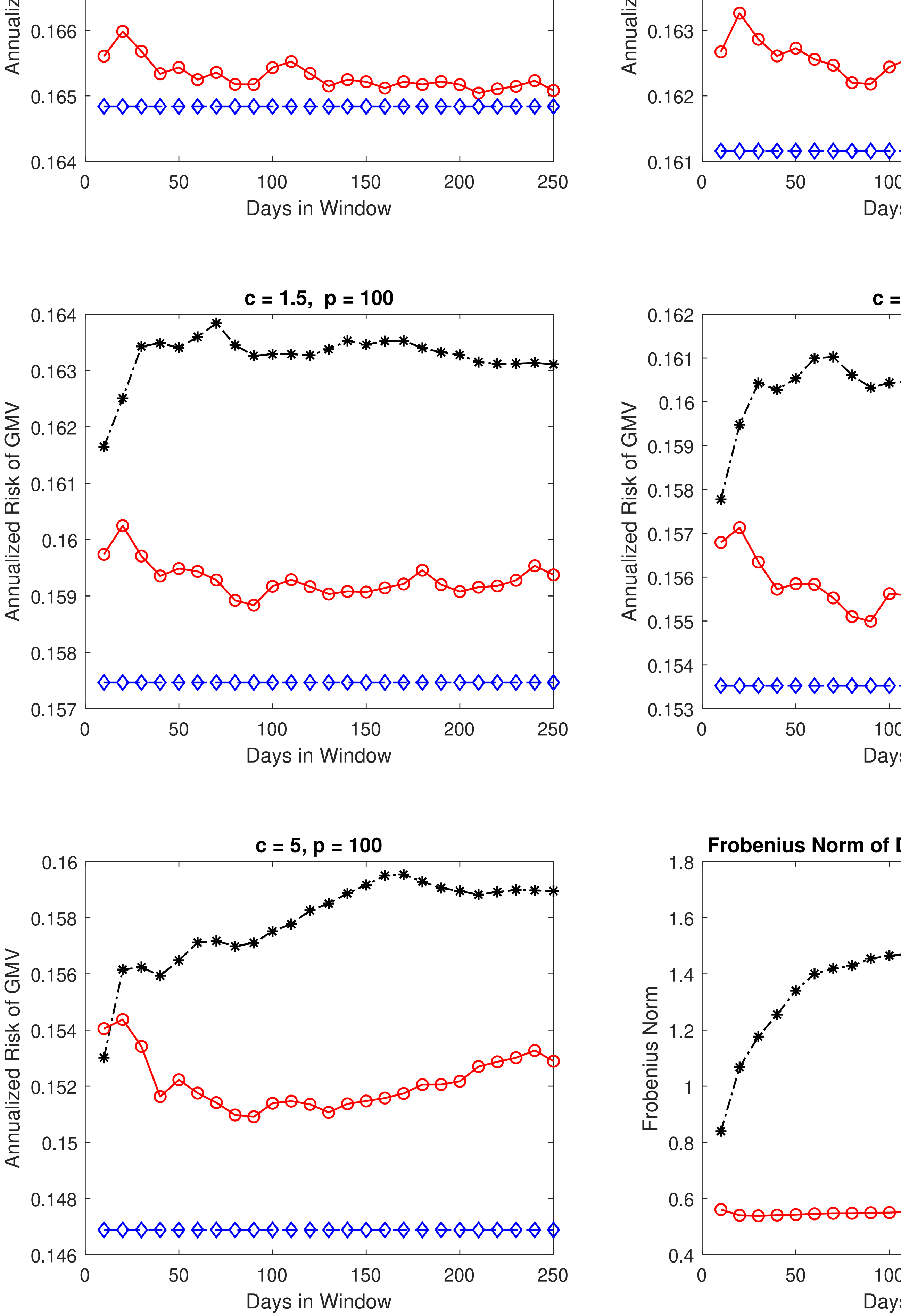
Note: FroNormDiff denotes the averages of $\|\widehat{\boldsymbol{\Sigma}}^*_{T,\tau} - \boldsymbol{\Sigma}^*_{N+1,\delta}\|_F / |\boldsymbol{\Sigma}^*_{N+1,\delta}\|_F \times 100$ and $\|\frac{1}{N}\sum_{k=1}^N \widetilde{\boldsymbol{\Sigma}}_{k,\delta} - \boldsymbol{\Sigma}^*_{N+1,\delta}\|_F / |\boldsymbol{\Sigma}^*_{N+1,\delta}\|_F \times 100$ over 504 investment days. WeightDiff denotes the averages of $|\widehat{\mathbf{w}}^c_{\text{KerAR1}} - \widehat{\mathbf{w}}^c_{\text{Oracle}}|/p \times 100$ and $|\widehat{\mathbf{w}}^c_{\text{Aver}} - \widehat{\mathbf{w}}^c_{\text{Oracle}}|/p \times 100$.

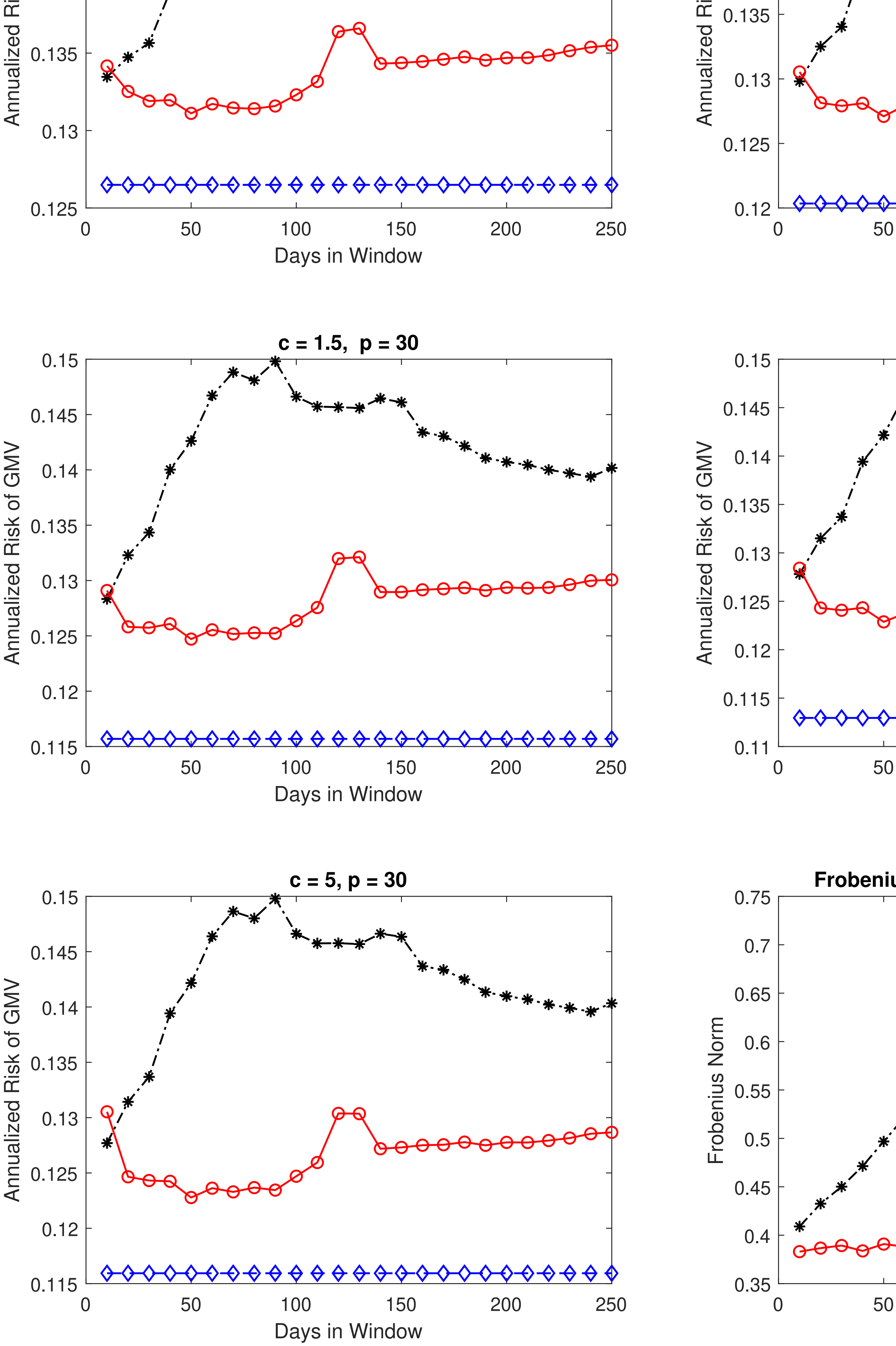

c = 1.5, p = 30
c = 5, p = 30
Frobenius Norm of
Annualized Risk of GMV
Frobenius Norm
Days in Window

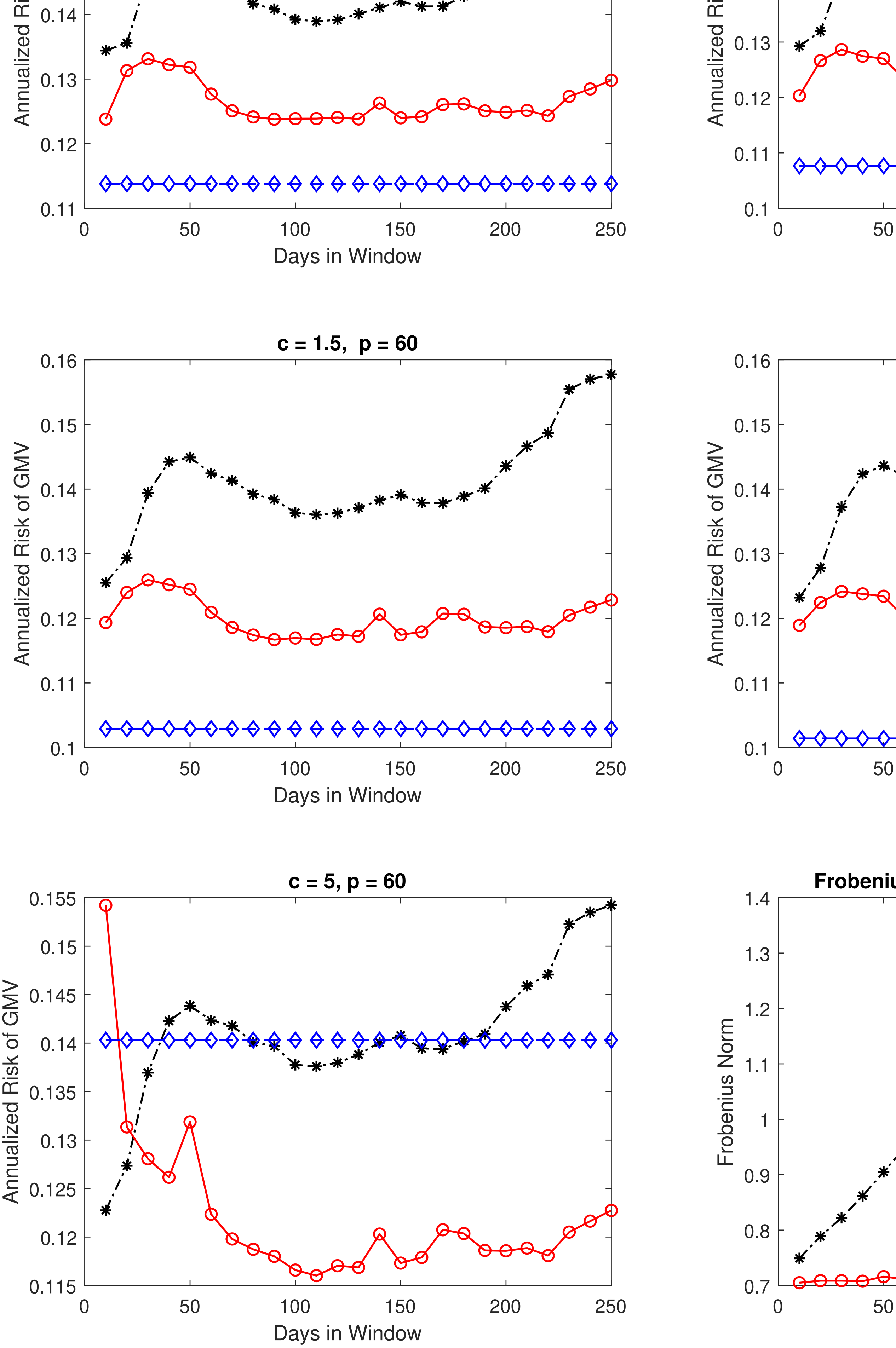

c = 1.5, p = 100
c = 5, p = 100
Frobenius Norm of D
Annualized Risk of GMV
Frobenius Norm
Days in Window

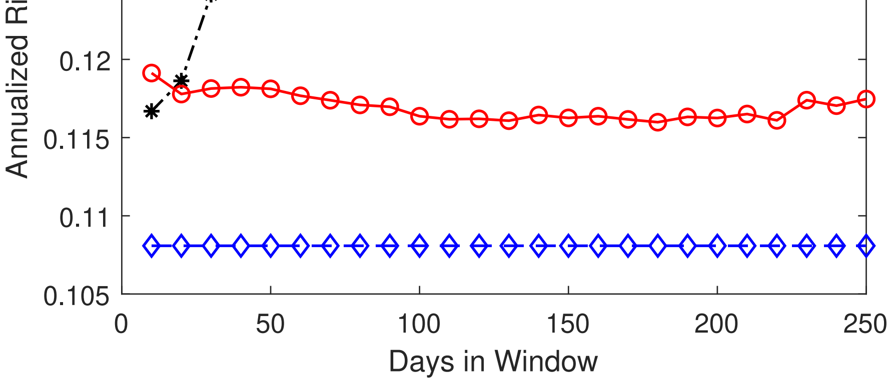


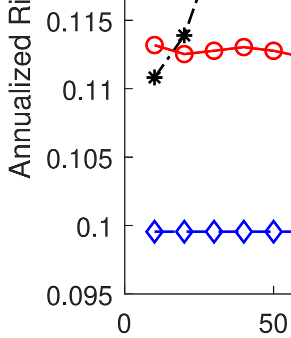
c = 1.5, p = 30
c = 5, p = 30
Frobenius Norm of
Annualized Risk of GMV
Days in Window
Frobenius Norm

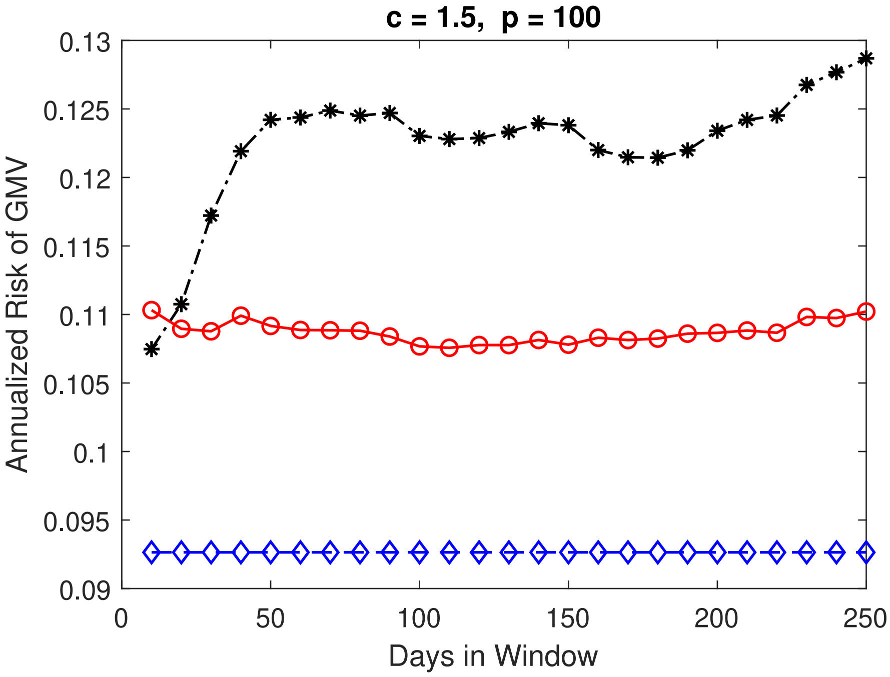

c = 1.5, p = 100
c = 5, p = 100
Frobenius Norm of D
Annualized Risk of GMV
Frobenius Norm
Days in Window

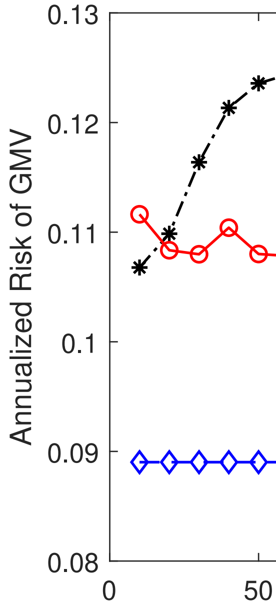

c = 1.5, p = 30
c = 5, p = 30
Frobeniu
Annualized Risk of GMV
Frobenius Norm
Days in Window

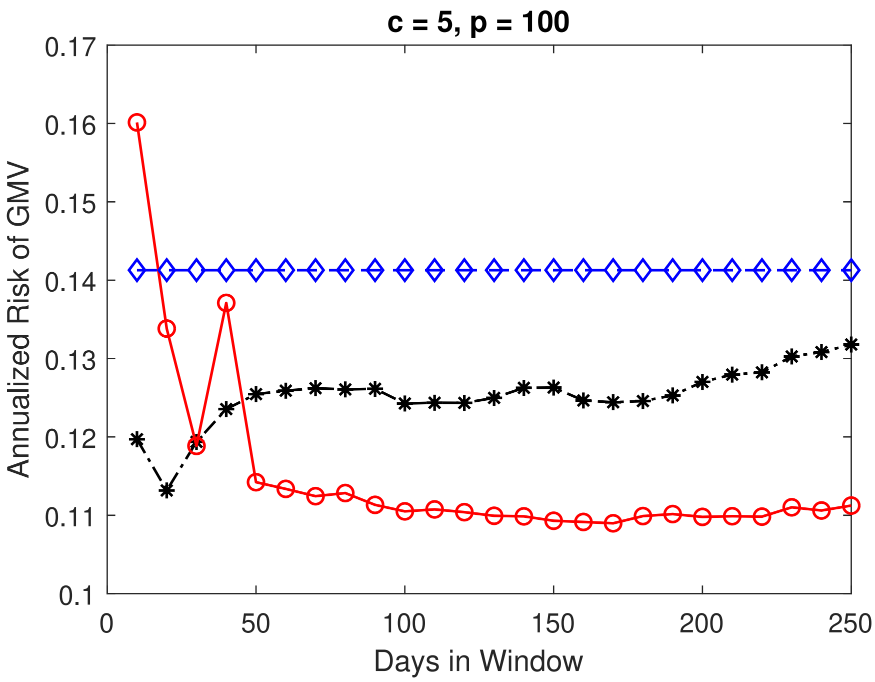

c = 1.5, p = 60
c = 5, p = 60
Annualized Risk of GMV
Days in Window
Frobenius Norm


Annualized Ri
Days in Window


Annualized Ri


c = 1.5, p = 100
Annualized Risk of GMV
Days in Window


Annualized Risk of GMV


c = 5, p = 100
Annualized Risk of GMV
Days in Window

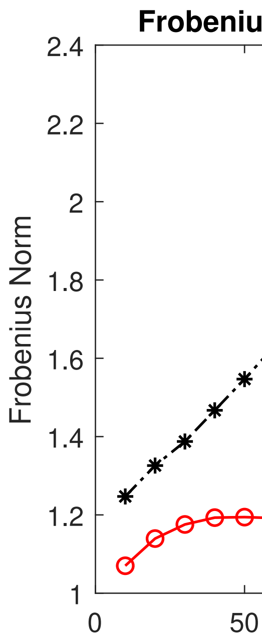

Frobeniu
Frobenius Norm

Table 2: Out-of-sample performances of different daily-rebalanced strategies for the GMV portfolio with different values for the norm constraint c during all trading days between 01/01/2009 to 06/30/2017.

		$1/p$	SP	FF3	LS	NL-SF	DCC	Aver	KerAR1
					$p = 30$				
$c = 1$	AV	9.14	5.45	4.59	4.65	5.66	5.04	6.19	7.42
	SD	19.04	14.64	14.53	15.09	14.16	14.48	13.57	**13.15**
	SR	0.48	0.37	0.32	0.31	0.40	0.35	0.46	0.56
$c = 1.5$	AV	9.14	6.56	4.97	4.71	6.73	5.59	6.80	7.99
	SD	19.04	14.56	14.52	14.90	13.86	14.18	13.43	**12.48**
	SR	0.48	0.45	0.34	0.32	0.49	0.39	0.51	0.64
$c = 2$	AV	9.14	8.07	6.47	5.13	7.71	5.60	6.90	8.04
	SD	19.04	14.90	15.13	15.12	14.12	14.34	13.37	**12.30**
	SR	0.48	0.54	0.43	0.34	0.55	0.39	0.52	0.65
					$p = 60$				
$c = 1$	AV	10.13	9.27	7.26	8.96	7.54	10.03	7.90	9.10
	SD	18.87	19.88	17.70	18.35	17.59	19.07	14.60	**13.38**
	SR	0.54	0.47	0.41	0.49	0.43	0.53	0.54	0.68
$c = 1.5$	AV	10.13	8.50	5.59	8.22	7.17	8.26	8.42	9.26
	SD	18.87	16.36	16.15	15.76	15.27	15.26	13.94	**12.57**
	SR	0.54	0.52	0.35	0.52	0.47	0.54	0.60	0.74
$c = 2$	AV	10.13	8.78	6.45	8.18	7.15	7.80	8.51	8.84
	SD	18.87	15.93	15.71	15.33	14.73	14.49	13.72	**12.37**
					$p = 100$				
	SR	0.54	0.55	0.41	0.53	0.49	0.54	0.62	0.71
$c = 1$	AV	10.08	7.42	7.88	8.15	7.31	9.01	8.28	8.69
	SD	18.92	15.30	15.02	14.39	14.09	14.66	12.42	**11.84**
	SR	0.53	0.49	0.52	0.57	0.52	0.61	0.67	0.73
$c = 1.5$	AV	10.08	7.80	5.55	7.28	6.68	7.12	8.70	8.60
	SD	18.92	13.67	13.86	12.77	12.47	12.46	11.72	**10.90**
	SR	0.53	0.57	0.40	0.57	0.54	0.57	0.74	0.79
$c = 2$	AV	10.08	7.69	6.45	6.23	6.76	6.55	8.40	8.20
	SD	18.92	13.29	13.84	12.46	12.36	12.52	11.64	**10.82**
	SR	0.53	0.58	0.47	0.50	0.55	0.52	0.72	0.76

Note: AV, SD, and SR denote the average, standard deviation, and sharp ratio, respectively, of 8.5 years' daily out-of-sample returns of portfolios. AV and SD are annualized and given as percents. The smallest number in each row labeled by SD is reported in boldface. 250 days' range of historical daily log-returns are used for SP, FF3, LS, NonLS, and NL-SF. 30 days' range of intra-daily log-returns are used for Aver and KerAR1.