A Linear Weight Estimator for Dynamic Global Minimum Variance Portfolio Allocation*

Ekaterina Kazak† Yifan Li‡ Ingmar Nolte§ Sandra Nolte¶

University of Manchester University of Manchester Lancaster University Lancaster University

Abstract

This paper introduces a linear weight estimation (LWE) framework as a novel semi-parametric method for dynamic global minimum variance portfolio (GVMP) allocation. The LWE method assumes a dynamic linear model for the *ex ante* optimal GMVP weights. Based on a time series of daily realized covariance estimates, the LWE model parameters can easily be estimated in closed form using the method of moments. Importantly, we prove that the estimated LWE portfolio weights directly and uniquely minimize a finite sample estimate of the unconditional portfolio variance, which is not achieved by most of the existing methods in the literature. Empirical results demonstrate that LWE outperforms competing estimators in terms of out-of-sample portfolio variance measures. LWE can also be extended to incorporate controls for investment constraints, which further balances its economic performance and transaction costs.

Keywords: portfolio allocation, high-frequency finance, realized measures, forecasting

JEL classification: C51, C58, G11, C32

*Financial support by the MEiN Program of the University of Konstanz is gratefully acknowledged. We are grateful to Peter Reinhard Hansen, Jia Li, Jun Yu, Yichong Zhang for their invaluable suggestions. We would like to thank the participants of the Konstanz-Lancaster Econometrics workshop German Statistical Week, CFE conference in London and SoFiE conference in Cambridge, and SMU econometrics workshop for their helpful comments. All remaining errors are ours.

†Department of Economics, School of Social Sciences, Arthur Lewis Building, Oxford Road, Manchester M13 9PL, email: ekaterina.kazak@manchester.ac.uk.

‡Corresponding author: Alliance Manchester Business School, Booth Street West, Manchester M15 6PB, UK. Phone: +44 (0) 161 306 6402, email: yifan.li@manchester.ac.uk.

§Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. Phone: +44 (0)1524 592644, email: i.nolte@lancaster.ac.uk.

¶Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, UK. Phone: +44 (0)1524 522020, email: s.nolte@lancaster.ac.uk.

1 Introduction

Estimating and forecasting the covariance matrix of an N dimensional return process is considered to be a cornerstone of any portfolio optimization problem. The availability of high-frequency data led to the development of precise nonparametric realized covariance estimators[1], which fundamentally changed the portfolio allocation problem. More specifically, realized covariances serve as an accurate proxy for the otherwise latent covariances among the universe of assets, which provide an observable benchmark for the prediction of future covariance matrices. Exploiting the realized covariances, most of the existing papers typically solve the portfolio allocation problem by the following two-step approach: (1) predicting the N-by-N realized covariance matrix using time-series methods; (2) solving for the optimal portfolio weights based on the predicted covariance matrix, usually referred to as the 'plug-in' estimator.

Nevertheless, forecasting a time series of N-by-N positive definite realized covariance matrices is by itself a Herculean task. One needs to overcome the curse of dimensionality as the number of time series to be predicted grows quadratically with N. In addition, the positive definiteness of the predicted realized covariances is required for a unique solution of the portfolio allocation problem. Several approaches have been proposed to tackle these two issues. One strand of literature suggests to use a structured and simplified forecasting model designed to reduce the dimensionality while preserving the positive definiteness. For example, a random walk model used in Fan et al. (2016), Aït-Sahalia and Xiu (2017), the Wishart-based models of Gourieroux et al. (2009), Noureldin et al. (2012), Golosnoy et al. (2012), Jin and Maheu (2013), the Cholesky-decomposition method of Chiriac and Voev (2011), the HAR-DRD model of Oh and Patton (2016) and its extensions in Bollerslev et al. (2018), and the score-driven model of Opschoor et al. (2018). Shrinkage and factor-based methods are proposed to smooth the forecasting errors in large dimension, see Ledoit and Wolf (2004), Bauer and Vorkink (2011), Fan et al. (2016), Hautsch et al. (2015), Aït-Sahalia and Xiu (2017), Ledoit and Wolf (2017), Engle et al. (2019), Hautsch and Voigt (2019), Ledoit and Wolf (2020), Callot et al. (2021), Ding et al. (2021) among many others.

To reduce the dimensionality of covariance matrix prediction, an emerging stream of literature proposes to model and forecast the N-by-1 realized weights of certain *ex post* optimal weight vectors, which can be computed given the realized covariance matrices (Golosnoy et al., 2019; Cipollini et al., 2021). In addition to the much reduced dimensionality, this approach is further motivated by the fact that portfolio allocation only requires the knowledge of a forward-looking weight vector without forecasting the full covariance matrix. Consequently, Cipollini et al. (2021) show significant improvements in the global minimum variance portfolio (GMVP) performance by forecasting the realized weights

[1] E.g., Barndorff-Nielsen et al. (2011), Zhang (2011), Bibinger and Mykland (2016), Varneskov (2016), Li et al. (2022), among others.

using a simple VARMA(1,1) model over various conventional two-step approaches.

Despite these methodological developments in forecasting realized covariances or weights, there is still a salient but often ignored flaw in the dynamic portfolio allocation exercise. Specifically, the existing forecasting models typically minimize certain distance measures between the true and the predicted values (e.g., matrix or vector norms, the multivariate QLIKE of Patton and Sheppard (2009), see also Laurent et al. (2013) for further multivariate loss functions), which does not translate into an *ex ante* optimal choice of portfolio weights in utility terms. For instance, a mean-squared optimal forecast of the realized GMVP weights is its conditional mean, which in general does not minimize the *ex ante* portfolio variance (see Eq. (2.12) and its implications). As discussed in Section 3.3 of Cipollini et al. (2021), a forecasting model for realized weights that simultaneously minimizes the *ex ante* portfolio variance is theoretically possible but empirically infeasible to apply, as one needs to optimize a highly nonlinear objective function, which is computationally costly especially when N is large. To our best knowledge, a feasible solution to the dynamic portfolio choice problem which directly optimizes the portfolio performance metric remains an open question.

Motivated by the above discussions, this paper proposes a novel linear weight estimation (LWE) framework for dynamic GMVP allocation. In detail, we directly model the *ex ante* optimal GMVP weight vector as a linear function of historical realized weight vectors. We show that the model parameters can be estimated via a standard method of moments that minimizes the unconditional portfolio variance, which yields a semi-parametric estimator for the forward-looking GMVP weights. This allows us to establish asymptotic properties of our estimator using standard econometric tools. Importantly, the number of parameters scales linearly with N, and the estimator is fully linear, and therefore very efficient to compute even for large N. By modifying the loss function of our method, one can also impose bounds on the portfolio weights or regularization techniques to tailor the resulting weight vector according to the investment constraints of the investor, such as those adopted in Jagannathan and Ma (2003), Brodie et al. (2009), Li (2015), Yen (2016), among others.

Our method is closely related to Brandt et al. (2009) who propose to model the *ex ante* optimal portfolio weights as a linear combination of firm characteristics. However, their weight model is static in nature and homogeneous across all stocks, while our model allows for greater flexibility with dynamic portfolio weights with stock-specific weights dynamics. Cipollini et al. (2021) consider a dynamic VARMA-type model for portfolio weights, but the estimated parameters do not guarantee optimal GMVP portfolio performance, as discussed above. Strikingly, in a dynamic daily GVMP allocation study, our empirical findings based on 250 most liquid constituents of the S&P 500 index show that our LWE estimator consistently outperforms the dynamic weight model of Cipollini et al. (2021), and several widely applied dynamic portfolio choice methods, with a smaller out-of-sample average realized portfolio variance uniformly over different choices of N. This result is robust to various investment constraints such as portfolio exposure and short-selling constraints. Transaction costs of the LWE method can be effectively controlled by a ridge-type penalized LWE estimator, which further illustrates the practicality of the proposed method.

The remainder of the paper is organized as follows: Section 2 describes the GMVP allocation problems. Section 3 introduces the novel LWE approach, investigates its theoretical properties, and discusses various extensions. Section 4 examines its performance in an empirical study together with the penalized LWE. Section 5 concludes.

2 The Dynamic GMVP Allocation Problem

We formally state the dynamic GMVP allocation problem in this section and discuss some problems with the existing approaches. Our setting is similar to those in Aït-Sahalia and Xiu (2017), Cipollini et al. (2021), and the importance of the GMVP allocation problem is emphasised in DeMiguel et al. (2009, 2014).

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, we consider a market with N distinct assets[2], whose N-dimensional price process is denoted as $(P_t)_{t\geq 0}$. We denote the t-th trading day as intervals of the form $[t-1, t]$, and it is understood that market closure periods are discarded. For each interval $[t-1, t]$, we shall assume the following structure for the log-return process $r_s = \ln(P_s/P_{t-1})$ with $s \in [t-1, t]$:

Assumption 1. *On $[t-1, t]$, the N-dimensional log-return process takes the following form:*

$$r_s = \int_{t-1}^{s} b_u du + \int_{t-1}^{s} \Theta_u^* dW_u, \tag{2.1}$$

where b_u is an $\mathcal{F}_{t-1}$-adapted locally bounded process representing the spot drift of r_s, Θ_u is an optional element-wise locally bounded stochastic processes called the spot covolatility processes, and W_u is an N-dimensional Wiener process. Denote the quadratic variation process of r_s on $[t-1, t]$ as $\Sigma_t^* = \int_{t-1}^{t} \Theta_s^* (\Theta_s^*)' ds$, we assume that for all $t \in \mathbb{N}$ and $\omega \in \mathbb{R}^N$, it holds almost surely that $0 < \omega' \Sigma_t^* \omega < \infty$ and that Σ_t^* is strictly stationary.

In essence, r_t is a section of an N-dimensional continuous semi-martingale on $[t-1, t]$ with the conditional moments $\mathrm{E}_{t-1}[r_t] = \mathrm{E}_{t-1}[\int_{t-1}^{t} b_s ds] := \mu_t$ and $\mathrm{V}_{t-1}[r_t] = \mathrm{E}_{t-1}[\Sigma_t^*]$ by the Ito isometry, where $\mathrm{E}_{t-1}[\cdot] = \mathrm{E}[\cdot|\mathcal{F}_{t-1}]$ and $\mathrm{V}_{t-1}[\cdot] = \mathrm{V}[\cdot|\mathcal{F}_{t-1}]$ are the $\mathcal{F}_{t-1}$-conditional expectation and variance operators, respectively. The assumption on Σ_t^* ensures that there are no perfectly colinear assets on the market, and the variance of any portfolio over $[t-1, t]$ is finite. The semi-martingale model is

[2] Throughout this paper, we shall assume that N is a fixed finite natural number, so that we are in the conventional fixed dimension setting. The high-dimensional case with $N \to \infty$ can have drastically different theoretical properties and is beyond the scope of this paper, which is left for future research.

typical in the high-frequency literature (Aït-Sahalia and Jacod, 2014), and the continuous assumption is only for notational convenience and should not be viewed as a restriction here, as the quadratic covariation matrix is also well-defined in the presence of jumps (see, for example, Chapter 1 of Aït-Sahalia and Jacod (2014)), albeit with more cumbersome expressions. Importantly, the availability of high-frequency data allows us to construct precise realized covariance estimators for Σ_t^* on a daily basis, which serves as an accurate observable proxy for Σ_t^* that depicts the covariance structure among the N assets. We shall therefore make the following assumption about an observed realized covariance estimator:

Assumption 2. *For each t, the Realized Covariance (RC) estimator Σ_t of Σ_t^* can be constructed from high-frequency observations of $(P_t)_{t\geq 0}$, which has the representation:*

$$\Sigma_t = \Sigma_t^* + S_t, \tag{2.2}$$

where S_t satisfies $\mathrm{E}_{t-1}[S_t] = \mathbf{0}_{N\times N}$, for all t.

Effectively, Assumption 2 assumes that the estimation error of RC, namely $S_t = \Sigma_t - \Sigma_t^*$, is a matrix-valued martingale difference sequence. This assumption is well-grounded in the literature of high-frequency realized measures. For example, the state-of-the-art flat-top realized kernel of Varneskov (2016) shows that, on the interval $[t-1, t]$, the estimation error S_t (after some appropriate scaling) converges to a zero-mean Gaussian process independent of the filtration $\mathcal{F}$ as the number of high-frequency observation increases. Therefore, Assumption 2 simply reflects the asymptotic properties of a desirable RC estimator, which should be robust to the presence of measurement errors and asynchronous trading to eliminate the asymptotic bias.[3]

Importantly, for any choice of weight vector $w \in \mathbb{R}^N$, Assumption 2 ensures that:

$$\mathrm{E}_{t-1}[w'\Sigma_t w] = \mathrm{E}_{t-1}[w'\Sigma_t^* w] + w'\,\mathrm{E}_{t-1}[S_t]w = \mathrm{E}_{t-1}[w'\Sigma_t^* w] = \mathrm{V}_{t-1}[w'r_t|\mathcal{F}_{t-1}], \tag{2.3}$$

so that minimizing the conditional realized portfolio variance is equivalent to a minimization of the true conditional portfolio variance. This provides an observable optimization target, from which the optimal GMVP weights can be obtained. Consequently, for two weight vectors w_1 and w_2, we always have:

$$\mathrm{E}_{t-1}[w_1'\Sigma_t w_1] \lesseqgtr \mathrm{E}_{t-1}[w_2'\Sigma_t w_2] \Leftrightarrow \mathrm{V}_{t-1}[w_1' r_t|\mathcal{F}_{t-1}] \lesseqgtr \mathrm{V}_{t-1}[w_2' r_t|\mathcal{F}_{t-1}], \tag{2.4}$$

so that under Assumption 2, the noisy proxy Σ_t still delivers consistent rankings of portfolio variance for different portfolio weights in the spirit of Hansen and Lunde (2006).

[3]Several RC estimators have been proposed to achieve asymptotic unbiasedness, including Barndorff-Nielsen et al. (2011), Bibinger and Mykland (2016), Lunde et al. (2016), Varneskov (2016), Boudt et al. (2017), Li et al. (2022) among others.

We now describe the dynamic GMVP optimization problem in detail. At the beginning of day $t-1$, a representative investor invests all her capital into the N assets based on the information set $\mathcal{F}_{t-1}$ by choosing an $\mathcal{F}_{t-1}$-adapted vector of weights w_t that sums up to one. She holds the portfolio till time t (the end of day t) and closes her position, and the procedure iterates indefinitely on day t, $t+1$, etc. This daily trading horizon ensures that the investor avoids holding the assets during market down times, such as overnight and weekends, whose associated risk is challenging to quantify because the high-frequency data required to construct valid realized risk measures is no longer available.

To choose the weight vector w_t, the investor solves the following one-day-ahead variance minimization problem:

$$\min_{w\in\mathcal{W}} \mathrm{V}[w'r_t|\mathcal{F}_{t-1}] = \min_{w\in\mathcal{W}} \mathrm{E}_{t-1}[w'\Sigma_t w], \tag{2.5}$$

where $\mathcal{W} = \{w \in \mathbb{R}^N : \iota' w = 1\}$, and ι is an N-by-1 vector of ones. The above formulation is consistent with a power expected utility framework with an infinite risk aversion coefficient. Viewing w as an argument rather than a random variable and by the linearity of the expectation operator, we see that:

$$\min_{w\in\mathcal{W}} \mathrm{E}_{t-1}[w'\Sigma_t w] \Leftrightarrow \min_{w\in\mathcal{W}} w'\Omega_t w, \tag{2.6}$$

where $\Omega_t := \mathrm{E}_{t-1}[\Sigma_t^*]$ is the expected variance-covariance matrix of the assets. Given Ω_t, the above problem has a well-known analytical solution:

$$w_t^{\star} = \frac{\Omega_t^{-1}\iota}{\iota'\Omega_t^{-1}\iota}. \tag{2.7}$$

The weight vector $w_t^{\star}$ is only optimal at a single time point t. The optimal dynamic GMVP strategy is defined by collecting the optimal GMVP weights $\{w_t^{\star}\}_{t=1,2,\ldots}$ over time, which solves the following unconditional GMVP problem:

$$\min_{w_t\in\mathcal{F}_{t-1}\cap\mathcal{W}} \mathrm{E}[w_t\Sigma_t w_t] = \min_{w_t\in\mathcal{F}_{t-1}\cap\mathcal{W}} \mathrm{E}[w_t\Omega_t w_t]. \tag{2.8}$$

Intuitively, $\{w_t^{\star}\}_{t=1,2,\ldots}$ yields the smallest average $\mathcal{F}_{t-1}$-conditional variance of the portfolio $w_t'r_t$ among all dynamic GMVP strategies $\{w_t\}_{t=1,2,\ldots}$, which is the ideal strategy for the investor. Unfortunately, $\{w_t^{\star}\}_{t=1,2,\ldots}$ is an infeasible optimal strategy as Ω_t is in general an unknown function of the information set $\mathcal{F}_{t-1}$.

To arrive at a feasible dynamic GMVP strategy, we may impose some structure on Ω_t and compute $w_t^{\star}$ accordingly. In general, one can consider a (potentially misspecified) parametric model $\Omega_t(\theta_0;\mathcal{F}_{t-1})$ for Ω_t, where θ_0 is some time-invariant parameter vector that best approximates the data generating process. Under the assumption that the parametric model is correctly specified, one immediately sees that:

$$w_t^{\star} = w_t(\theta_0), \quad \text{where } w_t(\theta) = \frac{\Omega_t(\theta;\mathcal{F}_{t-1})^{-1}\iota}{\iota'\Omega_t(\theta;\mathcal{F}_{t-1})^{-1}\iota}. \tag{2.9}$$

Under the correct model specification, θ_0 is the solution to the following two unconditional minimization problems:

$$\begin{aligned} &\min_\theta \mathrm{E}[||\Sigma_t - \Omega_t(\theta;\mathcal{F}_{t-1})||_F^2] && \text{(the forecasting problem) and,} \\ &\min_\theta \mathrm{E}\left[w_t(\theta)'\Sigma_t w_t(\theta)\right] && \text{(the minimum variance problem),} \end{aligned} \tag{2.10}$$

where $||\cdot||_F$ is the Frobenius norm. Intuitively, if $\Omega_t(\theta;\mathcal{F}_{t-1})$ correctly specifies the parametric family for the conditional mean dynamics of Σ_t^* such that $\mathrm{E}_{t-1}[\Sigma_t^*] = \mathrm{E}_{t-1}[\Sigma_t] = \Omega_t(\theta_0;\mathcal{F}_{t-1})$, then $\Omega_t(\theta_0;\mathcal{F}_{t-1})$ is the element-wise mean-squared optimal one-step-ahead forecast for Σ_t, thus θ_0 must solve the forecasting problem above. Simultaneously, the minimum variance problem is solved by construction. As all the quantities above are observed, one can easily construct finite sample analogs of the two minimization problems using sample means instead of expectations.

The dual interpretation of θ_0 in Eq. (2.10) motivates the so-called 'plug-in' approach in the literature. In essence, this approach first solves the finite sample forecasting problem to obtain the parameter estimate $\hat{\theta}^{plug-in}$. This estimate is then plugged into the optimal GMVP weight vector in Eq. (2.9), which gives $w_t(\hat{\theta}^{plug-in})$ as the estimated GMVP weight at time t. As the number of observations grows, $\hat{\theta}^{plug-in}$ becomes closer to θ_0 so that the minimum variance problem is also solved simultaneously.

Nevertheless, there is an inherent problem with the plug-in approach, in addition to the challenging task of forecasting the N-by-N positive definite matrix time series. Regardless of whether the parametric model is correctly specified, the solutions to the finite sample versions of the forecasting problem and the minimum variance problem do not necessarily coincide due to the different objective functions. This implies that $\hat{\theta}^{plug-in}$ does not necessarily minimize the historical averaged portfolio variance, which is the goal of the GMVP allocation. Moreover, under model misspecification, the two minimization problems in Eq. (2.10) may not have the same limiting solution, so $\hat{\theta}^{plug-in}$ produces a suboptimal dynamic GMVP strategy within the parametric family, even with an increasing amount of data. Here, a potentially viable approach to avoid the inconsistencies between the two minimization problems in Eq. (2.10) is to estimate θ_0 directly from the finite sample minimum variance problem. However, this approach is rarely adopted in the existing literature, which is mainly due to the complex nonlinear functional form of the target function w.r.t. θ that greatly complicates the numerical optimization procedure, especially for large N.

Instead of attempting to solve the challenging forecasting problem, from the sequence of Σ_t we can construct the sequence of *ex post* optimal GMVP weights, or the *realized* weights, denoted by:

$$w_t^{\star} = \frac{\Sigma_t^{-1}\iota}{\iota'\Sigma_t^{-1}\iota}. \tag{2.11}$$

These weights $\{w_t^{\star}\}_{t=1:T}$ are $\mathcal{F}_t$-adaptive and only available at the end of day t. Due to the intimate connection between Σ_t and Ω_t, one may expect to learn about $w_t^{\bigstar}$ from the observed sequence of weights $\{w_t^{\star}\}_{t=1:T}$, which is an N-dimensional vector. As investors only require an estimate of $w_t^{\bigstar}$ at time $t-1$ to place orders for the trading activity at time t, this circumvents the problem of forecasting large positive semi-definite matrices.

Given the sequence $\{w_t^{\star}\}_{t=1,2,\ldots}$, a simple estimator of $w_t^{\bigstar}$ would be the $\mathcal{F}_{t-1}$ optimal conditional forecast of $w_t^{\star}$ in the mean-squared sense, namely $\mathrm{E}_{t-1}[w_t^{\star}]$. As discussed in Cipollini et al. (2021), this forecast can be easily computed using an ARMA-type model. Unfortunately, this estimator also does not solve the unconditional GMVP problem, since:

$$\mathrm{E}_{t-1}[w_t^{\star}] = \mathrm{E}_{t-1}\left[\frac{\Sigma_t^{-1}\iota}{\iota'\Sigma_t^{-1}\iota}\right] \neq \frac{\Omega_t^{-1}\iota}{\iota'\Omega_t^{-1}\iota} = w_t^{\bigstar}, \tag{2.12}$$

due to the nonlinearity of the matrix inverses, unless some restrictive distributional assumptions are imposed on Σ_t. For example, the results in Okhrin and Schmid (2006) imply that $\mathrm{E}_{t-1}[w_t^{\star}] = w_t^{\bigstar}$ if Σ_t is conditionally Wishart with the conditional mean Ω_t. However, this is unlikely to hold for the observed realized covariance matrices. Therefore, an MSE-optimal prediction model for realized weights also does not directly solve the unconditional GMVP problem.

3 A Linear Framework for Optimal GMVP Weight Vector

This paper proposes a linear weight estimation (LWE) framework to solve the unconditional GMVP problem in Eq. (2.8) by modelling the optimal weight vector $w_t^{\bigstar}$ with a linear model. As $w_t^{\bigstar}$ must sum up to one, it suffices to only model the first $N-1$ (or any $N-1$ weights by permuting $w_t^{\bigstar}$) weights, which we denote by $\tilde{w}_t^{\bigstar}$. To fix ideas, we introduce the subspace $\tilde{\mathcal{W}} = \{\tilde{w} \in \mathbb{R}^{N-1} : \tilde{w} = Zw, w \in \mathcal{W}\}$, where $Z = [I_{N-1}, \mathbf{0}_{(N-1)\times 1}]$ is the mapping matrix that removes the last element from w. This induces the following linear mappings between $\mathcal{W}$ and $\tilde{\mathcal{W}}$:

$$\forall w \in \mathcal{W}, \tilde{w} \in \tilde{\mathcal{W}} : \tilde{w} = Zw, w = \tilde{Z}\tilde{w} + \boldsymbol{e}_N, \tag{3.1}$$

where $\tilde{Z} = [I_{N-1}, -Z\iota]'$, and $\boldsymbol{e}_N = (0, \ldots, 0, 1)'$ is the standard basis for the N-th dimension. Note, that by reducing one dimension from $\mathcal{W}$, we effectively eliminate the effect of the sum up constraint such that $\tilde{\mathcal{W}} = \mathbb{R}^{N-1}$. However, for every $\tilde{w}$, there exists $\boldsymbol{e}_N + \tilde{Z}\tilde{w} \in \mathcal{W}$ such that $Z(\boldsymbol{e}_N + \tilde{Z}\tilde{w}) = \tilde{w}$, so $\mathcal{W}$ and $\tilde{\mathcal{W}}$ are isomorphic. Importantly, $\mathcal{W}$ is only closed under affine transformations, i.e., for $w_1, w_2 \in \mathcal{W}$, $\lambda_1 w_1 + \lambda_2 w_2 \in \mathcal{W}$ iff $\lambda_1 + \lambda_2 = 1$. However, $\tilde{\mathcal{W}}$ is closed under linear transformations such that for any matrices $\Lambda_1, \Lambda_2 \in \mathbb{R}^{(N-1)\times(N-1)}$, $\tilde{w}_1, \tilde{w}_2 \in \tilde{\mathcal{W}}$ implies $\Lambda_1 \tilde{w}_1 + \Lambda_2 \tilde{w}_2 \in \tilde{\mathcal{W}}$. This property allows us to safely ignore the sum up constraint in building our linear model, which significantly simplifies the model construction.

It is convenient to consider the space generated by the zero-investment portfolio (ZIP) weights, defined by $\mathcal{V} = \{v \in \mathbb{R}^N : \iota' v = 0\}$. It is clear that for any $w_1, w_2 \in \mathcal{W}$, $w_1 - w_2 \in \mathcal{V}$, so $\mathcal{V}$ naturally contains the modelling errors for weight vectors. We shall also define its $N-1$ dimensional subspace as $\tilde{\mathcal{V}} = \{\tilde{v} : Zv, v \in \mathcal{V}\}$. For every $\tilde{v} \in \tilde{\mathcal{V}}$, there exists $\tilde{Z}\tilde{v} \in \mathcal{V}$ with $Z\tilde{Z}\tilde{v} = \tilde{v}$, so $\mathcal{V}$ and $\tilde{\mathcal{V}}$ are isomorphic. The dimension reduction does not alter the interpretation of the space $\tilde{\mathcal{V}}$, since for $\tilde{w}_1, \tilde{w}_2 \in \tilde{\mathcal{W}}$, $\tilde{w}_1 - \tilde{w}_2 \in \tilde{\mathcal{V}}$. In essence, $\tilde{\mathcal{W}} = \tilde{\mathcal{V}} = \mathbb{R}^{N-1}$ which is closed under addition and linear transformation.

We begin with a conditional moment condition which characterizes the relationship between the dimension-reduced weight vectors $\tilde{w}_t^{\bigstar}$ and $\tilde{w}_t^{\star}$:

Proposition 1. *Denote $\tilde{\epsilon}_t = \tilde{w}_t^{\star} - \tilde{w}_t^{\bigstar}$ and $\tilde{\Sigma}_t = \tilde{Z}'\Sigma_t\tilde{Z}$. The following identity holds for all t:*

$$\mathrm{E}_{t-1}[\tilde{\Sigma}_t \tilde{\epsilon}_t] = \mathbf{0}_{(N-1)\times 1}. \tag{3.2}$$

We can naturally interpret $\tilde{\epsilon}_t$ as the realization error between the *ex post* and the *ex ante* GMVP weight vectors. Proposition 1 shows that, although $\mathrm{E}_{t-1}[\tilde{\epsilon}_t] \neq 0$ in general, the transformed realization error $\tilde{\Sigma}_t \tilde{\epsilon}_t$ is in fact a martingale difference sequence. Therefore, given a parametric model for the unobserved $\tilde{w}_t^{\bigstar}$, Proposition 1 provides us with the required moment conditions for parameter identification, which will be exploited repeatedly in this paper.

In view of the above discussions, we consider the following linear predictive model for some $\mathcal{F}_t$-predictable element of $\tilde{\mathcal{W}}$:

$$\tilde{w}_t(\boldsymbol{\beta}) = b_0 + \sum_{j=1}^{K} B_j \tilde{X}_{jt} = \boldsymbol{X}_t \boldsymbol{\beta}, \quad b_0, \tilde{X}_{jt} \in \tilde{W}, \quad B_j \in \mathbb{R}^{(N-1)\times(N-1)}, \tag{3.3}$$

where:

$$\boldsymbol{X}_t = [I_{N-1}, \tilde{X}'_{1t} \otimes I_{N-1}, \ldots, \tilde{X}'_{Kt} \otimes I_{N-1}]_{(N-1)\times b(K)}, \quad \boldsymbol{\beta} = [b_0; \mathrm{vec}(B_1); \cdots ; \mathrm{vec}(B_K)]_{b(K)\times 1},$$

in which $b(K) = K(N-1)^2 + N - 1$ is the total number of active parameters to be estimated, and $\mathrm{vec}(A)$ is the vectorization operator for a matrix A. Notice the requirement that $\tilde{X}_{jt} \in \tilde{\mathcal{W}}$, which suggests that there exists some $X_{jt} \in \mathcal{W}$ such that $X_{jt} = \boldsymbol{e}_N + \tilde{Z}\tilde{X}_{jt}$. Intuitively, X_{jt} is the full N-by-1 predictive weight vector that enters into the model. The sum-up constraint on X_{jt} allows us to reduce the dimension of the predictive vector and include only $\tilde{X}_{jt}$ in the model, which simplifies the model construction. So if one would like to include a general N-by-1 vector $\check{X}_{jt} \in \mathbb{R}^N$, one should first standardize it and set $X_{jt} = \check{X}_{jt}/(\iota' \check{X}_{jt})$ which conforms with the structure of $\mathcal{W}$. The full N-by-1 version of the model is thus understood as:

$$w_t(\boldsymbol{\beta}) = \boldsymbol{e}_N + \tilde{Z}b_0 + \sum_{j=1}^{K} \tilde{Z}B_j\tilde{X}_{jt} = \boldsymbol{e}_N + \tilde{Z}\boldsymbol{X}_t\boldsymbol{\beta}. \tag{3.4}$$

Intuitively, b_0 is a vector of baseline unconditional weights understood as the intercept of the model. The sequence of vectors $\{\tilde{X}_{jt}\}_{j=1:K}$ are K sets of $\mathcal{F}_t$-**predictable** explanatory variables, i.e., $\{\tilde{X}_{jt}\}_{j=1:K}$ are observed at time $t-1$. These explanatory variables determine the time-varying component of $\tilde{w}_t(\boldsymbol{\beta})$ through the linear structure $\boldsymbol{X}_t\boldsymbol{\beta}$.

The linear model in Eq. (3.3) is highly flexible as it allows each predictor X_{jt} to have a different impact to each component of $w_t(\boldsymbol{\beta})$. However, in the fully specified model, there are $K(N-1)^2+N-1$ free parameters to be estimated, so we still suffer from the curse of dimensionality as N increases. Instead, one can consider the following diagonal-type specification:

$$\tilde{w}_t(\boldsymbol{\beta}) = b_0 + \sum_{j=1}^{K} \Big(B_j - \frac{\tilde{\iota}b_j' - b_{jN}\tilde{\iota}\tilde{\iota}'}{N}\Big)\tilde{X}_{jt}, \tag{3.5}$$

where B_j are $(N-1)$-by-$(N-1)$ diagonal matrices such that $\tilde{\iota}'B_j = b_j$, $\tilde{\iota} = Z\iota$ is an $(N-1)$-by-1 vector of ones, and b_{jN} is a scalar parameter capturing the impact of the N-th element of X_{jt} through the sum up constraint. In this case, there are $b(K) = KN + N - 1$ parameters in total, which considerably reduces the number of parameters to be estimated. The diagonal specification can be written compactly as $\tilde{w}_t(\boldsymbol{\beta}) = \boldsymbol{X}_t\boldsymbol{\beta}$ with the following $\boldsymbol{X}_t$ and $\boldsymbol{\beta}$:

$$\boldsymbol{X}_t = [I_{N-1}, \tilde{X}^*_{1t}, \ldots, \tilde{X}^*_{Kt}], \quad \boldsymbol{\beta} = [b_0; b_1; b_{1N}; \ldots; b_K; b_{KN}],$$

where:

$$\tilde{X}^*_{jt} = \left[\text{diag}^{\text{-1}}(\tilde{X}_{jt}) - \frac{\tilde{\iota}\tilde{X}'_{jt}}{N}, \frac{\tilde{\iota}\tilde{\iota}'\tilde{X}_{jt}}{N}\right]_{(N-1)\times N}, \tag{3.6}$$

and $\text{diag}^{\text{-1}}(a)$ converts the vector a into a diagonal matrix.

The peculiar $O(N^{-1})$ terms in Eq. (3.5) requires some discussion. This term ensures that the resulting model does not depend on the ordering of the assets, i.e., the model is permutation invariant. Clearly, if one simply specifies Eq. (3.5) without the $O(N^{-1})$ terms, the resulting model depends on which asset is left out from the N-variate system. One way to introduce permutation invariance is to construct the model in a 'jackknife' fashion, i.e., averaging all N distinct diagonal models constructed by leaving out the N assets one at a time. After some simple algebra, one can show that such averaged model is identical to the one in Eq. (3.5), which justifies its unusual specification.

To be even more parsimonious, one can consider a scalar specification for Eq. (3.3):

$$\boldsymbol{X}_t = [I_{N-1}, X_{1t}, \ldots, X_{Kt}], \quad \boldsymbol{\beta} = [b_0; b_K], \tag{3.7}$$

where b_K is a K-by-1 vector capturing the impact of the K predictor weights, so $b(K) = N + K - 1$. Intuitively, in this specification, $\tilde{w}_t(\boldsymbol{\beta})$ can be interpreted as a weighted average of the weights $\tilde{X}_{jt}$ and the baseline weight vector b_0, which is clearly permutation invariant.

Given a suitably parametrized linear model, we shall make the following assumption for the data generating process of $w_t^\star$:

Assumption 3. *Assume that, for some fixed $K > 0$, the $(N-1) \times b(K)$ matrix-valued $\mathcal{F}_t$-predictable process $\boldsymbol{X}_t$ is strictly stationary with full rank, and that the matrix $\mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t]$ is finite and invertible. The process $\boldsymbol{X}_t$ satisfies the following equation for all t:*

$$\tilde{w}_t^\star = \tilde{w}_t(\boldsymbol{\beta}_0) + \tilde{\eta}_t = \boldsymbol{X}_t \boldsymbol{\beta}_0 + \tilde{\eta}_t, \tag{3.8}$$

where $\boldsymbol{\beta}_0 \in \mathrm{int}(\mathcal{B})$ and $\mathrm{int}(\mathcal{B})$ denotes the interior of the parameter space $\mathcal{B}$, a compact subset of $\mathbb{R}^{b(K)}$. The process $\tilde{\eta}_t \in \tilde{\mathcal{V}}$ is a strictly stationary and $\mathcal{F}_t$-predictable process satisfying $\mathrm{E}[\tilde{\eta}_t] = \mathbf{0}_{(N-1)\times 1}$, $\mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \tilde{\eta}_t] = \mathbf{0}_{b(K)\times 1}$. We further require the following law of large number and central limit results to hold as $T \to \infty$:

$$\begin{aligned} &T^{-1} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t \xrightarrow{p} \mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t], \quad T^{-1} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t (\tilde{\eta}_t + \tilde{\epsilon}_t) \xrightarrow{p} \mathbf{0}_{b(K)\times 1}, \\ &T^{-1/2} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t (\tilde{\eta}_t + \tilde{\epsilon}_t) \xrightarrow{d} \mathcal{N}(\mathbf{0}_{b(K)\times 1}, \boldsymbol{C}_\infty), \end{aligned} \tag{3.9}$$

where $\boldsymbol{C}_\infty$ is the long-run variance-covariance matrix of $T^{-1/2} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t (\tilde{\eta}_t + \tilde{\epsilon}_t)$, which is assumed to be finite and almost surely positive definite.

Recall the definition of $\tilde{w}_t^\star$ in Eq. (2.7), Assumption 3 directly specifies the dynamic of $\tilde{w}_t^\star$ through the predictive linear model $\boldsymbol{X}_t \boldsymbol{\beta}_0$ and an additive modelling error term $\tilde{\eta}_t$ summarizing possible misspecification or omitted variables. In view of Proposition 1, the assumption $\mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \tilde{\eta}_t] = \mathbf{0}_{b(K)\times 1}$ is an identification condition, which states that $\tilde{\eta}_t$ should satisfy the same moment condition as the realization error $\tilde{\epsilon}_t$ (otherwise $\boldsymbol{\beta}_0$ cannot be recovered). Note that it does not appear possible to express $\boldsymbol{X}_t$ and $\boldsymbol{\beta}_0$ given a model for Ω_t, except from some trivial special cases (see Eq. (3.16) for instance). Therefore, by modelling the *ex ante* optimal weights directly, we are agnostic about the dynamic structure of Σ_t implied by the linear weight model.

The high-level asymptotic results assumed in Eq. (3.9) are quite standard in the literature of GMM estimation theory, which requires some technical conditions that guarantee the existence of certain moments and a restriction to the dependence of the associated processes (for example, a suitable mixing condition). For brevity, we omit the exact technical conditions here, which can be found in classic texts such as McLeish (1975), Domowitz and White (1982), Newey and McFadden (1994), Hall (1996), among many others.

The term $\tilde{\eta}_t + \tilde{\epsilon}_t$ in Eq. (3.9) requires some more discussion. Under Assumption 3, the realized GMVP weight can be expressed as a linear model:

$$\tilde{w}_t^{\star} = \boldsymbol{X}_t\boldsymbol{\beta}_0 + \tilde{\eta}_t + \tilde{\epsilon}_t, \tag{3.10}$$

where $\tilde{\eta}_t + \tilde{\epsilon}_t$ can be interpreted as the overall residual of fitting $\tilde{w}_t^{\star}$ by $\boldsymbol{X}_t\boldsymbol{\beta}_0$. For example, setting $X_{jt} = \tilde{w}_{t-j}^{\star}$ for $j \in \{1, \ldots, K\}$, we arrive at a VAR(K)-type model for $\tilde{w}_t^{\star}$, where $\boldsymbol{X}_t$ exploits the autoregressive structure in the realized GMVP weights, as advocated by Golosnoy et al. (2019), Cipollini et al. (2021):

$$\tilde{w}_t^{\star} = b_0 + \sum_{j=1}^{K} B_j \tilde{w}_{t-j}^{\star} + \tilde{\eta}_t + \tilde{\epsilon}_t. \tag{3.11}$$

Therefore, the central limit assumption in Eq. (3.9) can be interpreted as a condition on the long-run behaviour of the overall (weighted) residuals of the linear model. Here, one might attempt to estimate $\boldsymbol{\beta}_0$ by fitting the above model using least squares. However, from Eq. (2.12) we know that $\mathrm{E}[\tilde{\epsilon}_t] \neq 0$ in general, so an ordinary least square estimator is both biased and inconsistent. A more critical issue is that the OLS estimator of $\boldsymbol{\beta}_0$ does not directly optimize the unconditional GMVP problem in Eq. (2.8), thus it is not a viable approach here.

To construct a consistent estimator for $\boldsymbol{\beta}_0$ which solves Eq. (2.8), we start with the following characterization of $\boldsymbol{\beta}_0$ under Assumption 3:

Theorem 1. *Under Assumption 3, $\boldsymbol{\beta}_0$ is the unique minimizer of the following unconditional optimization problems:*

$$\min_{\boldsymbol{\beta} \in \mathcal{B}} \mathrm{E}[w_t(\boldsymbol{\beta})'\Sigma_t w_t(\boldsymbol{\beta})] \Leftrightarrow \min_{\boldsymbol{\beta} \in \mathcal{B}} \mathrm{E}[\tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t \tilde{u}_t(\boldsymbol{\beta})], \tag{3.12}$$

where $\tilde{\Sigma}_t = \tilde{Z}'\Sigma_t\tilde{Z}$ and $\tilde{u}_t(\boldsymbol{\beta}) = \tilde{w}_t^{\star} - \boldsymbol{X}_t\boldsymbol{\beta}$ is the residual of fitting $\tilde{w}_t^{\star}$ by $\boldsymbol{X}_t\boldsymbol{\beta}$. The solution is characterized by the following $b(K)$ unconditional moment conditions:

$$\mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta}_0)] = 0, \tag{3.13}$$

whose solution has the following explicit form:

$$\boldsymbol{\beta}_0 = \mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\boldsymbol{X}_t]^{-1}\,\mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\tilde{w}_t^{\star}]. \tag{3.14}$$

Theorem 1 suggests three interpretations of $\boldsymbol{\beta}_0$. First, one can interpret $\boldsymbol{\beta}_0$ as the parameter vector that minimizes the unconditional GMVP problem in Eq. (2.8). Due to the presence of possible misspecification given by η_t, it is not possible to exactly solve Eq. (2.5) for each t under the linear framework. However, unconditionally, η_t no longer contributes to the optimization problem by the orthogonality condition in Assumption 3. This allows us to express $\boldsymbol{\beta}_0$ as a function of the unconditional expectations of the **observed** time series $\tilde{\Sigma}_t$ and $\boldsymbol{X}_t$ without explicitly specifying their dynamic structures. This indicates that $w_t(\boldsymbol{\beta}_0)$ is the solution to the unconditional GMVP problem in Eq. (2.8) under a possibly misspecified linear model.

The second optimization problem in Eq. (3.12) suggests that $\boldsymbol{\beta}_0$ can also be interpreted as the solution of a generalized least square (GLS)-type problem $\tilde{w}_t^{\star} = \boldsymbol{X}_t\boldsymbol{\beta} + \tilde{u}_t(\boldsymbol{\beta})$ by minimizing the expectation of the distance measure $\mathrm{E}[\tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta})]$. However, this is not a conventional GLS problem because (1) $\mathrm{E}[\tilde{u}_t(\boldsymbol{\beta}_0)] = \mathrm{E}[\tilde{\epsilon}_t] \neq 0$ due to Eq. (2.12); (2) $\tilde{\Sigma}_t$ is non-deterministic and depends on $\tilde{u}_t(\boldsymbol{\beta}_0)$. As a result, standard asymptotic results for GLS regressions do not apply in this context.

The third interpretation is related to the set of moment conditions in Eq. (3.13), which resembles the moment conditions under the generalized method of moments (GMM) framework. In our case, the number of moment conditions equal the number of parameters to be estimated, thus one can interpret Eq. (3.13) as an exactly-identified special case of a GMM estimation problem. In this case, one can consistently estimate $\boldsymbol{\beta}_0$ by the standard method-of-moment estimator, which shall be discussed later on in this section.

It is worth discussing the property of the process that satisfies the moment condition, which we shall denote by $g_t = \boldsymbol{X}_t'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta}_0) = \boldsymbol{X}_t'\tilde{\Sigma}_t(\tilde{\eta}_t + \tilde{\epsilon}_t)$. Using Proposition 1, we see that:

$$\mathrm{E}_{t-1}[g_t] = \boldsymbol{X}_t'\tilde{\Omega}_t\tilde{\eta}_t, \quad \mathrm{E}[\boldsymbol{X}_t'\tilde{\Omega}_t\tilde{\eta}_t] = 0, \tag{3.15}$$

where the first relation is due to $\tilde{\Sigma}_t\tilde{\epsilon}_t$ being a martingale difference sequence (MDS), and the second relation follows from Assumption 3. It is worth noting that the process g_t is not a martingale difference sequence in general since $\tilde{\eta}_t$ is adapted to $\mathcal{F}_{t-1}$ and does not vanish after taking conditional expectations. Therefore, g_t is expected to be an autocorrelated zero-mean process unless $\eta_t = 0$ for all t, in which case g_t becomes an MDS. This property can be exploited to assess the overall goodness-of-fit of the linear model to the *ex ante* optimal GMVP weights.

As a sanity check, we show that when $\tilde{w}_t^{\bigstar} = \boldsymbol{\beta}_0$ is time-invariant, $\boldsymbol{\beta}_0$ reduces to the GMVP weight vector implied by the unconditional variance-covariance matrix:

Proposition 2. *Suppose $\tilde{w}_t^{\bigstar} = \boldsymbol{\beta}_0$ with $\boldsymbol{X}_t = I_{N-1}$, then $\boldsymbol{\beta}_0 = \frac{Z\,\mathrm{E}[\Sigma_t]^{-1}\iota}{\iota'\,\mathrm{E}[\Sigma_t]^{-1}\iota}$ solves Eq. (3.12).*

Intuitively, this result suggests that if the true $w_t^{\bigstar}$ is a constant plus a random innovation term, then the optimal GMVP weight vector is simply the plug-in vector computed from the unconditional expectation of Σ_t. In this simple setting, Σ_t admits the following representation for all t:

$$\Omega_t = \mathrm{E}_{t-1}[\Sigma_t] = \mathrm{E}[\Sigma_t]. \tag{3.16}$$

Importantly, this suggests that $\Sigma_t - \mathrm{E}[\Sigma_t]$ must be independent of $\mathcal{F}_{t-1}$ (i.e., it is a matrix-valued MDS), so if Σ_t presents a persistent autoregressive structure, then a constant model for $\tilde{w}_t^{\bigstar}$ must be misspecified. Interestingly, setting $\tilde{w}_t^{\bigstar} = \boldsymbol{\beta}_0 + \tilde{\eta}_t$ does not change the value of $\boldsymbol{\beta}_0$ as $\tilde{\eta}_t$ does not enter into $\boldsymbol{\beta}_0$. However, Eq. (3.16) no longer holds due to the time-varying Ω_t implied by a time-varying $\tilde{\eta}_t$.

We now discuss how to estimate $\boldsymbol{\beta}_0$ from the observed sequence $\{\Sigma_t\}_{t=1:T}$. As all the quantities in the expectations of Eq. (3.14) are observed, Theorem 1 provides a direct method-of-moment estimator for $\boldsymbol{\beta}_0$. Based on $\{\Sigma_t\}_{t=1:T}$ with $T > b(K)$, we propose the following estimator:

$$\hat{\boldsymbol{\beta}}_T = \left(\sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t\right)^{-1} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \tilde{w}_t^{\star}. \tag{3.17}$$

Note that the condition $T > b(K)$ ensures that the first sum is invertible[4]. It is easy to show that $\hat{\boldsymbol{\beta}}_T$ solves the finite-sample version of the optimization problem in Theorem 1:

$$\min_{\boldsymbol{\beta}\in\mathcal{B}} \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t(\boldsymbol{\beta})' \tilde{\Sigma}_t \tilde{u}_t(\boldsymbol{\beta}) \Leftrightarrow \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \tilde{u}_t(\boldsymbol{\beta}) = 0, \tag{3.18}$$

which is also the solution to the finite-sample minimum variance problem discussed in Eq. (2.10). Also, one should realize that the estimator $\hat{\boldsymbol{\beta}}_T$ only requires a linear specification in the weights, but does not rely on any specific assumption about the distribution of either r_t or Σ_t. Therefore, the LWE method is *semi-parametric* in nature.

We now discuss some properties of the estimator $\hat{\boldsymbol{\beta}}_T$. From Eq. (3.17) and using the decomposition $\tilde{w}_t^{\star} = \boldsymbol{X}_t \boldsymbol{\beta}_0 + \tilde{\eta}_t + \tilde{\epsilon}_t$, we can directly obtain:

$$\mathrm{E}[\hat{\boldsymbol{\beta}}_T] - \boldsymbol{\beta}_0 = \mathrm{E}\left[\left(\sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t\right)^{-1} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t (\tilde{\eta}_t + \tilde{\epsilon}_t)\right]. \tag{3.19}$$

As $\tilde{\epsilon}_t$ in general has a non-zero mean and depends on $\tilde{\Sigma}_t$, the above bias term is non-zero and does not simplify further. Nevertheless, as $T \to \infty$, Eq. (3.9) in Assumption 3 implies that:

$$\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 = \left(\sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t\right)^{-1} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t (\tilde{\eta}_t + \tilde{\epsilon}_t) \xrightarrow{p} \mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t]^{-1} \mathbf{0}_{b(K)\times 1} = \mathbf{0}_{b(K)\times 1}, \tag{3.20}$$

where the last equality follows from Proposition 1. This shows that the proposed estimator is consis-

[4] One should verify that $\boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t$ is of rank $N-1$ by Assumption 3, so we must sum over more than $b(K)$ terms to ensure that $\sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t$ is of full rank hence invertible.

tent, which also explains why the same property does not hold for an OLS estimator. The asymptotic normality of $\hat{\boldsymbol{\beta}}_T$ also holds from Eq. (3.9) and a direct application of the continuous mapping theorem:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t]^{-1} \boldsymbol{C}_\infty \, \mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t]^{-1}), \tag{3.21}$$

Note that a feasible version of the above CLT is available since both $\mathrm{E}[\boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t]^{-1}$ and $\boldsymbol{C}_\infty$ can be consistently estimated from observed data (under further assumptions about the existence of higher moments). In principle, this allows us to test whether certain (set of) elements from $\boldsymbol{\beta}_0$ are zero, which can be used to select the predictor vectors. It also allows us to construct confidence bounds for the estimated GMVP weight vector by a straightforward application of the delta method. Nevertheless, it is worth noting that such asymptotic result is only valid when $\hat{\beta}_0$ lies in the interior of the parameter space, which may not be the case when one imposes certain restrictions to the weight vector (see the discussion in the next section). In practice, one should rely on bootstrap-based methods to compute standard errors and confidence bounds of parameter estimates under such constraints, which is further discussed in Brandt and Santa-Clara (2006).

3.1 Out-of-Sample Evaluation of GMVP Strategies

Instead of statistical inference, practitioners are more interested in evaluating the out-of-sample performance of a given GMVP strategy. To this end, one can estimate $\hat{\boldsymbol{\beta}}_T$ based on a sample of $\{\Sigma_t, \boldsymbol{X}_t\}_{t=1:T}$. Since $\boldsymbol{X}_{T+1}$ is $\mathcal{F}_T$-adapted, the predicted GMVP weight vector for time $T+1$ is simply given by:

$$w_{T+1}(\hat{\boldsymbol{\beta}}_T) = \boldsymbol{e}_N + \tilde{Z}\boldsymbol{X}_{T+1}\hat{\boldsymbol{\beta}}_T. \tag{3.22}$$

Using a standard rolling window or extending window design, one can obtain predicted GMVP weight vectors $\{w_t(\hat{\boldsymbol{\beta}}_{t-1})\}_{t\in[T+1:T']}$ for the (pseudo) out-of-sample period $[T+1 : T']$. The out-of-sample performance of a GMVP strategy $\{w_t\}_{t\in[T+1:T']}$ can be evaluated by the average realized portfolio variance (ARPV) or the sample portfolio variance (SPV) over the out-of-sample period, defined as:

$$\mathrm{ARPV} = \frac{1}{T'-T}\sum_{t=T+1}^{T'} w_t'\Sigma_t w_t, \quad \mathrm{SPV} = \frac{1}{T'-T-1}\sum_{t=T+1}^{T'} w_t'(r_t r_t' - \bar{r}\bar{r}')w_t, \tag{3.23}$$

where $\bar{r} = (T'-T)^{-1}\sum_{t=T+1}^{T'} r_t$ is the out-of-sample sample mean of the return. These statistics serve as point estimates for the comparison of different GMVP strategies.

To formally test whether one strategy outperforms another, we consider a modified Diebold-Mariano (DM) test (Diebold and Mariano, 1995; Harvey et al., 1997) tailored to the context of out-of-sample GMVP realized variance comparison. Let $\{w_t^{(1)}\}_{t\in[T+1:T']}$ and $\{w_t^{(2)}\}_{t\in[T+1:T']}$ denote two GMVP strategies, where $\{w_t^{(1)}\}_{t\in[T+1:T']}$ is the benchmark model that is believed to have a superior performance. We construct the portfolio loss at time t as:

$$L_t(w) = w'\Sigma_t w - (w_t^{\star})'\Sigma_t w_t^{\star} = w'\Sigma_t w - \frac{1}{\iota'\Sigma_t\iota} \geq 0. \tag{3.24}$$

Intuitively, $L_t(w)$ measures the distance between the realized portfolio variance with weight w to the *ex post* GMVP variance, which is strictly non-negative. For a given GMVP strategy $\{w_t\}_{t\in[T+1:T']}$, its out-of-sample performance can be summarized by the mean loss $\mathrm{E}[L_t(w_t)]$, and it is easily seen that $\mathrm{E}[L_t(w_t^{\bigstar})]$ minimizes this mean loss by construction. As a pairwise prediction accuracy comparison test, the null hypothesis of the DM test is $\mathrm{E}[L_t(w_t^{(1)})] = \mathrm{E}[L_t(w_t^{(2)})]$ against the alternative hypothesis $\mathrm{E}[L_t(w_t^{(1)})] < \mathrm{E}[L_t(w_t^{(2)})]$. The test can be performed by computing the loss differentials for the out-of-sample period:

$$d_t = L_t(w_t^{(1)}) - L_t(w_t^{(2)}), \quad t \in [T+1:T'], \tag{3.25}$$

which should satisfy $\mathrm{E}[d_t] = 0$ under the null hypothesis. With the time series $\{d_t\}_{t\in[T+1:T']}$, the DM test can be performed by closely following the steps in Harvey et al. (1997), which is essentially a heteroscedasticity and autocorrelation-robust t-test for the out-of-sample mean of d_t.

3.2 Generalizations

3.2.1 Investment Constraints

Eq. (3.22) provides an unrestricted weight forecast based on the proposed linear model, where $\boldsymbol{X}_{T+1}\hat{\boldsymbol{\beta}}_T$ can in principle take any value in $\tilde{\mathcal{W}}$, including those that are practically infeasible. In practice, one might want to restrict the weight vector to satisfy certain investment constraints. We discuss two generalizations to the LWE method which aim to ensure that the weights are empirically investable.

We first consider a penalized version of the LWE method by augmenting the minimization problem in Eq. (3.18) with a ridge-type penalty function, i.e.:

$$\min_{\boldsymbol{\beta}\in\mathcal{B}} \frac{1}{T}\sum_{t=1}^{T} \tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta}) + \lambda f(w_{T+1}(\boldsymbol{\beta})), \tag{3.26}$$

where $\lambda > 0$ is a tuning parameter, and $f : \mathcal{W} \mapsto \mathbb{R}$ is a general quadratic function of some $w \in \mathcal{W}$:

$$f(w) := w'\boldsymbol{A}w + 2\boldsymbol{a}'w + C \tag{3.27}$$

for some positive definite N-by-N matrix $\boldsymbol{A}$, some vector $\boldsymbol{a} \in \mathbb{R}^N$, and some arbitrary constant C that does not depend on w. Intuitively, $f(w)$ should be chosen as an 'un-investability' measure of the weight vector w, so that Eq. (3.26) is jointly minimising a weighted average of both portfolio realized variance and the un-investability measure. The quadratic form of $f(w)$ is important, as it ensures that

Eq. (3.26) can still be solved in closed form with a unique minimum. This rules out other popular penalty functions, such as the LASSO of Tibshirani (1994), the SCAD of Fan and Li (2001), etc., which requires numerical algorithms to compute a solution and is thus not pursued here.

We now characterize the closed-form solution to Eq. (3.26) and its behaviour as $\lambda \to \infty$:

Proposition 3. *For any $\lambda \geq 0$ and $T > b(K)$, the unique minimizer of problem Eq. (3.26) is:*

$$\hat{\boldsymbol{\beta}}_{T,\lambda} = \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_t' \tilde{\Sigma}_t \boldsymbol{X}_t + \lambda \boldsymbol{X}_{T+1}' \tilde{\boldsymbol{A}} \boldsymbol{X}_{T+1} \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_t \tilde{\Sigma}_t \tilde{w}_t^{\star} - \lambda \boldsymbol{X}_{T+1}' \tilde{Z}' (\boldsymbol{A} \boldsymbol{e}_N + \boldsymbol{a}) \right), \quad (3.28)$$

where $\tilde{\boldsymbol{A}} = \tilde{Z}' \boldsymbol{A} \tilde{Z}$. It holds that $\lim_{\lambda \to 0} \hat{\boldsymbol{\beta}}_{T,\lambda} \equiv \hat{\boldsymbol{\beta}}_T$. As $\lambda \to \infty$, the solution $\hat{\boldsymbol{\beta}}_{T,\infty}$ is not necessarily unique, but all solutions must satisfy:

$$w^* = w_{T+1}(\hat{\boldsymbol{\beta}}_{T,\infty}), \quad (3.29)$$

where $w^* \in \mathcal{W}$ is the unique minimizer of $f(w)$:

$$w^* := \frac{\boldsymbol{A}^{-1}\iota}{\iota' \boldsymbol{A}^{-1} \iota} + \frac{\iota' \boldsymbol{A}^{-1} \boldsymbol{a}}{\iota' \boldsymbol{A}^{-1} \iota} \boldsymbol{A}^{-1} \iota - \boldsymbol{A}^{-1} \boldsymbol{a}. \quad (3.30)$$

Proposition 3 presents the closed-form solution of the penalized LWE estimator for a general $f(w)$ with some $\lambda \geq 0$, as well as the limiting characterization of the solution when $\lambda \to \infty$. Importantly, when $\lambda \to \infty$, the weight estimator w^* minimizes the given un-investability criteria, so the choice of λ here strikes a balance between minimizing the portfolio variance and the improving the investability of the portfolio weights.

We now give two concrete examples of $f(w)$. First, one can interpret the investability of the portfolio weights as the associated transaction cost, which is proportional to the total holdings in each asset in our day trading setup without overnight holdings. This is measured by the portfolio exposure, i.e., $||w||_1$, where $|| \cdot ||_p$ is the ℓ_p norm of a vector. Clearly, one cannot directly set $f(w) = ||w||_1$ or $||w||_1^2$, as the ℓ_1-norm is not quadratic in w. Nevertheless, by the well-known norm inequalities:

$$||w||_2^2 \leq ||w||_1^2 \leq N ||w||_2^2. \quad (3.31)$$

As a result, by setting $f(w) = N||w||_2^2$ with $\boldsymbol{A} = N\boldsymbol{I}_N$, $\boldsymbol{a} = \boldsymbol{0}_{N \times 1}$ and $C = 0$, the penalized LWE method shrinks the upper bound of $||w||_1^2$, which in turn mitigates the portfolio exposure, hence the transaction cost. It is worth noting that, as $w^* = \iota/N \in \mathcal{W}$ minimizes $f(w)$ in this case, the solution $\hat{\boldsymbol{\beta}}_{T,\infty}$ is any $\boldsymbol{\beta}$ that produces the equal weight vector.

As another example, a portfolio manager may want to minimize the portfolio variance while tracking some target weight vector, say $\tau_T \in \mathcal{W}$. For instance, one can set τ_T to be the portfolio weight of the previous period to reduce portfolio turnover, or set τ_T to be the portfolio holdings within (resp. without) a sector to increase (resp. decrease) the holdings in that sector. In this case, one can set $f(w) = ||w - \tau_T||_2^2$, with $\boldsymbol{A} = \boldsymbol{I}_N$, $\boldsymbol{a} = -\tau_T$ and $C = ||\tau_T||_2^2$. Then as $\lambda \to \infty$, any solution $\hat{\boldsymbol{\beta}}_{T,\infty}$ of Eq. (3.26) should produce the target weight vector $\tau_T = w_{T+1}(\hat{\boldsymbol{\beta}}_{T,\infty})$ by construction.

Second, we shall show that the LWE method can be augmented with linear inequality constraints in weights, i.e., a constrained LWE method. Unlike the penalized LWE method which provides a smooth transition from the unconstrained weight $w_{T+1}(\hat{\boldsymbol{\beta}}_T)$ to a target weight vector $w_{T+1}(\hat{\boldsymbol{\beta}}_{T,\infty})$, the constrained LWE method ensures that the portfolio weights must satisfy constraints of the form:

$$\boldsymbol{B} w_{T+1}(\hat{\boldsymbol{\beta}}_T) \leq \boldsymbol{b}, \tag{3.32}$$

where $\boldsymbol{B}_{k\times N}$ and $\boldsymbol{b}_{k\times 1}$ specifies $k \leq N$ linear inequalities that the weights must satisfy. For example, the choices $\boldsymbol{B} = -\boldsymbol{I}_N$ and $\boldsymbol{b} = \boldsymbol{0}_{N\times 1}$ correspond to a no-short-selling restriction, and a portfolio exposure constraint in the spirit of Jagannathan and Ma (2003) can also be implemented using this device. As Eq. (3.32) is just a linear inequality in terms of $\boldsymbol{\beta}$, in practice one can impose Eq. (3.32) when solving Eq. (3.18), and the resulting optimization problem is still strictly convex (hence a unique solution exists), but the solution can only be computed numerically, albeit very efficiently using standard quadratic programming algorithms.

3.2.2 A Nonlinear Weight Model for GMVP Weights

As a nonlinear generalization of the LWE method, one can consider the following non-linear model for the *ex ante* GMVP weight vector:

$$\tilde{w}_t^{\star} = f(\boldsymbol{\beta}_0; \boldsymbol{X}_t) + \tilde{\eta}_t, \tag{3.33}$$

where $f(\boldsymbol{\beta}; \boldsymbol{X}_t)$ is a twice-differentiable function of $\boldsymbol{\beta}$ given the $\mathcal{F}_t$-predictable explanatory variables $\boldsymbol{X}_t$, and $\tilde{\eta}_t$ satisfies the following identification assumption:

$$\mathrm{E}[\nabla_{\boldsymbol{\beta}} f(\boldsymbol{\beta}_0; \boldsymbol{X}_t)' \tilde{\Sigma}_t \tilde{\eta}_t] = 0, \tag{3.34}$$

where $\nabla_{\boldsymbol{\beta}} f(\boldsymbol{\beta}_0; \boldsymbol{X}_t)$ is the $(N-1)$-by-$b(K)$ Jacobian matrix of $f(\boldsymbol{\beta}; \boldsymbol{X}_t)$ evaluated at $\boldsymbol{\beta}_0$.

The non-linear framework incorporates most of the existing approaches for the GMVP weight selection problem, and our linear model is obviously a special case. For example, $f(\boldsymbol{\beta}_0; \boldsymbol{X}_t)$ can be a VARMA-type model for the observed GMVP weights similar to those used in Palandri (2022), where $\boldsymbol{X}_t$ are the past *ex post* and *ex ante* GMVP weight vectors and $\boldsymbol{\beta}_0$ summarizes their contributions to the model. One can also consider $f(\boldsymbol{\beta}_0; \boldsymbol{X}_t)$ as the plug-in weight computed from a predictive model of Ω_t, in which case $\boldsymbol{X}_t$ is a collection of past RC matrices and $\boldsymbol{\beta}_0$ becomes the model parameter for

Ω_t. Importantly, we emphasis that Eq. (3.34) is the correct moment condition for the estimation of these model parameters that solves the unconditional minimum variance problem in Eq. (2.10), i.e., it ensures that the parameter estimates minimize the unconditional portfolio variance, which is not achieved by other heuristic approaches such as least-square or likelihood-based inference attempting to predict the weight vector or the realized covariance matrix.

Under the non-linear framework, a result similar to Theorem 1 holds, which gives the following minimization problem and moment condition for the estimation of $\boldsymbol{\beta}_0$:

$$\min_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}[\tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta})] \Rightarrow \mathrm{E}[\nabla_{\boldsymbol{\beta}} f(\boldsymbol{\beta}_0;\boldsymbol{X}_t)'\tilde{\Sigma}_t\tilde{u}(\boldsymbol{\beta}_0)] = 0. \tag{3.35}$$

As this falls into the GMM framework with a non-linear moment condition, the asymptotic results for the parameter estimates can be established following e.g., Hall (1996). However, the non-linear optimization problem is in general not convex, thus a $\boldsymbol{\beta}_0$ that satisfies the moment condition above is not necessarily the global minimizer of the above optimization problem. This also creates problems for the parameter estimation in finite sample, as the parameters need to be estimated numerically, and convergence to the global minima is not guaranteed, especially when N or the number of parameters are large. Therefore, despite the additional flexibility enjoyed by the non-linear framework, we focus on the linear framework in this paper due to these theoretical and computational drawbacks.

3.2.3 Linear Weight Estimation for The Mean-Variance Portfolio

Finally, our approach can be generalized to model the efficient portfolio (mean-variance portfolio of N risky assets) or the mean-variance portfolio (maximum Sharpe ratio portfolio of N risky assets and a risk-free asset). For example, given the mean-variance conditional utility function:

$$U_t(w) = \mathrm{E}_{t-1}[w'r_t] - \frac{\gamma}{2}\mathrm{V}_{t-1}[w'r_t] = w'\mu_t - \frac{\gamma}{2}w'\Omega_t w, \tag{3.36}$$

where γ is the Arrow-Pratt risk-aversion coefficient. For the mean-variance portfolio allocation problem, the investor maximizes $U_t(w)$ for $w \in \mathbb{R}^N$ at time $t-1$. Notice that w does not have to sum up to 1 here when a risk-free asset is available. The *ex ante* and the realized portfolio optimal weight vectors are, respectively:

$$w_t^{\bullet} = \frac{1}{\gamma}\Omega_t^{-1}\mu_t, \quad w_t^{\circ} = \frac{1}{\gamma}\Sigma_t^{-1}r_t. \tag{3.37}$$

Write $\epsilon_t = w_t^{\bullet} - w_t^{\circ}$ and notice that a dimension reduction is not required in this case, one immediately sees that:

$$\mathrm{E}_{t-1}[\Sigma_t\epsilon_t] = \mathbf{0}_{N\times 1}, \tag{3.38}$$

which is the moment condition similar to Proposition 1 for the GMVP problem. In the spirit of the LWE estimator for GMVP, we parametrize $w_t^\bullet$ using a linear model:

$$w_t^\bullet = \boldsymbol{Y}_t \boldsymbol{\beta}_0 + \epsilon_t, \tag{3.39}$$

for some N-by-$b(K)$ matrices of explanatory weight vectors $\{\boldsymbol{Y}_t\}_{t=1,2,...}$ which satisfy $\mathrm{E}[\boldsymbol{Y}_t \Sigma_t \epsilon_t] = \mathbf{0}_{b(K)\times 1}$. Similar to the proof of Theorem 1, one can show that $\boldsymbol{\beta}_0 = \mathrm{E}[\boldsymbol{Y}_t' \Sigma_t \boldsymbol{Y}_t]^{-1}\,\mathrm{E}[\boldsymbol{Y}_t' \Sigma_t w_t^\circ]$ solves the following unconditional utility maximization problem, independent of the choice of γ:

$$\max_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}[U_t(\boldsymbol{Y}_t\boldsymbol{\beta} + \epsilon_t)] \Leftrightarrow \min_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}[\epsilon_t' \Sigma_t \epsilon_t]. \tag{3.40}$$

An estimator of $\boldsymbol{\beta}_0$ can therefore be constructed as:

$$\hat{\boldsymbol{\beta}}_T = \frac{1}{\gamma}\Big(\sum_{t=1}^{T} \boldsymbol{Y}_t' \Sigma_t \boldsymbol{Y}_t\Big)^{-1} \sum_{t=1}^{T} \Sigma_t^{-1} r_t. \tag{3.41}$$

Interestingly, this also allows us to construct the one-step-ahead tangency portfolio weight that does not depend on γ:

$$w_T^{tan} = \frac{\boldsymbol{Y}_{T+1}\hat{\boldsymbol{\beta}}_T}{\iota' \boldsymbol{Y}_{T+1}\hat{\boldsymbol{\beta}}_T} \in \mathcal{W}. \tag{3.42}$$

The main complication here is the choice of $\boldsymbol{Y}_t$ to explain the *ex ante* portfolio weights. Intuitively, $w_t^\bullet \propto \Omega_t^{-1}\mu_t$ is proportional to the variance-normalized conditional mean of each asset, whose dynamic properties are fundamentally different from the GMVP weights $w_t^\bigstar$ that only depend on Ω_t. Variations in the conditional mean of a cross-section of assets are typically explained by various asset pricing factors, or firm-level characteristics following Brandt et al. (2009). As the potential choice of $\boldsymbol{Y}_t$ is non-trivial and deserves individual investigation, we shall discuss the dynamic mean-variance portfolio allocation problem in a subsequent paper.

4 Empirical Evidence

In this section we evaluate the performance of our LWE approach and compare it with other well-known competitors using daily financial data. To conduct our analysis, we select the most liquid 250 stocks from the S&P 500 index in terms of the average number of transactions per day. We gather daily data covering the period from January 2015 to December 2022. To construct the realized covariance matrices, we apply the flat-top realized kernel of Varneskov (2016) to estimate the daily RC matrices using the tick-by-tick data of the 250 stocks.[5] This results in a time series comprising 1756

[5] See Section 5.1 and the Online Appendix A of Li et al. (2022) for the data cleaning procedures and the implementation details of the flat-top realized kernel.

observations. Descriptive statistics for all the stocks used can be found in Table A.1 in the Appendix. Notably, daily returns exhibit the expected characteristics of left skewness and leptokurticity. We aim at constructing daily forecasts of the GMVP weights and compare our proposed estimator with the competing approaches in terms of both economic and statistical performances.

As unrestricted portfolio weight estimates might result in practically infeasible positions such as large short-selling or excessive holdings of certain assets, we impose an exposure constraint on the absolute value of portfolio weights to ensure that the weight estimates are empirically investable. We consider three progressively more restrictive investment constraints to mimic realistic portfolio weights. Specifically, let $w^{(n)}$ denote the weight applied to the n-th asset, we shall assume for all $1 \leq n \leq N$: (1) $|w^{(n)}| \leq \max(5\%, 5/N)$; (2) $|w^{(n)}| \leq \max(3\%, 3/N)$; (3) $w^{(n)} \in [0, \max(3\%, 3/N)]$. In detail, constraints (1) and (2) restrict the total exposure to one asset to be 5% or 3% with short-selling, while constraint (3) further prohibits short-selling on top of the 3% exposure constraint.

We now discuss how to impose these constraints to an estimated weight vector $\hat{w}$ of some GMVP strategy. To this end, let us denote $\mathcal{W}^r$ as the collection of all weights that satisfy a particular constraint, and we impose the exposure constraint to $\hat{w}$ by finding some $\hat{w}^r \in \mathcal{W}^r$ that solves the following problem:

$$\hat{w}^r = \underset{w \in \mathcal{W}^r}{\arg\min} (w - \hat{w})'(w - \hat{w}). \tag{4.1}$$

Intuitively, $\hat{w}^r$ is the closest element of $\mathcal{W}^r$ to $\hat{w}$ in the Euclidean distance that satisfies the exposure constraint. As the problem is quadratic and $\mathcal{W}^r$ is convex for our three constraints, the solution $\hat{w}^r$ is unique and can be solved by standard quadratic programming algorithms very efficiently. It is worth noting that, for the LWE method, we can directly incorporate these exposure constraint by utilising Eq. (3.32). However, as this is in general not applicable to the competing methods considered here, we shall apply Eq. (4.1) to all the weight estimates to eliminate the potential impact of different exposure constraints.

We proceed to discuss the candidate GMVP strategies in our empirical study. For our LWE estimator, we consider the diagonal specification of Eq. (3.5) and the scalar specification of Eq. (3.7). The predictor weights in the linear model consist of a history of realized GMVP portfolio weights at daily, weekly and monthly horizon, which is inspired by the classic HAR model of Corsi (2009) and the persistent realized portfolio weights documented in Cipollini et al. (2021):

$$\tilde{X}_{t+1}^{(j)} = \frac{1}{j} \sum_{i=0}^{j-1} \tilde{w}_{t-i}^{\star}, \quad j \in \{1, 5, 22\}.$$

Intuitively, $\tilde{X}_{t+1}^{(1)}$, $\tilde{X}_{t+1}^{(5)}$ and $\tilde{X}_{t+1}^{(22)}$ are the corresponding dimension-reduced averaged realized GMVP weights available on time t at the daily, weekly and monthly horizon. Consequently, we have $K = 3$ explanatory variables in the LWE model, which implies $4N - 1$ and $N + 2$ numbers of parameters for the diagonal and the scalar specifications, respectively.

We compare our estimator to several popular estimators in the literature: (1) The static model denoted as 'sample', where the sample covariance estimator obtained from $T = 1000$ in-sample daily return observations is plugged into Eq. (2.7); (2) The random walk (denoted as RW) model with $\Omega_t \equiv \Sigma_t$ is assumed for the dynamics of realized covariances and therefore the Σ_{t-1} is used in Eq. (2.7) to construct the forecast of the portfolio weights. (3) and (4): We consider the dynamic RC forecast model by Bollerslev et al. (2018) with both constant and dynamic conditional correlation structures, denoted as HAR CCC and HAR DCC, respectively. The model is estimated on the history of RC matrices, where a one-step ahead forecast of the covariance matrix is defined as $\hat{\Omega}_{t+1|t} = \hat{D}_{t+1|t}\hat{R}_{t+1|t}\hat{D}_{t+1|t}$. The matrix of conditional volatilities, $\hat{D}_{t+1|t}$, is forecasted with univariate HAR models and the conditional correlations $\hat{R}_{t+1|t}$ are either modelled as sample covariances (CCC) or as a vector HAR (DCC). (5) Following Golosnoy et al. (2019) we implement the exponential smoothing of realized portfolio weights with the smoothing parameter $\lambda = 0.94$. This estimator is denoted as ES and resembles the Risk Metrics estimator, which is applied to realized portfolio weights instead of conditional volatility. (6) We consider the dynamic conditional weights (DCW) estimator by Cipollini et al. (2021). Similar to our approach, the DCW estimator uses the history of realized portfolio weights to build a linear VARMA model. We implemented the scalar and the diagonal VARMA(1,1) models[6] as suggested by the authors. It is worth noting that among all the methods, the LWE method is computationally much faster than the dynamic forecasting models such as HAR CCC/DCC and the DCW methods due to its closed-form solution.

To evaluate and interpret the performance of the GMVP strategies, in Table 4.1 we present the annualized average realized portfolio volatility in percentage, defined as $\sqrt{\text{ARPV} \cdot 252} \cdot 100\%$ with ARPV defined in Eq. (3.23), for all GMVP strategies under various investment constraints. The table clearly shows that, regardless of the choice of investment constraints, the scalar or diagonal LWE methods always have the smallest realized portfolio volatilities among all the competing strategies. The realized portfolio volatilities of the LWE strategies in general decrease as N becomes larger and as the investment constraint becomes less stringent, which are consistent with our expectation. The performances of the scalar and the diagonal LWE are largely similar, and the simple scalar specification appears to be more favourable under more stringent investment restrictions.

For the competing strategies, the sample covariance, the random walk and the HAR CCC strategies have overall the worst performance. Interestingly, the performances of these estimators deteriorate as N increases, suggesting instability in the estimated weights at large dimension, which can be improved by imposing investment constraints. For the remaining competing strategies, the DCW scalar and the exponential smoothing weight methods have comparably better performance, which is consis-

[6]Note that for the DCW method, the estimated weights need to be normalized to sum up to one.

Table 4.1: Annualized out-of-sample average realized portfolio volatility

N	LWE scalar	LWE diagonal	sample	RW	ES	HAR CCC	HAR DCC	DCW scalar	DCW diagonal
Panel 1: Unrestricted portfolio weights									
10	14.38	**14.37**	16.09***	14.63***	14.72**	15.07***	14.92**	14.40	14.41
20	*13.47*	**13.45**	15.83***	13.97***	13.86***	14.56***	14.08***	13.51**	13.52**
30	*12.88*	**12.85**	15.07***	13.77***	13.22***	14.41***	13.59***	12.95*	12.94*
40	**12.64**	*12.64*	14.99***	13.69***	13.07***	14.31***	13.44***	12.78**	12.74*
50	*12.57*	**12.54**	15.02***	13.89***	13.01***	14.50***	13.38***	12.74**	12.68**
60	*12.57*	**12.50**	14.99***	14.25***	12.98***	14.86***	13.44***	12.76***	12.69**
70	*12.59*	**12.54**	15.12***	14.77***	12.98***	15.39***	13.66***	12.79***	12.8***
80	*12.38*	**12.32**	15.17***	14.99***	12.88***	15.58***	13.46***	12.67***	12.71***
90	*12.28*	**12.23**	15.26***	15.50***	12.76***	15.97***	13.20***	12.58***	12.62***
100	*12.30**	**12.21**	15.24***	16.10***	12.81***	16.56***	13.15***	12.64***	12.69***
150	*12.37*	**12.28**	15.57***	22.86***	13.32***	21.68***	13.08***	13.12***	13.43***
200	**12.23**	*12.33*	15.91***	25.92***	13.38**	25.37***	14.70*	13.22***	13.53***
250	**11.58**	*11.72*	15.90***	18.92***	12.74**	20.45***	14.60*	12.84***	13.25***
Panel 2: $\forall n \leq N, \lvert w^{(n)} \rvert \leq \max(5\%, 5/N)$									
10	*14.39*	**14.37**	16.09***	14.63***	14.72**	14.86***	14.70**	14.41*	14.41*
20	**13.61**	*13.61*	15.78***	14.06***	13.89***	14.37***	13.93***	13.65*	13.67**
30	*13.06*	**13.05**	15.03***	13.72***	13.29***	14.02***	13.39***	13.13**	13.15**
40	**12.88**	*12.91*	14.91***	13.58***	13.18***	13.94***	13.32***	13.00**	13.02**
50	**12.85**	*12.85*	14.85***	13.67***	13.16***	14.03***	13.3***	12.99***	13.01**
60	*12.9*	**12.90**	14.80***	13.87***	13.16***	14.22***	13.37***	13.04***	13.05***
70	**12.93**	*12.94*	14.81***	14.11***	13.17***	14.37***	13.43***	13.08***	13.15***
80	*12.75*	**12.75**	14.77***	14.05***	13.07***	14.36***	13.26***	12.96***	13.04***
90	**12.61**	*12.65*	14.7***	14.12***	12.93***	14.38***	13.12***	12.85***	12.94***
100	**12.65**	*12.66*	14.62***	14.28***	12.95***	14.59***	13.13***	12.90***	13.00***
150	**12.5**	*12.60*	15.28***	15.91***	13.25***	16.02***	13.01***	13.18***	13.56***
200	**12.26**	*12.34*	15.49***	17.08***	13.29***	17.27***	13.05***	13.22***	13.51***
250	**11.59**	*11.77*	15.53***	16.30***	12.69***	16.61***	12.90***	12.85***	13.25***
Panel 3: $\forall n \leq N, \lvert w^{(n)} \rvert \leq \max(3\%, 3/N)$									
10	**14.52**	*14.53*	16.03***	14.73***	14.76***	14.84***	14.70***	14.55	14.57*
20	*13.88*	**13.88**	15.69***	14.24***	14.06***	14.47***	14.10***	13.92**	13.92*
30	*13.37*	**13.34**	14.88***	13.83***	13.53***	14.06***	13.59***	13.43***	13.44***
40	**13.20**	*13.22*	14.75***	13.74***	13.41***	14.01***	13.54***	13.30**	13.31**
50	**13.20**	*13.21*	14.70***	13.82***	13.40***	14.09***	13.57***	13.30**	13.31**
60	**13.28**	*13.30*	14.70***	14.00***	13.44***	14.23***	13.67***	13.38**	13.36**
70	**13.35**	*13.37*	14.71***	14.21***	13.48***	14.40***	13.78***	13.44**	13.46***
80	**13.20**	*13.23*	14.61***	14.12***	13.42***	14.34***	13.65***	13.36***	13.38***
90	**13.06**	*13.12*	14.54***	14.11***	13.27***	14.29***	13.53***	13.23***	13.26***
100	**13.11**	*13.16*	14.46***	14.23***	13.31***	14.44***	13.57***	13.29***	13.33***
150	**12.81**	*13.00*	15.21***	15.05***	13.33***	15.21***	13.39***	13.35***	13.73***
200	**12.45**	*12.61*	15.13***	15.55***	13.21***	15.9***	13.37***	13.27***	13.55***
250	**11.78**	*12.03*	15.33***	15.09***	12.72***	15.55***	13.23***	12.88***	13.33***
Panel 4: $\forall n \leq N, w^{(n)} \in [0, \max(3\%, 3/N)]$									
10	14.72	*14.72*	15.88***	14.89***	14.90**	14.96***	14.88**	14.74*	**14.71**
20	**14.04**	*14.04*	15.28***	14.26***	14.16***	14.43***	14.22***	14.05	14.05
30	*13.63*	**13.61**	14.52***	13.89***	13.71***	14.05***	13.74***	13.66**	13.65**
40	**13.51**	*13.52*	14.32***	13.79***	13.63***	13.97***	13.69***	13.56**	13.55*
50	**13.58**	*13.60*	14.32***	13.89***	13.68***	14.05***	13.74***	13.64**	13.62
60	**13.65**	*13.67*	14.35***	14.00***	13.74***	14.19***	13.83***	13.71*	13.68
70	**13.76**	*13.78*	14.46***	14.16***	13.84***	14.30***	13.97***	13.84**	13.82**
80	*13.68*	**13.67**	14.45***	14.13***	13.75***	14.31***	13.89***	13.75***	13.74***
90	*13.62*	**13.59**	14.39***	14.10***	13.67***	14.28***	13.78***	13.69***	13.70***
100	*13.67***	**13.63**	14.39***	14.21***	13.69***	14.36***	13.8***	13.73***	13.75***
150	*13.62*	**13.59**	14.52***	14.65***	13.89***	14.86***	13.76***	13.90***	14.05***
200	*13.75***	**13.63**	14.71***	15.11***	14.16**	15.37***	13.87**	14.18***	14.26***
250	*13.64**	**13.59**	14.83***	15.36***	14.04***	15.66***	13.92***	14.12***	14.27***

Numbers in the table correspond to the annualized average realized GMVP portfolio volatility in percentage. For each GMVP strategy and portfolio size N, the portfolio standard deviations are computed over an evaluation horizon $H = 756$ observations and an in-sample estimation window length $T = 1000$. Numbers in **bold** (*italic*) correspond to the smallest (second smallest) value for a given portfolio size N. The stars next to numbers correspond to the p-values of the Diebold-Mariano test discussed in Section 3.1, where rejection of the test indicates that the corresponding GMVP strategy produces larger realized portfolio variance that that of the benchmark LWE diagonal model. *** - $p < 0.01$, ** - $p < 0.05$ and * - $p < 0.1$.

tent with the findings in Golosnoy et al. (2019), Cipollini et al. (2021). Nevertheless, they still produce significantly larger realized portfolio volatilities than our LWE methods, regardless of the choice of N and investment constraints. Also, all competing strategies have somewhat inflated portfolio variances for $N > 100$ compared to $N \leq 100$, suggesting that these methods may not fully exploit the diversification effect when the dimension is large, which is clearly not the case for the LWE method.

Table 4.2 presents a commonly used but less precise measure for comparing the performance of different GMVP strategies: the standard deviation of the out-of-sample portfolio returns computed as $\sqrt{SPV \cdot 252} \cdot 100\%$, where SPV is defined in equation Eq. (3.23). The results are less clear cut than those in Table 4.1 and depend on the choice of investment constraint, but the LWE estimators still deliver consistently small portfolio standard deviations, especially when N is large and when the investment constraint is imposed. Here we do not expect the LWE method to consistently beat all the competitors, because the LWE method does not minimize the SPV of the portfolio by design. As the realized portfolio standard deviation based-on high-frequency prices is a more accurate measure of portfolio variance than the sample standard deviation using daily returns, we argue that Table 4.1 shows a more accurate description of the relative performances of all GMVP strategies.

To understand when the LWE method outperforms the competing strategies, in Fig. 1 we plot a weekly moving average of out-of-sample realized GMVP standard deviation differences obtained from two relatively good competing strategies (DCW diagonal and HAR DCC) to that from the LWE diagonal method. The figure clearly shows that, for different choices of N and years in the out-of-sample period, the LWE method consistently produces smaller realized portfolio volatilities on average. More importantly, a substantial reduction of realized portfolio volatility is achieved by the LWE method in the first half of 2020, when the US market crashed due to the covid outbreak. In detail, during the market turbulence period, the LWE method can reduce the annualized portfolio standard deviation by as much as 1% for $N = 100$ and 2% for $N = 200$. This evidence clearly demonstrates the empirical importance of the LWE method, which provides a highly effective portfolio risk reduction during periods with large market fluctuations.

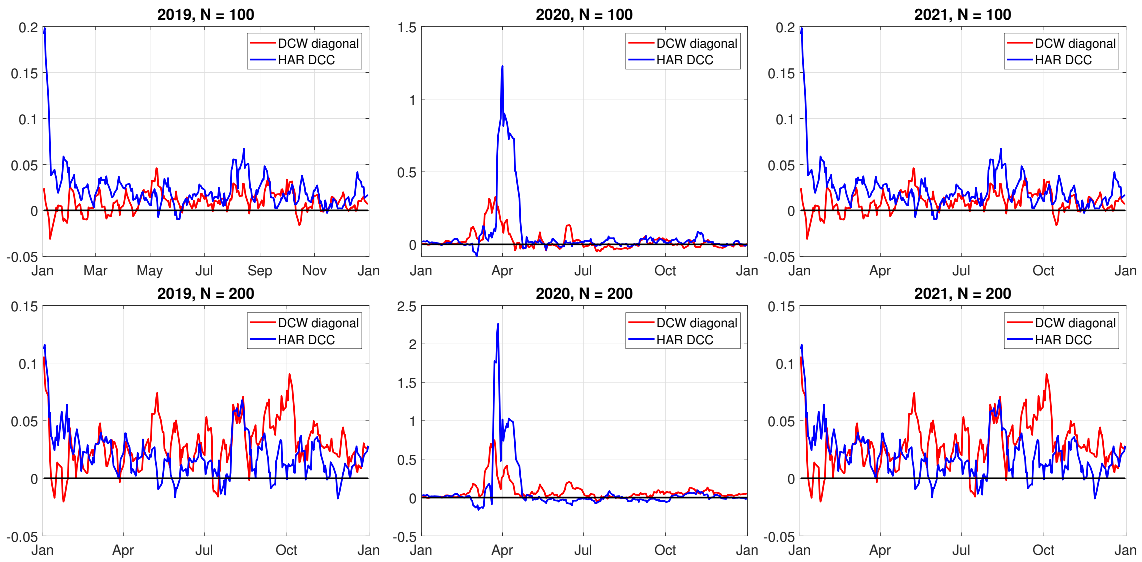
We proceed to analyse the transaction costs associated with the various GMVP strategies, which sheds light on the empirical feasibility of these methods. For our day-trading GMVP strategy, we do not hold assets overnight, thus the transaction cost at day t is proportional to the portfolio exposure, defined as $||w_t||_1$ for some portfolio weights w_t. Intuitively, as the GMVP weights sum up to 1, the portfolio exposure is 1 if short-selling is not allowed, thus the excess exposure over 1 captures the degree of short-selling required to achieve the GMVP allocation at time t. We report the average portfolio exposures for all the GMVP strategies in Table 4.3. The table shows that, apart from Panel 4 where all strategies have unit exposure due to the no short-selling constraint, the DCW and the ES strategies have overall the smallest portfolio exposure, followed by the LWE strategies, the HAR

Table 4.2: Annualized out-of-sample standard deviation of portfolio returns

N	LWE scalar	LWE diagonal	sample	RW	ES	HAR CCC	HAR DCC	DCW scalar	DCW diagonal
Panel 1: Unrestricted portfolio weights									
10	*12.74*	12.95	13.53	12.93	12.93	13.09	12.92	12.80	**12.62**
20	12.32	12.42	13.35	13.14	12.31	13.28	12.84	*12.24*	**12.07**
30	11.62	11.88	12.35	12.89	*11.49*	13.24	12.20	11.52	**11.40**
40	*11.20*	11.57	12.58	12.36	11.27	12.47	11.79	11.24	**11.16**
50	*11.26*	11.63	12.56	12.80	11.35	12.88	11.73	11.29	**11.26**
60	**11.46**	12.00	12.65	13.17	11.63	13.72	12.11	11.50	*11.47*
70	*11.69*	12.27	12.94	13.75	11.86	13.94	12.32	11.71	**11.65**
80	11.75	12.28	13.20	14.25	11.78	14.66	12.56	*11.63*	**11.62**
90	12.05	12.68	13.42	14.91	12.22	15.11	12.91	*12.01*	**11.99**
100	**11.81**	12.42	13.47	15.26	12.10	15.79	12.79	*11.89*	12.02
150	**12.16**	*12.48*	14.11	22.18	12.87	22.48	12.73	12.58	12.96
200	**12.11**	*12.35*	14.56	22.36	13.06	22.48	14.29	12.66	13.02
250	**11.66**	12.33	13.70	18.42	12.32	19.30	14.39	*12.27*	12.68
Panel 2: $\forall n \leq N, \lvert w^{(n)} \rvert \leq \max(5\%, 5/N)$									
10	*12.69*	12.83	13.53	12.80	12.92	12.95	12.85	12.74	**12.59**
20	12.31	12.37	13.27	13.04	12.32	12.97	12.78	*12.25*	**12.10**
30	11.68	11.85	12.30	12.43	*11.59*	12.58	12.18	11.64	**11.55**
40	**11.24**	11.48	12.48	11.68	11.42	11.92	11.85	11.35	*11.32*
50	**11.38**	11.63	12.39	11.89	11.54	12.15	11.78	*11.47*	11.50
60	**11.61**	12.09	12.47	12.30	11.90	12.55	12.16	*11.78*	11.79
70	**11.84**	12.24	12.54	12.54	12.14	12.74	12.48	*11.97*	11.98
80	**11.80**	12.11	12.65	12.70	12.02	13.04	12.54	*11.89*	11.95
90	**12.13**	12.35	12.68	13.14	12.41	13.14	12.86	*12.24*	12.32
100	**11.89**	12.20	12.60	13.14	12.24	13.35	12.74	*12.11*	12.34
150	**12.01**	*12.14*	13.59	14.33	12.74	14.68	12.58	12.59	13.12
200	**12.03**	*12.15*	13.96	15.23	12.98	14.89	12.72	12.66	12.98
250	**11.62**	*12.08*	13.28	15.31	12.36	15.49	12.72	12.27	12.69
Panel 3: $\forall n \leq N, \lvert w^{(n)} \rvert \leq \max(3\%, 3/N)$									
10	*12.80*	12.87	13.52	13.02	12.95	12.90	12.92	12.82	**12.69**
20	12.49	12.55	13.14	12.91	12.50	13.02	12.83	*12.45*	**12.28**
30	*11.96*	12.05	12.30	12.36	11.98	12.60	12.21	11.96	**11.82**
40	**11.54**	11.73	12.38	11.81	11.77	12.08	11.94	11.68	*11.62*
50	**11.62**	11.90	12.31	12.05	11.90	12.22	11.98	11.77	*11.76*
60	**11.81**	12.20	12.44	12.44	12.20	12.61	12.38	12.03	*11.97*
70	**12.03**	12.32	12.48	12.71	12.39	12.76	12.69	12.22	*12.21*
80	**11.96**	12.27	12.52	12.78	12.26	12.91	12.79	*12.12*	12.18
90	**12.15**	12.44	12.68	12.94	12.54	12.91	13.02	*12.40*	12.45
100	**12.05**	12.35	12.71	12.97	12.46	13.20	13.04	*12.33*	12.45
150	**12.04**	*12.32*	13.62	13.54	12.75	13.77	12.81	12.71	13.11
200	**11.75**	*12.17*	13.56	13.82	12.49	13.92	12.87	12.49	12.80
250	**11.60**	*12.07*	13.22	13.85	12.42	14.34	12.81	12.31	12.73
Panel 4: $\forall n \leq N, w^{(n)} \in [0, \max(3\%, 3/N)]$									
10	**12.89**	12.91	13.55	13.01	12.95	13.13	13.06	13.00	*12.90*
20	12.46	12.46	13.26	12.83	**12.40**	12.99	12.81	12.48	*12.44*
30	*11.98*	11.99	12.36	12.20	**11.92**	12.33	12.27	12.00	11.98
40	**11.69**	*11.69*	12.46	11.86	11.82	12.08	12.00	11.77	11.79
50	**11.93**	*11.95*	12.48	12.16	11.96	12.35	12.06	11.95	11.97
60	**12.12**	12.23	12.75	12.49	12.23	12.65	12.30	*12.19*	12.22
70	**12.35**	12.40	12.91	12.70	*12.38*	12.84	12.48	12.39	12.42
80	*12.35*	**12.34**	13.02	12.71	12.39	12.93	12.51	12.43	12.43
90	*12.50*	**12.50**	13.04	12.96	12.54	13.10	12.61	12.59	12.64
100	**12.42**	*12.47*	13.01	13.02	12.52	13.08	12.57	12.58	12.70
150	**12.25**	*12.39*	13.13	13.29	12.77	13.45	12.50	12.78	13.06
200	12.58	**12.33**	13.52	13.95	13.08	14.02	*12.57*	13.10	13.17
250	*12.26*	**12.21**	13.29	14.02	12.79	14.29	12.64	12.84	13.04

Numbers in the table correspond to the annualized average sample standard deviation of out-of-sample portfolio returns in percentage. For each portfolio the sample standard deviation is computed over an evaluation horizon $H = 756$ observations and an in-sample estimation window length $T = 1000$. Numbers in **bold** (*italic*) correspond to the smallest (second smallest) number in each row.

Figure 1: Relative performance of the LWE method to two competing strategies



The figure plots the 5-day moving average of the out-of-sample realized portfolio volatility (in percentage) of DCW diagonal or HAR DCC minus that from the LWE diagonal approach. The y-axis is the out-of-sample realized portfolio volatility differences, and the x-axis is the date.

strategies, and the sample and RW strategies in ascending order. The small portfolio exposure of the DCW and ES strategies is consistent with the findings in Golosnoy et al. (2019), Cipollini et al. (2021), which could be due to the smoothing effect of the vector ARIMA-type forecasting models for the weights. Nevertheless, the LWE method still outperforms the remaining competing models with a competitive average portfolio exposure. In fact, we shall show that one can further reduce the transaction cost of the LWE method via penalization, which is elaborated in the following section.

4.1 Penalizing the LWE Method for Transaction Cost Reduction

As discussed in Section 3.2.1, one can consider a penalized LWE method to reduce transaction costs. For the current day trading setup, we shall set $f(w) = N||w||_2^2$ to penalize the portfolio exposure, which directly translates to a lower transaction costs. One needs to choose a tuning parameter λ which controls for the degree of shrinkage: the larger λ, the closer the estimated weights to the equally weighted portfolio (with unit portfolio exposure), and hence less transaction cost. Obviously, this leads to a trade-off between transaction cost and portfolio variance, since any penalized solution is no longer minimizing the unconditional portfolio variance.

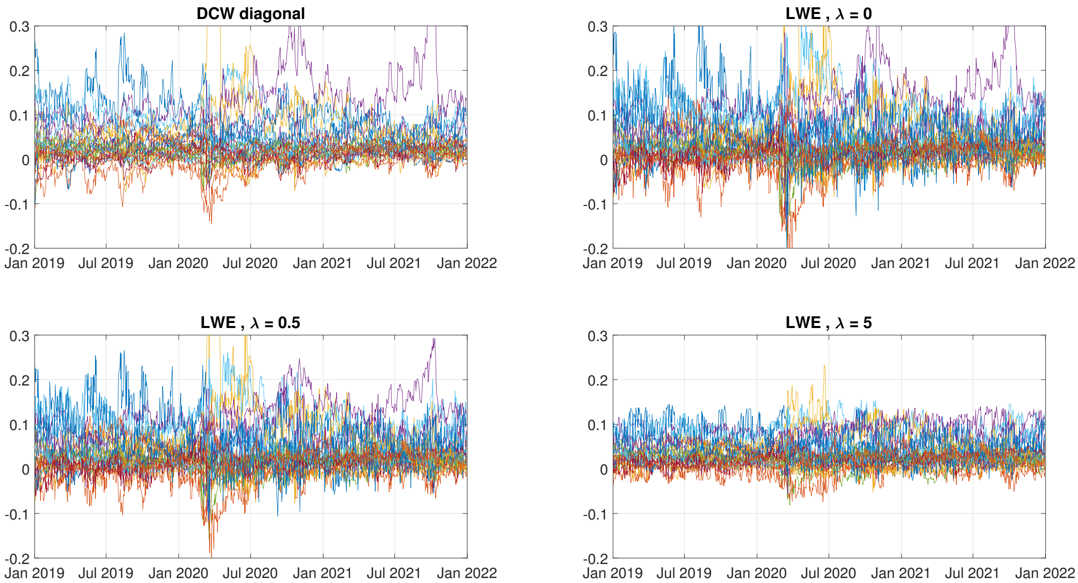
To showcase the effect of penalization on the time series of portfolio weights, Figure 2 plots the time series of the forecasted weights during the out-of-sample period for unrestricted estimators (the upper panels) and penalized LWE diagonal estimators (lower panels) without any investment constraints. As can be seen from the upper panels of the figure, the estimated weights of the unrestricted

Table 4.3: Out-of-sample average portfolio exposure

N	LWE scalar	LWE diagonal	sample	RW	ES	HAR CCC	HAR DCC	DCW scalar	DCW diagonal
Panel 1: Unrestricted portfolio weights									
10	1.09	1.09	1.21	1.15	1.07	1.14	1.09	*1.07*	**1.06**
20	1.15	1.13	1.53	1.33	1.12	1.33	1.16	*1.11*	**1.09**
30	1.25	1.23	1.78	1.63	1.19	1.63	1.29	*1.17*	**1.15**
40	1.38	1.34	2.08	1.93	1.28	1.93	1.43	*1.24*	**1.22**
50	1.50	1.46	2.32	2.23	1.39	2.24	1.57	*1.32*	**1.29**
60	1.64	1.59	2.57	2.55	1.51	2.55	1.70	*1.41*	**1.38**
70	1.74	1.69	2.88	2.89	1.61	2.89	1.84	*1.48*	**1.44**
80	1.85	1.77	3.11	3.24	1.70	3.23	1.95	*1.53*	**1.49**
90	1.92	1.86	3.44	3.60	1.79	3.60	2.04	*1.57*	**1.54**
100	2.01	1.93	3.64	4.00	1.89	3.99	2.13	*1.62*	**1.57**
150	2.48	2.31	4.74	6.26	2.38	6.12	2.55	*1.85*	**1.83**
200	2.78	2.67	5.84	7.32	2.60	7.02	2.94	**1.92**	*1.94*
250	2.97	2.93	6.59	6.45	2.45	6.19	3.23	*1.86*	**1.84**
Panel 2: $\forall n \leq N, \lvert w^{(n)}\rvert \leq \max(5\%, 5/N)$									
10	1.09	1.09	1.21	1.14	1.07	1.14	1.08	*1.07*	**1.06**
20	1.14	1.12	1.52	1.31	1.11	1.31	1.15	*1.10*	**1.08**
30	1.23	1.21	1.76	1.56	1.18	1.57	1.26	*1.16*	**1.13**
40	1.34	1.30	2.04	1.81	1.25	1.82	1.38	*1.22*	**1.20**
50	1.45	1.40	2.21	2.04	1.34	2.05	1.50	*1.29*	**1.26**
60	1.57	1.51	2.39	2.26	1.46	2.27	1.62	*1.38*	**1.35**
70	1.65	1.58	2.59	2.47	1.53	2.48	1.73	*1.43*	**1.40**
80	1.75	1.65	2.73	2.66	1.61	2.66	1.82	*1.48*	**1.45**
90	1.82	1.72	2.90	2.84	1.69	2.84	1.90	*1.53*	**1.49**
100	1.90	1.78	3.02	3.01	1.78	3.00	1.96	*1.57*	**1.53**
150	2.42	2.19	4.14	4.55	2.29	4.48	2.44	*1.84*	**1.81**
200	2.74	2.59	5.23	5.65	2.55	5.50	2.85	**1.92**	*1.93*
250	2.95	2.86	5.98	5.77	2.42	5.54	3.16	*1.86*	**1.84**
Panel 3: $\forall n \leq N, \lvert w^{(n)}\rvert \leq \max(3\%, 3/N)$									
10	1.07	1.07	1.18	1.11	1.07	1.11	1.07	*1.06*	**1.05**
20	1.11	1.09	1.46	1.25	1.09	1.25	1.12	*1.08*	**1.07**
30	1.17	1.16	1.61	1.44	1.14	1.45	1.20	*1.12*	**1.11**
40	1.27	1.23	1.82	1.61	1.20	1.63	1.30	*1.18*	**1.16**
50	1.36	1.32	1.90	1.76	1.27	1.77	1.39	*1.24*	**1.22**
60	1.46	1.41	1.99	1.89	1.36	1.90	1.48	*1.31*	**1.29**
70	1.51	1.46	2.09	2.00	1.42	2.01	1.56	*1.35*	**1.33**
80	1.58	1.51	2.16	2.10	1.48	2.10	1.62	*1.39*	**1.37**
90	1.63	1.56	2.25	2.18	1.55	2.18	1.67	*1.43*	**1.41**
100	1.69	1.60	2.30	2.26	1.60	2.26	1.71	*1.47*	**1.44**
150	2.22	2.03	3.23	3.37	2.10	3.33	2.20	*1.77*	**1.73**
200	2.60	2.44	4.13	4.29	2.42	4.20	2.64	*1.90*	**1.89**
250	2.84	2.73	4.86	4.71	2.36	4.56	2.98	*1.85*	**1.83**
Panel 4: $\forall n \leq N, w^{(n)} \in [0, \max(3\%, 3/N)]$									
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
150	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Numbers in the table correspond to the portfolio exposures averaged across $H = 756$ out-of-sample observations for portfolios of size N (in rows). Numbers in bold and italic correspond to the smallest and second smallest exposure for a given portfolio size N correspondingly.

Figure 2: Time series plots of DCW weight estimates and LWE diagonal with $\lambda = 0, 0.5, 5$.



Different panels on the plot correspond to the time series of estimated weights over the 3 years of the out-of-sample period for $N = 30$. The upper left panel corresponds to the unrestricted diagonal DCW weight estimator and other panels correspond to the LWE diagonal estimator with different values of the tuning parameter λ.

LWE estimator ($\lambda = 0$) are more volatile over time compared to the diagonal DCW weights, which explains its relatively high portfolio exposure as reported in Table 4.3. With the increase in the tuning parameter λ, the penalized LWE weights become smoother over time, resulting in a drastically smaller portfolio exposure. This clearly shows the effectiveness of the penalization to reduce portfolio exposure.

We proceed to examine the performances of the penalized LWE estimators relative to some well-established covariance estimators designed with the aim to stabilize the intertemporal portfolio weights and reduce transaction costs. As the performances of the penalized LWE diagonal and scalar estimators are similar, we only report the diagonal version for conciseness. The first competing method is a plug-in estimator with shrinkage to the market covariance matrix, which is estimated using T in-sample returns (Ledoit and Wolf, 2004). For the weight forecast, we use the last available estimate $\hat{\Sigma}_t$. The second method is a plug-in estimator with non-linear eigenvalue shrinkage of the covariance matrix, also estimated using T in-sample returns (Ledoit and Wolf, 2020). Again, the last available estimate $\hat{\Sigma}_t$ is used for the weight forecast. The third method is the daily equally weighted portfolio. We present the out-of-sample performances of the methods in Table 4.4.

Several interesting conclusions can be drawn from Table 4.4. First, as λ increases, the performance of the penalized LWE estimator measured in realized portfolio volatility always deteriorates (Panel 1), in exchange of a lower portfolio exposure (Panel 3). The portfolio return standard deviation

Table 4.4: Performances of the penalized LWE estimator and the competing strategies

N	LWE λ =0.1	LWE λ =0.5	LWE λ =1	LWE λ =3	LWE λ =5	LW 2004	LW 2020	1/N
Panel 1: out-of-sample annualized average portfolio realized volatility								
10	**14.36**	*14.37*	14.38	14.47	14.57	16.02	16.08	18.48
20	**13.44**	*13.46*	13.51	13.73	13.90	15.60	15.77	16.07
30	**12.84**	*12.86*	12.94	13.22	13.43	14.85	15.00	15.31
40	**12.63**	*12.67*	12.78	13.13	13.36	14.68	14.87	15.04
50	**12.53**	*12.60*	12.74	13.14	13.39	14.63	14.85	15.01
60	**12.49**	*12.58*	12.75	13.20	13.47	14.59	14.79	15.07
70	**12.51**	*12.60*	12.76	13.24	13.54	14.65	14.84	15.18
80	**12.29**	*12.38*	12.56	13.09	13.42	14.65	14.85	15.11
90	**12.19**	*12.30*	12.49	13.03	13.36	14.63	14.83	14.99
100	**12.18**	*12.31*	12.52	13.09	13.42	14.60	14.79	15.03
150	**12.21**	*12.41*	12.66	13.31	13.67	14.66	14.86	15.43
200	**12.16**	*12.34*	12.65	13.44	13.86	14.48	14.70	15.72
250	**11.64**	*11.89*	12.30	13.38	13.95	14.23	14.37	16.19
Panel 2: out-of-sample annualized standard deviation of portfolio returns								
10	12.94	12.90	*12.87*	**12.87**	12.92	13.54	13.53	16.82
20	12.38	*12.29*	**12.26**	12.33	12.44	13.31	13.32	14.16
30	11.81	*11.69*	**11.67**	11.81	11.96	12.31	12.32	13.56
40	11.46	*11.30*	**11.28**	11.50	11.70	12.43	12.51	13.34
50	11.51	**11.39**	*11.42*	11.69	11.89	12.44	12.49	13.32
60	11.83	**11.64**	*11.67*	11.95	12.15	12.54	12.57	13.47
70	12.07	**11.83**	*11.83*	12.09	12.30	12.71	12.76	13.62
80	12.05	**11.80**	*11.81*	12.09	12.30	12.92	13.01	13.59
90	12.39	*12.06*	**12.02**	12.22	12.39	13.03	13.18	13.59
100	12.15	**11.91**	*11.93*	12.21	12.41	13.06	13.18	13.67
150	12.09	**11.92**	*12.01*	12.40	12.65	13.36	13.61	14.12
200	12.02	**11.85**	*11.96*	12.44	12.75	13.41	13.60	14.38
250	12.03	**11.74**	*11.82*	12.41	12.81	12.97	12.80	14.71
Panel 3: out-of-sample average portfolio exposure								
10	1.09	1.08	1.08	*1.06*	**1.05**	1.19	1.21	**1.00**
20	1.13	1.11	1.10	*1.06*	**1.05**	1.45	1.50	**1.00**
30	1.22	1.19	1.16	*1.10*	**1.06**	1.65	1.72	**1.00**
40	1.33	1.27	1.23	*1.13*	**1.08**	1.90	2.00	**1.00**
50	1.43	1.36	1.30	*1.17*	**1.11**	2.07	2.19	**1.00**
60	1.56	1.47	1.39	*1.22*	**1.14**	2.25	2.39	**1.00**
70	1.65	1.53	1.43	*1.23*	**1.14**	2.51	2.64	**1.00**
80	1.72	1.58	1.47	*1.24*	**1.15**	2.71	2.82	**1.00**
90	1.80	1.63	1.51	*1.26*	**1.15**	2.92	3.05	**1.00**
100	1.86	1.68	1.54	*1.27*	**1.16**	3.07	3.20	**1.00**
150	2.20	1.93	1.73	*1.36*	**1.21**	3.91	3.88	**1.00**
200	2.52	2.15	1.89	*1.43*	**1.25**	4.61	4.47	**1.00**
250	2.74	2.27	1.95	*1.42*	**1.23**	5.07	4.78	**1.00**

For each strategy and N, panel 1 reports the annualized average realized GMVP portfolio volatility in percentage. Panel 2 reports the annualized average sample standard deviation of out-of-sample portfolio returns in percentage. Panel 3 reports the portfolio exposures averaged over the out-of-sample period. All the measures are computed over an evaluation horizon $H = 756$ observations and an in-sample estimation window length $T = 1000$. Numbers in bold and italic correspond to the smallest and second smallest number in each row (in panel 3, the ranking does not include the $1/N$ portfolio).

seems to first decrease then increase with λ (Panel 2), reflecting the differences of the two objective functions. Importantly, regardless of the choice of λ, the LWE method comfortably beats the two shrinkage-based methods and the equally weighted portfolio in minimizing the portfolio variance. The LWE method even has considerably lower portfolio exposure than the shrinkage-based methods, which clearly demonstrates the practicability of our method. Also, the LWE method with $\lambda = 1$ still delivers smaller realized portfolio volatility than those of the DWE methods in Table 4.1 on average, but with a comparable transaction cost. One can further impose investment constraints by combining Eq. (4.1) with the penalized LWE estimator. The results are qualitatively similar and are available upon request.

In practice, the choice of λ depends on the practitioner's trade-off between portfolio performance measures and the transaction cost, which can be fine-tuned to deliver a desirable level of transaction cost and at the same time minimising the unconditional realized portfolio variance. Distinct from all its competing methods, this unique feature of our LWE method provides the investor with great flexibility to tailor the portfolio allocations while preserving certain optimality of the target performance metric.

5 Conclusions

This paper presents a novel semi-parametric linear portfolio weight estimation approach (LWE) that offers several advantages over existing methods. By assuming a dynamic linear model directly for the *ex ante* optimal GMVP weights, the LWE model avoids restrictive assumptions about return distributions or high-dimensional covariance matrix dynamics. The parameters of the LWE model can be estimated by directly minimizing the unconditional GMVP problem whose solution is unique and in closed form. Consequently, the LWE model can accommodate rich dynamics in the optimal GMVP weights with a very fast computation speed, even in relatively large dimensions. The theoretical properties of the LWE parameter estimates, such as consistency and asymptotic normality, can be established by exploiting its method-of-moment nature. The theoretical results provide a basis for rigorous statistical inference and hypothesis tests. Several extensions of the LWE model are also discussed, which provide useful tools to improve ivestability of the portfolio weights.

In the empirical application, the LWE is shown to outperform competing estimators that rely on realized covariance forecasting and shrinkage of covariance matrices together with alternative approaches employed in GMVP realized weight forecasting. The LWE exhibits superior forecasting precision and economic measures of portfolio performance. Furthermore, the LWE can be extended to incorporate controls for portfolio rebalancing that leads to a reduction in transaction costs.

As a potential avenue for future research, it would be interesting to explore the use of machine learning techniques to improve the univariate forecasting model employed in the LWE method. This could potentially enhance the forecasting accuracy and performance of the estimator. Following the same vein, one can also extend the LWE method to solve the dynamic mean-variance portfolio allocation problem. Overall, the LWE approach presented in this paper offers a promising framework for dynamic portfolio weight estimation, as it provides robust results, theoretical support, and practical advantages for portfolio management and decision-making.

References

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

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.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, 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(2):149–169.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101.

Bibinger, M. and Mykland, P. A. (2016). Inference for Multi-dimensional High-frequency Data with an Application to Conditional Independence Testing. *Scandinavian Journal of Statistics*, 43(4):1078–1102.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207(1):71–91.

Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R., and Sauri, O. (2017). Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics*, 196(2):347–367.

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance*, 61(5):2187–2217.

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the Cross-Section of equity returns. *Review of Financial Studies*, 22(9):3411–3447.

Brodie, J., Daubechies, I., Mol, C. D., Giannone, D., and Loris, I. (2009). Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences*, 106(30):12267–12272. PMID: 19617537.

Callot, L., Caner, M., Önder, A. Ö., and Ulaşan, E. (2021). A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*, 39(2):520–531.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Cipollini, F., Gallo, G. M., and Palandri, A. (2021). A dynamic conditional approach to forecasting portfolio weights. *International Journal of Forecasting*, 37(3):1111–1126.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

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

DeMiguel, V., Nogales, F. J., and Uppal, R. (2014). Stock return serial dependence and Out-of-Sample portfolio performance. *Review of Financial Studies*, 27(4):1031–1073.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Ding, Y., Li, Y., and Zheng, X. (2021). High dimensional minimum variance portfolio estimation under statistical factor models. *Journal of Econometrics*, 222(1):502–515.

Domowitz, I. and White, H. (1982). Misspecified models with dependent observations. *Journal of Econometrics*, 20(1):35–58.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Fan, J., Furger, A., and Xiu, D. (2016). Incorporating Global Industrial Classification Standard Into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator With High-Frequency Data. *Journal of Business & Economic Statistics*, 34(4):489–503.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012). The conditional autoregressive wishart model for multivariate stock market volatility. *Journal of Econometrics*, 167(1):211–223.

Golosnoy, V., Gribisch, B., and Seifert, M. I. (2019). Exponential smoothing of realized portfolio weights. *Journal of Empirical Finance*, 53:222–237.

Gourieroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart Autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.

Hall, A. R. (1996). *Generalized Method of Moments: Advanced Text in Econometrics*, volume 100-101. Oxford University Press, New York.

Hansen, P. R. and Lunde, A. (2006). Consistent ranking of volatility models. *Journal of Econometrics*, 131(1):97–121.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hautsch, N., Kyj, L. M., and Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2):263–290.

Hautsch, N. and Voigt, S. (2019). Large-scale portfolio allocation under transaction costs and model uncertainty. *Journal of Econometrics*, 212(1):221–240.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684.

Jin, X. and Maheu, J. M. (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics*, 11(2):335–369.

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.

Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.

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.

Ledoit, O. and Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Annals of Statistics*, 48(5):3043–3065.

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business & Economic Statistics*, 33(3):381–392.

Li, Y., Nolte, I., Vasios, M., Voev, V., and Xu, Q. (2022). Weighted Least Squares Realized Covariation Estimation. *Journal of Banking & Finance*, 137:106420.

Lunde, A., Shephard, N., and Sheppard, K. (2016). Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels and Their Application to Portfolio Choice. *Journal of Business and Economic Statistics*, 34(4):504–518.

McLeish, D. L. (1975). A Maximal Inequality and Dependent Strong Laws. *The Annals of Probability*, 3(5):829–839.

Newey, W. K. and McFadden, D. B. T. H. o. E. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 27(6):907–933.

Oh, D. H. and Patton, A. J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2):349–366.

Okhrin, Y. and Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of Econometrics*, 134(1):235–256.

Opschoor, A., Janus, P., Lucas, A., and Van Dijk, D. (2018). New HEAVY Models for Fat-Tailed

Realized Covariances and Returns. *Journal of Business & Economic Statistics*, 36(4):643–657.

Palandri, A. (2022). Rank-invariance conditions for the comparison of volatility forecasts. *The Econometrics Journal*, 25(1):155–175.

Patton, A. J. and Sheppard, K. (2009). Evaluating Volatility and Correlation Forecasts. In Mikosch, T., Kreiß, J. P., Davis, R., and Andersen, T. G., editors, *Handbook of Financial Time Series*, pages 810–838. Springer Berlin Heidelberg.

Tibshirani, R. (1994). Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B*, 58:267–288.

Varneskov, R. T. (2016). Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices. *Journal of Business & Economic Statistics*, 34(1):1–22.

Yen, Y.-M. (2016). Sparse weighted-norm minimum variance portfolios. *Review of Finance*, 20(3):1259–1287.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

A Appendix

A.1 Technical Proofs

Proof of Proposition 1. We shall abuse the notation and write 0 as a vector of zeros with appropriate dimension throughout this proof. Start with the observation that $\tilde{Z}'\iota = 0$. Now notice that $\tilde{Z}\tilde{\epsilon}_t = \tilde{Z}(\tilde{w}_t^{\star} - \tilde{w}_t^{\bigstar}) = w_t^{\star} - w_t^{\bigstar} \in \mathcal{V}$. Therefore:

$$\mathrm{E}_{t-1}[\tilde{\Sigma}_t\tilde{\epsilon}_t] = \mathrm{E}_{t-1}[\tilde{Z}'\Sigma_t(w_t^{\star} - w_t^{\bigstar})] = \mathrm{E}_{t-1}\left[\frac{\tilde{Z}'\iota}{\iota'\Sigma_t^{-1}\iota}\right] - \frac{\tilde{Z}'\iota}{\iota'\Omega_t^{-1}\iota} = 0. \tag{A.1}$$

This completes the proof. □

Proof of Theorem 1. Assumption 3 implies that $\boldsymbol{\beta}_0$ is the unique optimal solution of the following problem for every t by the optimality of $w_t^{\bigstar}$ and the strict convexity of the optimization problem:

$$\min_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}_{t-1}[(w_t(\boldsymbol{\beta}) + \eta_t)'\Sigma_t(w_t(\boldsymbol{\beta}) + \eta_t)]. \tag{A.2}$$

Since $\boldsymbol{\beta}_0$ is optimal and unique for every t, it must also be the unique optimal solution of the following problem below by the law of total expectation:

$$\min_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}[(w_t(\boldsymbol{\beta}) + \eta_t)'\Sigma_t(w_t(\boldsymbol{\beta}) + \eta_t)]. \tag{A.3}$$

To prove the first minimization problem of Eq. (3.12), note that:

$$\mathrm{E}[(w_t(\boldsymbol{\beta}) + \eta_t)'\Sigma_t(w_t(\boldsymbol{\beta}) + \eta_t)] = \mathrm{E}[w_t(\boldsymbol{\beta})'\Sigma_t w_t(\boldsymbol{\beta})] + 2\,\mathrm{E}[w_t(\boldsymbol{\beta})'\Sigma_t\eta_t] + \mathrm{E}[\eta_t'\Sigma_t\eta_t]. \tag{A.4}$$

Clearly the last term is independent of $\boldsymbol{\beta}$ and can be ignored in the optimization. For the second from last term, using the definition of $w_t(\boldsymbol{\beta})$ and recall Assumption 3:

$$\mathrm{E}[w_t(\boldsymbol{\beta})'\Sigma_t\eta_t] = \mathrm{E}[\boldsymbol{e}_N'\Sigma_t\eta_t] + \boldsymbol{\beta}'\,\mathrm{E}[\boldsymbol{X}_t'\tilde{Z}'\Sigma_t\tilde{Z}\tilde{\eta}_t] = \mathrm{E}[\boldsymbol{e}_N'\Sigma_t\eta_t], \tag{A.5}$$

which is also independent of $\boldsymbol{\beta}$. The above result directly implies that:

$$\min_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}[(w_t(\boldsymbol{\beta}) + \eta_t)'\Sigma_t(w_t(\boldsymbol{\beta}) + \eta_t)] \Leftrightarrow \min_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}[w_t(\boldsymbol{\beta})'\Sigma_t w_t(\boldsymbol{\beta})], \tag{A.6}$$

which is the claimed result in the first minimization problem of Eq. (3.12). For the second minimization problem in Eq. (3.12), notice that:

$$\begin{aligned} w_t(\boldsymbol{\beta})'\Sigma_t w_t(\boldsymbol{\beta}) &= (w_t(\boldsymbol{\beta}) - w_t^{\star} + w_t^{\star})'\Sigma_t(w_t(\boldsymbol{\beta})' - w_t^{\star} + w_t^{\star}) \\ &= \underbrace{(w_t^{\star} - w_t(\boldsymbol{\beta}))'}_{=\tilde{u}_t(\boldsymbol{\beta})'\tilde{Z}'}\Sigma_t w_t(\boldsymbol{\beta})(w_t^{\star} - w_t(\boldsymbol{\beta})) + (w_t^{\star})'\Sigma_t w_t^{\star} \\ &= \tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta}) + (w_t^{\star})'\Sigma_t w_t^{\star}. \end{aligned} \tag{A.7}$$

where the cross term vanishes due to $(w_t(\boldsymbol{\beta})' - w_t^{\star})'\Sigma_t w_t^{\star} \propto (w_t(\boldsymbol{\beta})' - w_t^{\star})'\iota = 0$. As the last quantity does not depend on $\boldsymbol{\beta}$, minimizing $\mathrm{E}[w_t(\boldsymbol{\beta})'\Sigma_t w_t(\boldsymbol{\beta})]$ is equivalent to minimizing $\mathrm{E}[\tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta})]$, which proves the equivalence of the two optimizations in Eq. (3.12) as desired.

Eq. (3.13) can be obtained by taking the first-order condition of $\min_{\boldsymbol{\beta}\in\mathcal{B}} \mathrm{E}[\tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta})]$ and set it to zero. Finally, Eq. (3.17) can be obtained by solving for $\boldsymbol{\beta}_0$ explicitly from Eq. (3.13):

$$\mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta}_0)] = 0 \Leftrightarrow \mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\boldsymbol{X}_t]\boldsymbol{\beta}_0 = \mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\tilde{w}_t^{\star}] \Leftrightarrow \boldsymbol{\beta}_0 = \mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\boldsymbol{X}_t]^{-1}\,\mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\tilde{w}_t^{\star}].$$

and Assumption 3 ensures that $\mathrm{E}[\boldsymbol{X}_t'\tilde{\Sigma}_t\boldsymbol{X}_t]$ is invertible, so the solution exists and is unique. This completes the proof. $\square$

Proof of Proposition 2. Note that when $\tilde{w}_t^{\bigstar} = \boldsymbol{\beta}_0$ and $\boldsymbol{X}_t = I_{N-1}$, $\boldsymbol{\beta}_0$ satisfies the following equation by Eq. (3.17):

$$\mathrm{E}[\tilde{\Sigma}_t]\boldsymbol{\beta}_0 = \mathrm{E}[\tilde{\Sigma}_t\tilde{w}_t^{\star}] = \mathrm{E}\left[\frac{\tilde{Z}'\Sigma_t\tilde{Z}Z\Sigma_t^{-1}\iota}{\iota'\Sigma_t^{-1}\iota}\right]. \tag{A.8}$$

We shall prove that the above equation holds when $\boldsymbol{\beta}_0 = \frac{Z\,\mathrm{E}[\Sigma_t]^{-1}\iota}{\iota'\,\mathrm{E}[\Sigma_t]^{-1}\iota}$, which implies that:

$$\frac{\tilde{Z}'\,\mathrm{E}[\Sigma_t]\tilde{Z}Z\,\mathrm{E}[\Sigma_t]^{-1}\iota}{\iota'\,\mathrm{E}[\Sigma_t]^{-1}\iota} = \mathrm{E}\Big[\frac{\tilde{Z}'\Sigma_t\tilde{Z}Z\Sigma_t^{-1}\iota}{\iota'\Sigma_t^{-1}\iota}\Big]. \tag{A.9}$$

To prove this result, let $\boldsymbol{S}$ be an arbitrary N-by-N positive definite matrix, and consider the following two equivalent optimization problems, where the equivalence follows from Eq. (3.12):

$$\min_{\boldsymbol{\beta}\in\mathcal{B}}(\boldsymbol{e}_N + \tilde{Z}\boldsymbol{\beta})'\boldsymbol{S}(\boldsymbol{e}_N + \tilde{Z}\boldsymbol{\beta}) \Leftrightarrow \min_{\boldsymbol{\beta}\in\mathcal{B}}\Big(\frac{Z\boldsymbol{S}^{-1}\iota}{\iota'\boldsymbol{S}^{-1}\iota} - \boldsymbol{\beta}\Big)'\tilde{Z}'\boldsymbol{S}\tilde{Z}\Big(\frac{Z\boldsymbol{S}^{-1}\iota}{\iota'\boldsymbol{S}^{-1}\iota} - \boldsymbol{\beta}\Big). \tag{A.10}$$

As the optimization problem is strictly convex with a unique solution, their solutions must coincide. Equating the solutions obtained from both problems yields:

$$-\tilde{Z}'\boldsymbol{S}\boldsymbol{e}_N = \frac{\tilde{Z}'\boldsymbol{S}\tilde{Z}Z\boldsymbol{S}^{-1}\iota}{\iota'\boldsymbol{S}^{-1}\iota}. \tag{A.11}$$

As the choice of $\boldsymbol{S}$ is arbitrary, one can apply the above relation to both sizes of Eq. (A.10) by setting $\boldsymbol{S} = \mathrm{E}[\Sigma_t]$ and $\boldsymbol{S} = \Sigma_t$, respectively:

$$-\tilde{Z}'\,\mathrm{E}[\Sigma_t]\boldsymbol{e}_N = \frac{\tilde{Z}'\,\mathrm{E}[\Sigma_t]\tilde{Z}Z\,\mathrm{E}[\Sigma_t]^{-1}\iota}{\iota'\,\mathrm{E}[\Sigma_t]^{-1}\iota} = \mathrm{E}\Big[\frac{\tilde{Z}'\Sigma_t\tilde{Z}Z\Sigma_t^{-1}\iota}{\iota'\Sigma_t^{-1}\iota}\Big] = -\,\mathrm{E}[\tilde{Z}'\Sigma_t\boldsymbol{e}_N], \tag{A.12}$$

but the leftmost and the rightmost expressions are clearly equivalent. This completes the proof. □

Proof of Proposition 3. Let us write the objective function in Eq. (3.26) as:

$$G_T(\boldsymbol{\beta};\lambda) = \frac{1}{T}\sum_{t=1}^{T}\tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta}) + \lambda f(\boldsymbol{e}_N + \tilde{Z}\boldsymbol{X}_{T+1}\boldsymbol{\beta}), \tag{A.13}$$

which is clearly a quadratic function of $\boldsymbol{\beta}$. We shall now verify that, for all finite $\lambda \geq 0$, the above function is strictly convex in $\boldsymbol{\beta}$, thus a unique minimizer exists. To wit, we can compute the Hessian of the optimization problem:

$$\frac{\partial^2 G_T(\boldsymbol{\beta};\lambda)}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} = \underbrace{\frac{2}{T}\sum_{t=1}^{T}\boldsymbol{X}_t'\tilde{\Sigma}_t\boldsymbol{X}_t}_{(I)} + 2\lambda\underbrace{\boldsymbol{X}_{T+1}'\tilde{\boldsymbol{A}}\boldsymbol{X}_{T+1}}_{(II)} \succ 0, \tag{A.14}$$

where $\tilde{\boldsymbol{A}} = \tilde{Z}'\boldsymbol{A}\tilde{Z}$. Note that (I) is of full rank and hence positive definite by the assumption that $T > b(K)$. For (II), $\tilde{\boldsymbol{A}}$ with rank $N-1$ is positive definite by assumption, and the quadratic structure suggests that the second term is positive semi-definite, although not positive definite as it is rank deficient due to the fact that $b(K) > N-1$ in general. As a result, for any finite λ, $(I) + 2\lambda(II)$ is positive definite, hence $G_T(\boldsymbol{\beta};\lambda)$ is strictly convex, and any minimizer of $G_T(\boldsymbol{\beta};\lambda)$ must be unique.

In fact, the rank deficiency of (II) also hinted that the solution of $G_T(\boldsymbol{\beta};\lambda)$ may not be unique as $\lambda\to\infty$, which we shall discuss later in this proof.

For any finite λ, we now derive the unique minimizer of $G_T(\boldsymbol{\beta};\lambda)$. This can be done by setting the first order condition to zero:

$$\begin{aligned}\frac{\partial G_T(\boldsymbol{\beta};\lambda)}{\partial\boldsymbol{\beta}}&=\frac{2}{T}\sum_{t=1}^{T}(\boldsymbol{X}_t'\tilde{\Sigma}_t\boldsymbol{X}_t\boldsymbol{\beta}-\boldsymbol{X}_t\tilde{\Sigma}_t\tilde{w}_t^{\star})+2\lambda(\boldsymbol{X}_{T+1}'\tilde{\boldsymbol{A}}\boldsymbol{X}_{T+1}\boldsymbol{\beta}+\boldsymbol{X}_{T+1}'\tilde{Z}'(\boldsymbol{A}\boldsymbol{e}_N+\boldsymbol{a}))=\mathbf{0}_{b(K)\times1}\\ \Leftrightarrow\hat{\boldsymbol{\beta}}_{T,\lambda}&=\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t'\tilde{\Sigma}_t\boldsymbol{X}_t+\lambda\boldsymbol{X}_{T+1}'\tilde{\boldsymbol{A}}\boldsymbol{X}_{T+1}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\tilde{\Sigma}_t\tilde{w}_t^{\star}-\lambda\boldsymbol{X}_{T+1}'\tilde{Z}'(\boldsymbol{A}\boldsymbol{e}_N+\boldsymbol{a})\right),\end{aligned}\tag{A.15}$$

which is the desired expression. One should immediately verify that the matrix inverse is well-defined for any finite λ by Eq. (A.14), and $\hat{\boldsymbol{\beta}}_{T,0}=\hat{\boldsymbol{\beta}}_T$ where $\hat{\boldsymbol{\beta}}_{T,\lambda}$ is clearly continuous at $\lambda=0$, which proves the claim as $\lambda\to0$.

We now turn to the case $\lambda\to\infty$. First, one should notice that as $\lambda\to\infty$, the matrix inverse in $\hat{\boldsymbol{\beta}}_{T,\lambda}$ becomes defective due to the aforementioned rank deficiency problem of (II). This results in a positive semi-definite Hessian of the limiting problem $\lim_{\lambda\to\infty}G_T(\boldsymbol{\beta};\lambda)$. As a result, there can be many minima $\hat{\beta}_{T,\infty}$ that minimize $\lim_{\lambda\to\infty}G_T(\boldsymbol{\beta};\lambda)$, all with the same minimized objective function. We shall now characterize all solutions to this problem. Notice that, as λ diverges, the minimization problem is equivalent to the minimization of the penalty function:

$$\min_{\boldsymbol{\beta}\in\mathcal{B}}\lim_{\lambda\to\infty}\frac{1}{T}\sum_{t=1}^{T}\tilde{u}_t(\boldsymbol{\beta})'\tilde{\Sigma}_t\tilde{u}_t(\boldsymbol{\beta})+\lambda f(w_{T+1}(\boldsymbol{\beta}))\Leftrightarrow\min_{\boldsymbol{\beta}\in\mathcal{B}}f(w_{T+1}(\boldsymbol{\beta})).\tag{A.16}$$

However, the penalty function $f(w)$ has a well-defined unique minimum w^* in Eq. (3.30) by the positive definiteness of $\boldsymbol{A}$, which can be solved by the standard Lagrange multiplier method. Therefore, if any optimal $\hat{\boldsymbol{\beta}}_{T,\infty}$ exists, it must satisfy $w^*=w_{T+1}(\hat{\boldsymbol{\beta}}_{T,\infty})$. We now show that such $\hat{\boldsymbol{\beta}}_{T,\infty}$ always exists, so w^* is always attained. Write $\tilde{w}^*=Zw^*\in\tilde{\mathcal{W}}$ as the dimension-reduced optimal weight, then $w^*=w_{T+1}(\hat{\boldsymbol{\beta}}_{T,\infty})$ is equivalent to the existence of $\hat{\boldsymbol{\beta}}_{T,\infty}$ such that $\tilde{w}^*=\boldsymbol{X}_{T+1}\hat{\boldsymbol{\beta}}_{T,\infty}$ holds. As $\boldsymbol{X}_{T+1}$ has rank $b(K)\geq N-1$ by Assumption 3, where the $N-1$ comes from the constant term in Eq. (3.3) that is always included, $\hat{\boldsymbol{\beta}}_{T,\infty}$ solves the underdetermined system of equations $\boldsymbol{X}_{T+1}\boldsymbol{\beta}-\tilde{w}^*=\mathbf{0}_{(N-1)\times1}$, which either has zero or infinitely many solutions. Nevertheless, in the notation of Eq. (3.3), the following is always a trivial solution to the above problem for all $b(K)\geq N-1$, which does not depend on the choice of $\boldsymbol{X}_{T+1}$:

$$\hat{\boldsymbol{\beta}}_{T,\infty}=[\tilde{w}^*;\mathbf{0}_{(b(K)-N+1)\times1}].\tag{A.17}$$

Therefore, for any $b(K)>N-1$, there must exist infinitely many solutions $\hat{\boldsymbol{\beta}}_{T,\infty}$, which jointly satisfy $\tilde{w}^*=\boldsymbol{X}_{T+1}\hat{\boldsymbol{\beta}}_{T,\infty}$ and hence $w^*=w_{T+1}(\hat{\boldsymbol{\beta}}_{T,\infty})$. This completes the proof. $\square$

A.2 Descriptive statistic

Table A.1: Descriptive statistic

	Returns					RC			Returns					RC	
Ticker	Min	Mean	Max	Skewness	Kurtosis	Av. var.	Av. corr.	Ticker	Min	Mean	Max	Skewness	Kurtosis	Av. var.	Av. corr.
AAPL	-7.76	0.06	8.32	-0.17	6.09	2.12	0.22	SYY	-17.31	0.01	17.57	-0.13	33.88	2.61	0.22
MSFT	-6.52	0.05	7.41	-0.23	6.54	1.95	0.22	EXC	-8.69	0.05	14.07	0.31	13.71	2.15	0.16
AMZN	-8.86	-0.02	7.10	-0.28	5.70	2.40	0.19	VLO	-13.58	-0.04	11.47	-0.09	7.44	4.94	0.15
GOOG	-6.18	0.02	4.99	-0.47	5.55	1.89	0.22	COF	-11.03	-0.01	11.75	-0.38	9.47	2.97	0.25
TSLA	-13.68	0.03	14.68	0.28	5.83	7.30	0.10	O	-20.01	-0.01	11.65	-2.30	39.84	2.20	0.15
BRK_B	-5.09	-0.04	5.12	-0.21	7.52	1.10	0.36	ADM	-10.24	0.00	5.91	-0.36	6.83	1.99	0.24
UNH	-8.08	0.03	8.36	0.02	8.41	2.29	0.19	AIG	-14.48	-0.01	10.70	-0.47	13.37	2.86	0.25
JNJ	-7.96	0.00	6.07	-0.89	11.22	1.34	0.23	TRV	-11.90	0.01	10.16	-0.30	16.41	1.74	0.26
NVDA	-11.14	0.06	10.74	-0.20	6.25	4.50	0.14	FCX	-21.73	-0.06	16.60	0.04	9.05	7.92	0.12
XOM	-7.16	-0.04	8.07	-0.11	6.93	2.14	0.25	HCA	-11.38	0.04	22.73	0.77	23.03	4.87	0.16
PG	-6.48	0.04	7.08	-0.05	11.62	1.36	0.20	ECL	-12.10	0.00	12.48	-0.48	21.38	1.87	0.24
JPM	-5.44	0.00	8.17	0.07	6.51	2.16	0.29	NXPI	-11.60	-0.04	16.07	0.42	10.36	4.00	0.14
V	-7.00	-0.01	6.36	-0.24	7.11	1.80	0.25	STZ	-12.14	0.00	10.82	-0.33	13.17	2.47	0.18
HD	-9.26	0.04	7.25	-0.31	9.43	1.90	0.26	ADSK	-10.78	0.06	8.61	-0.30	6.26	3.57	0.16
PFE	-6.59	-0.02	5.78	-0.31	7.25	1.78	0.21	PSX	-12.75	-0.05	10.28	-0.31	8.30	3.70	0.19
CVX	-18.70	-0.03	7.90	-0.98	22.52	2.35	0.24	MAR	-15.12	0.04	9.56	-0.44	10.72	3.62	0.21
MA	-11.27	-0.02	9.17	-0.56	9.88	2.11	0.24	XEL	-14.61	0.05	8.04	-0.91	17.85	1.74	0.14
ABBV	-10.92	0.01	7.19	-0.47	7.35	2.73	0.17	DLTR	-12.88	0.01	10.65	-0.15	11.76	2.90	0.16
LLY	-7.89	0.05	10.26	0.43	9.28	2.15	0.17	WMB	-29.10	-0.04	31.02	-0.09	39.05	5.53	0.17
KO	-9.41	0.01	6.29	-0.74	12.67	1.20	0.23	DLR	-11.02	0.03	9.28	-0.67	9.98	2.48	0.11
MRK	-7.92	-0.03	7.24	-0.16	7.31	1.79	0.21	TEL	-9.32	0.00	6.80	-0.54	7.47	2.32	0.26
PEP	-12.89	0.03	9.02	-0.68	25.25	1.40	0.22	A	-5.11	0.02	7.25	0.01	5.26	2.10	0.23
BAC	-10.24	-0.02	8.46	-0.14	7.27	2.41	0.26	MNST	-7.35	0.01	12.44	0.24	8.14	2.70	0.16
COST	-6.60	0.04	6.04	-0.04	7.32	1.44	0.22	PRU	-13.15	-0.02	10.38	-0.55	10.18	2.81	0.26
VZ	-5.00	0.00	7.48	0.27	7.20	1.42	0.21	MSI	-8.47	0.04	9.54	-0.39	10.02	2.09	0.21
TMO	-7.20	0.01	6.76	-0.42	6.64	2.06	0.21	ALL	-7.75	0.02	7.54	0.01	10.19	1.80	0.28
AVGO	-10.76	0.02	14.75	0.07	8.05	3.34	0.17	CTSH	-8.10	0.01	11.58	0.10	9.51	2.24	0.23
ABT	-9.16	0.03	7.03	-0.44	8.66	1.82	0.23	EA	-6.83	0.03	12.11	0.14	6.40	3.31	0.13
MCD	-6.46	0.03	10.05	0.70	15.09	1.41	0.23	YUM	-9.85	0.03	12.05	0.15	13.78	1.99	0.22
ADBE	-7.89	0.04	7.64	-0.33	6.34	2.52	0.18	JCI	-8.43	0.00	5.76	-0.47	5.73	2.26	0.25
DIS	-10.86	-0.04	7.79	-0.25	10.09	1.92	0.26	HPQ	-10.54	0.07	14.43	0.17	10.16	3.03	0.21
WMT	-10.63	0.03	7.18	-0.07	13.40	1.39	0.20	AFL	-9.76	0.01	22.09	2.39	60.19	1.94	0.30
CSCO	-7.37	0.05	10.31	0.12	9.95	1.85	0.26	BAX	-7.94	0.01	9.35	0.14	9.69	1.83	0.20
ACN	-5.77	0.06	12.11	0.43	12.11	1.67	0.27	ED	-14.32	0.03	14.27	-0.80	30.17	1.74	0.13
CRM	-8.40	0.00	11.95	-0.12	7.09	3.14	0.16	KMI	-16.67	-0.09	11.03	-0.57	12.47	3.43	0.19
DHR	-10.99	0.01	6.43	-0.43	10.48	1.66	0.25	SPG	-22.80	-0.07	18.53	-1.89	34.21	3.74	0.15
BMY	-7.77	-0.02	6.01	-0.42	6.32	2.19	0.18	DVN	-16.12	-0.07	17.29	0.07	6.48	8.14	0.12
NEE	-7.19	0.04	9.29	0.16	10.88	1.86	0.14	PH	-10.26	0.00	9.23	-0.09	7.22	2.90	0.23
PM	-8.80	0.02	7.25	-0.27	9.44	1.81	0.19	HSY	-12.99	0.05	15.20	0.71	30.58	1.89	0.16
WFC	-9.13	-0.02	8.67	-0.29	8.96	2.35	0.27	PEG	-5.69	0.04	7.34	0.12	6.98	1.96	0.15
T	-7.09	-0.04	6.28	-0.29	7.76	1.39	0.24	BK	-7.60	0.00	7.20	-0.21	6.04	2.15	0.27
INTC	-7.11	0.06	12.63	0.46	9.31	2.40	0.22	KR	-8.25	0.05	12.19	0.25	8.61	2.82	0.15
QCOM	-10.66	0.01	20.48	1.03	21.00	2.64	0.20	WEC	-20.23	0.06	10.21	-2.16	44.15	2.00	0.13
TXN	-6.86	0.04	10.00	0.05	6.72	2.12	0.24	ILMN	-9.66	0.04	10.75	-0.12	5.68	4.50	0.12
UPS	-6.10	0.01	8.23	0.42	9.32	1.73	0.27	TWTR	-20.93	-0.08	9.78	-0.52	7.64	6.75	0.10
NKE	-6.39	0.01	9.38	-0.03	6.85	2.01	0.22	FAST	-11.75	0.04	8.69	-0.22	8.98	2.63	0.21
UNP	-7.08	0.02	7.37	0.08	6.25	2.23	0.24	PCAR	-7.71	0.01	8.74	0.21	6.91	2.50	0.24
AMGN	-7.31	0.00	7.95	0.17	6.53	2.52	0.17	PPG	-8.36	0.01	7.90	-0.18	8.24	2.17	0.24
IBM	-6.72	0.00	5.58	-0.15	7.18	1.50	0.30	NUE	-6.29	-0.03	8.65	0.09	4.58	3.21	0.20
MDT	-6.95	-0.03	5.36	-0.35	6.25	1.83	0.24	CMI	-10.97	-0.01	11.51	0.11	9.24	2.51	0.23
CVS	-7.00	-0.02	8.03	-0.10	6.59	2.39	0.22	DFS	-23.14	-0.02	14.29	-1.83	30.68	3.18	0.25
AMT	-8.62	0.03	11.17	0.17	10.69	1.98	0.16	HES	-11.10	-0.02	12.78	0.20	5.37	6.01	0.14
LOW	-16.45	0.03	8.14	-1.17	17.24	2.38	0.23	AVB	-11.52	0.01	11.18	-0.26	15.80	2.36	0.14
HON	-7.20	0.00	8.78	0.04	9.54	1.65	0.32	AMP	-9.98	0.00	13.34	0.42	11.70	3.27	0.24
ORCL	-6.32	0.05	8.53	0.23	8.85	1.73	0.25	ROST	-13.78	0.00	9.20	-0.51	10.64	2.88	0.19
INTU	-8.36	0.07	12.53	-0.19	8.85	2.39	0.20	HAL	-17.77	-0.07	11.46	-0.52	9.81	5.79	0.15
COP	-8.97	-0.03	15.00	0.28	7.86	4.05	0.18	WY	-11.52	0.00	16.46	0.11	15.14	2.68	0.22
MS	-7.62	-0.01	9.84	0.10	6.69	2.75	0.25	TSN	-6.84	0.01	20.87	1.14	24.05	2.62	0.17
GS	-7.64	0.00	8.44	-0.01	7.18	2.25	0.26	EBAY	-6.80	0.04	7.53	0.02	5.26	2.53	0.19
LMT	-8.54	-0.01	11.90	0.14	14.39	1.76	0.23	EQR	-8.66	0.01	8.99	-0.23	9.47	2.40	0.15
SCHW	-8.24	0.02	10.87	-0.03	6.32	3.11	0.22	OKE	-27.77	-0.06	24.06	-1.31	27.87	5.55	0.17
CAT	-10.19	0.01	6.38	-0.19	6.77	2.50	0.24	GLW	-12.40	0.04	10.96	0.01	10.51	2.40	0.25
SBUX	-7.22	0.02	8.21	-0.11	7.24	1.97	0.23	DTE	-8.51	0.03	11.37	0.12	12.87	1.94	0.14
C	-8.03	-0.05	11.12	-0.13	8.17	2.74	0.25	DHI	-10.13	-0.01	7.54	-0.38	5.99	4.07	0.16
PLD	-11.04	0.03	6.80	-0.63	9.78	2.12	0.18	EIX	-13.07	0.04	8.85	-0.91	15.05	2.57	0.13
ADP	-6.80	0.06	11.18	0.26	12.83	1.96	0.27	ROK	-11.84	0.03	9.01	-0.04	8.80	2.80	0.22
MDLZ	-5.46	0.01	5.91	0.15	6.39	1.66	0.22	FITB	-8.98	-0.01	14.19	0.29	9.27	3.39	0.22
AXP	-8.20	-0.02	13.22	0.37	13.54	2.04	0.29	STT	-7.78	0.02	12.35	0.03	7.86	2.92	0.24
CI	-9.84	0.02	12.03	0.12	8.12	3.07	0.16	AEE	-12.77	0.05	14.06	-0.07	18.98	1.99	0.13
NFLX	-9.71	0.02	11.32	0.11	5.50	4.58	0.12	TSCO	-6.78	0.03	8.02	0.12	5.05	3.22	0.16
ZTS	-10.28	0.03	10.52	0.06	10.12	2.27	0.18	ETR	-8.65	0.02	11.52	-0.01	10.88	2.09	0.13
DUK	-10.41	0.02	9.12	-0.10	13.08	1.58	0.15	LYB	-12.58	-0.05	14.36	0.08	8.66	3.71	0.19
CB	-7.19	0.02	8.01	0.20	9.70	1.82	0.26	LH	-14.53	0.00	10.62	-0.71	15.39	2.76	0.17
DE	-7.16	0.04	9.45	0.20	6.94	2.41	0.23	HIG	-18.60	0.02	25.57	2.02	61.48	2.69	0.25
MMC	-7.93	0.05	8.60	-0.10	10.36	1.46	0.30	LUV	-11.62	-0.06	8.57	-0.39	6.60	4.06	0.17
GILD	-8.17	-0.05	7.37	-0.01	6.37	2.79	0.16	FE	-19.51	0.02	13.14	-1.02	28.23	2.45	0.14
BA	-16.43	-0.07	10.55	-0.73	13.21	3.69	0.22	ABC	-9.03	0.01	10.24	-0.14	8.20	3.01	0.17
AMAT	-8.71	0.01	9.63	-0.11	5.43	3.58	0.17	VTR	-22.39	-0.02	15.62	-1.88	33.43	3.80	0.13
MO	-12.16	0.01	8.66	-0.87	13.08	1.94	0.19	LEN	-9.45	-0.03	13.31	0.16	8.19	4.25	0.16
SO	-9.60	0.04	14.33	0.62	22.97	1.59	0.15	PPL	-11.73	-0.01	8.56	-0.47	11.61	1.89	0.16
CCI	-8.06	0.02	10.58	0.34	9.68	1.96	0.15	FANG	-15.95	-0.02	16.24	-0.03	6.96	9.04	0.10
MMM	-10.29	-0.01	8.31	-0.59	12.24	1.51	0.30	CMS	-12.54	0.05	10.55	-0.33	15.99	1.85	0.14
CME	-12.30	0.03	10.26	-0.55	12.70	2.10	0.19	VMC	-14.41	-0.02	15.82	0.14	12.92	3.62	0.16
GE	-9.83	-0.09	11.12	0.00	7.57	3.27	0.23	DAL	-22.62	-0.11	8.75	-1.22	14.99	4.93	0.17

NOC	-8.33	0.02	10.74	0.02	10.36	2.13	0.21	K	-10.26	0.02	8.84	-0.04	12.27	1.84	0.16
CL	-7.66	0.02	8.78	0.19	11.22	1.48	0.21	PWR	-20.74	0.01	10.70	-1.10	18.18	3.37	0.20
SYK	-9.49	0.01	9.08	-0.66	9.55	2.02	0.22	RF	-12.98	-0.01	12.60	-0.08	8.93	3.77	0.20
TJX	-13.97	-0.01	7.47	-0.74	13.64	2.40	0.22	URI	-12.76	0.00	13.63	-0.15	7.03	5.70	0.16
TGT	-8.40	0.01	8.51	-0.06	7.68	2.20	0.19	IR	-10.83	0.03	9.27	-0.15	9.39	2.92	0.24
PNC	-8.73	0.02	10.45	0.04	8.49	2.41	0.26	SWK	-13.87	-0.01	16.20	0.11	17.34	2.65	0.23
D	-11.74	0.03	13.68	-0.15	22.08	1.68	0.15	MOS	-24.32	-0.08	15.78	-0.50	12.95	5.41	0.15
MU	-10.10	-0.08	9.90	-0.22	4.39	5.54	0.13	CAG	-14.08	0.02	16.91	0.20	20.87	2.39	0.17
USB	-9.23	-0.01	11.23	0.00	11.17	2.04	0.30	EXPD	-7.25	0.06	9.26	-0.13	8.08	2.32	0.19
CSX	-8.86	0.06	9.16	0.18	7.16	2.45	0.23	KEY	-16.25	-0.02	15.34	-0.01	13.77	3.46	0.21
ATVI	-8.95	-0.03	8.28	-0.32	5.65	3.39	0.14	DGX	-9.00	0.02	9.91	-0.17	9.67	2.41	0.17
EOG	-9.58	-0.03	10.77	0.32	5.90	5.06	0.15	PFG	-8.24	0.01	12.99	-0.04	8.76	3.22	0.24
EW	-9.84	0.03	12.16	-0.09	7.55	2.97	0.15	KMX	-12.04	0.03	9.32	-0.19	7.87	3.73	0.18
EL	-6.27	0.02	7.84	-0.09	6.69	2.08	0.20	IP	-8.92	-0.01	10.11	0.15	7.36	2.73	0.22
HUM	-12.89	0.05	18.07	0.91	19.13	3.28	0.13	SWKS	-10.60	-0.01	12.01	-0.03	6.14	4.19	0.15
AON	-10.85	0.03	9.62	-0.34	14.10	1.84	0.23	AKAM	-9.35	0.01	16.34	0.24	12.01	2.91	0.17
WM	-11.94	0.03	10.43	-0.08	22.71	1.35	0.25	MRO	-15.93	-0.13	13.86	-0.01	5.85	8.54	0.12
FIS	-7.64	0.02	8.55	-0.18	7.22	2.05	0.24	GRMN	-12.40	0.00	8.98	-0.49	10.33	2.49	0.20
DG	-7.82	0.03	11.18	0.13	9.04	2.25	0.17	INCY	-15.13	-0.03	11.15	-0.07	5.55	6.98	0.08
FDX	-7.43	-0.02	12.20	0.13	9.18	2.54	0.24	CAH	-8.38	-0.01	7.03	-0.41	5.90	2.83	0.20
FISV	-7.92	0.04	6.02	-0.43	7.52	2.08	0.23	EXPE	-15.83	-0.01	11.21	-0.31	9.75	4.41	0.15
LRCX	-8.61	0.01	9.89	-0.11	6.01	3.86	0.16	VFC	-8.86	-0.01	8.89	-0.33	8.51	2.76	0.21
NSC	-11.58	0.04	10.74	-0.03	9.72	2.54	0.22	NTAP	-10.48	0.06	11.60	0.05	7.00	3.23	0.19
OXY	-20.75	-0.08	15.04	-0.15	12.85	5.96	0.16	STX	-18.73	0.05	10.45	-0.58	9.38	4.33	0.16
ITW	-9.88	0.04	11.95	0.08	13.37	1.88	0.30	BBY	-11.71	0.04	9.28	-0.30	6.31	3.90	0.16
ETN	-12.89	-0.01	14.45	-0.24	17.40	2.31	0.27	IRM	-11.12	-0.01	9.36	-0.44	7.92	2.77	0.16
GD	-6.75	0.00	7.77	-0.03	7.10	1.92	0.25	AES	-13.20	0.02	15.84	0.04	10.86	3.39	0.15
PXD	-9.52	-0.01	11.99	0.07	5.06	5.79	0.13	WDC	-11.88	-0.07	10.74	-0.26	5.19	5.27	0.15
AEP	-12.96	0.03	10.71	-0.50	17.81	1.79	0.14	OMC	-11.46	-0.03	6.30	-0.43	9.62	2.47	0.21
NEM	-15.04	0.00	17.75	0.15	10.57	4.07	0.05	CHRW	-7.78	0.05	7.02	-0.14	5.64	2.39	0.17
GM	-9.35	-0.08	9.22	0.04	6.19	3.24	0.20	MAS	-11.01	0.02	9.62	-0.36	8.37	2.65	0.22
SLB	-11.20	-0.06	13.08	0.12	7.19	4.49	0.17	KIM	-14.96	-0.03	14.41	-0.46	12.36	4.21	0.16
EMR	-12.18	-0.02	12.81	-0.08	13.64	2.21	0.28	CTXS	-7.39	0.03	11.36	0.26	8.18	2.72	0.16
MCK	-9.74	0.00	11.61	-0.18	7.79	3.29	0.17	LVS	-13.70	-0.06	10.66	-0.09	6.86	4.19	0.16
SRE	-14.09	0.01	10.12	-0.28	18.50	2.13	0.15	UAL	-29.92	-0.12	10.69	-1.44	18.30	6.88	0.14
KMB	-7.48	0.03	9.15	-0.08	10.74	1.68	0.17	NI	-11.93	0.05	13.39	-0.01	14.23	2.16	0.14
GIS	-10.25	0.05	10.71	0.06	12.66	1.79	0.18	L	-18.31	0.00	9.48	-1.16	27.55	2.29	0.27
KLAC	-12.90	0.05	12.59	-0.08	9.15	3.37	0.17	EMN	-7.89	0.00	8.73	0.03	5.86	3.52	0.20
MPC	-20.76	-0.04	14.51	-0.39	10.82	5.72	0.15	HST	-7.30	-0.04	12.49	0.35	8.32	3.96	0.18
F	-7.36	-0.09	10.00	0.17	6.62	3.10	0.22	APA	-35.82	-0.13	15.06	-1.24	19.90	8.83	0.12
MET	-12.33	-0.02	9.93	-0.32	9.58	2.60	0.26	MGM	-23.15	-0.06	10.36	-1.14	14.75	5.99	0.16

The table reports the sample moments of the returns used in Section 4 and time series average realized variances and correlations with the other assets. Minimum, mean and maximum of returns are scaled by 100 and average realized variances by 10^4.