Journal of Financial Econometrics, 2024, Vol. 22, No. 1, 94–118
https://doi.org/10.1093/jjfinec/nbac029
Advance Access Publication Date: 8 August 2022
Article

An Enhanced Factor Model for Portfolio Selection in High Dimensions*

Fangquan Shi[1], Lianjie Shu [2], and Xinhua Gu[2]

[1]School of Finance, Nanjing Audit University, Jiangsu, China and [2]Faculty of Business Administration, University of Macau, Macau, China

Address correspondence to Lianjie Shu, Faculty of Business Administration, University of Macau, Macau, China, or e-mail: ljshu@um.edu.mo.

Received January 26, 2021; revised July 3, 2022; editorial decision July 27, 2022; accepted July 27, 2022

Abstract

This article extends Fama and French (FF) models of observed factors by introducing latent factors (LFs) to further extract information from FF residual returns. A diagonally dominant (DD) rather than a diagonal or sparse matrix structure is adopted in this study to estimate remaining covariance between disturbance terms. Such an enhanced factor (EF) model provides a more comprehensive analysis for portfolio selection in high dimensions and also has certain advantages of estimation stability and computational efficiency. It is shown that the proposed EF–DD approach achieves overall better performance than competing models in terms of portfolio variance and the net Sharpe ratio.

Key words: covariance matrices, diagonally dominant structures, factor models, Fama and French models, latent factors, minimum variance portfolios (MVPs)

JEL classification: C13, C58, G11

Since estimation errors of mean returns have a more serious impact on portfolio weights than those of asset covariances (Merton 1980; Kan and Zhou 2007), much of finance research has deviated from Markowitz's mean–variance strategies and turned to global minimum-variance portfolios (MVPs) (Jagannathan and Ma 2003; Kolm, Tütüncü, and Fabozzi 2014). Yet the problem remains because of estimation error for the covariance matrix Σ of asset returns, especially in the case of high dimensions. Numerous approaches have been proposed to deal with this problem and they can be grouped into two categories: factor-free and factor-based models.

The first category is motivated by three ideas for error reduction under no factor analysis. The first idea is to strike a balance between bias and variance by shrinking sample covariances toward a target matrix (Ledoit and Wolf 2003, 2004, 2017). The second idea is to improve

* The authors thank the editor, an associate editor, and two referees for their valuable comments.

the performance of portfolios by imposing no-short-sale or norm constraints on their weights that would otherwise have taken large positions (positive or negative) if derived from sample covariance matrices (Jagannathan and Ma 2003; DeMiguel et al. 2009). The third idea is to avoid computing imprecisely the inverse covariance matrix Σ^{-1} by shrinking it directly and estimating it through graphical lasso (Glasso) treatment (Goto and Xu 2015).

The second category uses a small number of common factors to explain co-movements of a large universe of correlated asset returns. Such mapping from a large asset space to a much smaller factor space makes it possible to estimate efficiently various parameters arising from high-dimensional data sets. This method is practically useful because financial investors routinely handle a large number of assets (Bai and Ng 2002). Factor models have a long history in finance. A sample of recent research includes Han (2006); Rajaratnam, Massam, and Carvalho (2008); Page and Taborsky (2011); Callot, Kock, and Medeiros (2017); Bender, Blackburn, and Sun (2019); De Nard, Ledoit, and Wolf (2021); Ding, Li, and Zheng (2021); Giglio, Liao, and Xiu (2021); and Li et al. (2021).

Various models of the two categories are proposed in the literature and it is of interest to see how their performance stacks up against each other. However, there is a lack of comprehensive comparison between these models and our work intends to fill this gap. We focus on factor analysis, with asset returns split up into factor-related returns and asset-specific returns (residual terms). Two fundamental issues need to be addressed when using factor models to estimate Σ: what factors should be included to have asset returns well explained? and what structure for the residual returns should be specified to get covariance matrices accurately estimated? Two types of factors are separately used in the literature and they are either unobservable latent factors (LFs) or observed factors as in Fama and French (FF) (1993, 2015, 2018), with some residual structures assumed for estimation yet remaining to be improved.

The LFs and FF factors are organically blended in our enhanced factor (EF) model as a unified framework. LF models improve estimation accuracy for noisy data but with no intuition for financial trading (Bai and Ng 2006). FF models are preferred by both researchers interested in economic fundamentals and practitioners looking for intuitive indicators (Aharoni, Grundy, and Zeng 2013; Feng, Giglio, and Xiu 2020), yet there is no consensus over the nature or number of observed factors to be used except the market index (De Nard, Ledoit, and Wolf 2021). It is impossible to rely only on observed factors for fully describing complex variations in asset returns, especially when data sets are large and noisy. Some of these variations are left unexplained as unobservable residuals and this motivates us to utilize LFs to extract information further from FF residual returns. In such an EF model, the LFs make up the weakness of pre-specified FF factors and play a complementary role in accounting for asset returns.

Two typical assumptions are made in different factor models for the structure of disturbance covariances yet with some problems. A diagonal structure is assumed in exact factor models, allowing no correlations among residuals (FF 2018). A sparse structure, capturing those correlations, is assumed in approximate factor models, with off-diagonals of residual covariances shrunk toward zero (Fan, Liao, and Mincheva 2013). Estimation instability might arise under shrinkage since high thresholding is required to make the covariance matrix positive definite (Fan, Liao, and Liu 2016). This issue can be addressed by applying nonlinear shrinkage (NLS) to residual covariances (Ledoit and Wolf 2017), but this method tends to be computationally expensive. To alleviate this problem, our article suggests that a new structure, called the diagonal-dominance (DD) structure, should be employed for residual or disturbance covariances. Under this DD structure, each diagonal entry of a covariance matrix is large relative to other entries in the same row, so that computation complexity and estimation stability can be balanced.

The DD structure has recently been proposed in statistics but not applied yet to finance research. Our study invokes this new structure for portfolio finance because of its three "nice" properties (Ke, Xue, and Yang 2020). First, high-dimensional portfolios usually involve strongly correlated assets. The DD setting is a natural generalization of the diagonal structure due to its permission for mild correlation between idiosyncratic disturbances. Second, the DD structure ensures well-conditioning for the inverse of disturbance covariances because a strict DD matrix is non-singular (Horn and Johnson 2012). The DD matrix enhances estimation stability without relying on high thresholding for its positive definiteness. Third, the DD structure is computationally efficient as it is nearly tuning free, with its only tuning parameter set equal to one. These "nice" properties make it very attractive to apply the DD treatment to portfolio selection in high dimensions.

The aforementioned two fundamental issues are addressed in our EF model, with three enhancements achieved effectively. First, factor returns in the EF model are attributed to both LFs and FF factors. The LFs are introduced to improve the prediction of an FF model, and its residual returns are further decomposed into LFs-driven returns and idiosyncratic returns. Second, the EF model, powered by a DD structure, can deal well with interactive effects when unobservable common factors have heterogeneous impacts on cross sections of idiosyncratic errors. Correlation and heteroskedasticity are jointly taken into account to promote efficient estimation (Bai 2009; Bai and Liao 2017). Third, numbers of FF factors and LFs in the EF model are chosen in a scientific rather than *ad hoc* manner. The number of FF factors is often pre-specified in the finance literature, but now selected in the EF model via cross-validation, with the number of LFs set in a data-driven manner. Such an adaptive way to choose the two types of factors in the EF-DD setting can lead to a stable and precise estimate of Σ. This setting with the three enhancements is expected to produce better-performing portfolios than other competing estimators. As shown later, this expectation turns out to be true when estimated with real-world data sets.

This article proceeds as follows. Section 1 proposes the EF model along with its factor choice and estimation strategy. Section 2 discusses the estimation of disturbance covariances under different structures together with numerical simulation for structure selection. Section 3 sets up the scheme for comparing performance among various portfolio strategies according to certain standard criteria. Section 4 reports empirical results on the out-of-sample performance of the proposed approach and its competitors. Some concluding remarks are given in Section 5.

1 The EF Model

1.1 Formulation of the Model

Consider a portfolio composed of p assets, with y_{it} denoting the random return on asset $i = \{1, 2, \ldots, p\}$ at time $t = \{1, 2, \ldots, n\}$. A typical FF factor model takes the form of

$$y_t = Af_t^o + g_t, \tag{1}$$

where $y_t = (y_{1t}, \dots, y_{pt})'$ is a $p \times 1$ vector of asset returns, $f_t^o = (f_{1t}^o, \dots, f_{mt}^o)'$ is an $m \times 1$ random vector of m observed common factors that are independent of each other and have pervasive effects on all assets, and A is a $p \times m$ loading matrix of the observed factors. Additionally, $g_t = (g_{1t}, \dots, g_{pt})'$ is a $p \times 1$ vector of residual returns (or asset-specific returns) after removing the observed effects on the return data y_t.

Under the classical model assumptions, the covariance matrix Σ of y_t is given by

$$\Sigma = \Sigma_A + \Sigma_g,$$

where $\Sigma_A = A\Sigma_{f^o}A'$, Σ_{f^o} and Σ_g are covariance matrices of f_t^o and g_t, respectively. In the FF model of Equation (1), the factor loadings are obtained by the ordinary least squares (OLS) estimator. The covariance matrix Σ_{f^o} is estimated by the sample covariance matrix of f_t^o. The estimation of the covariance matrix Σ_g depends on the structure of residual returns g_t. Specifically, Σ_g is assumed to be a diagonal matrix in an exact factor model, but a bounded matrix by L^1 or L^2 in an approximate factor model (De Nard, Ledoit, and Wolf 2021).

The observed FF factors alone may not account well for asset returns that are usually noisy with complex variations. It is thus reasonable to invoke unobservable LFs to capture the remaining part of those variations after the effects of the observed factors have been taken out. Suppose the residual returns g_t are governed by LFs in such a way as

$$g_t = Bf_t^u + u_t,$$

where $f_t^u = (f_{1t}^u, \dots, f_{kt}^u)'$ is a $k \times 1$ vector of LFs that are common to all assets, B is a $p \times k$ loading matrix for the LFs, and $u_t = (u_{1t}, \dots, u_{pt})'$ is a $p \times 1$ vector of disturbances (or error terms). As usual, $k \ll p$ for high-dimensional data.

LFs are unknown and must be approximated from the data. The two popular methods to approximate the LFs are the maximum-likelihood estimator and principal components analysis (PCA) (Bai and Shi 2011). The PCA is used in our work since this method provides consistent and concise estimation for LFs under certain regular conditions (Fan, Liao, and Mincheva 2013). Note that B and f_t^u may not be separately identifiable even though $B\mathrm{Cov}(f_t^u)B'$ ($\triangleq L$) can be asymptotically identified. The reason is that, in computing L, the pair (B, f_t^u) is equivalent to the pair (BH^{-1}, Hf_t^u) for any non-singular square matrix H. To resolve such ambiguity, assume $\mathrm{Cov}(f_t^u) = I$ for normalization without loss of generality; also, suppose the columns of B are orthogonal to each other. The covariance matrix Σ_g then has a canonical decomposition

$$\Sigma_g = BB' + \Sigma_u, \tag{2}$$

where Σ_u is the covariance matrix of disturbances or error terms u_t. The factor models with both observable and LFs have also been considered in other contexts; see, for example, Bai and Liao (2017) and Giglio, Liao, and Xiu (2021).

The effects of factors f_t^u are pervasive since eigenvalues of the $k \times k$ matrix, $B'B/p$, are uniformly bounded away from zero and approach infinity as $p \to \infty$. Let $\{\tilde{b}_1, \dots, \tilde{b}_k\}$ be the k columns of B, $\{\lambda_1, \dots, \lambda_k\}$ the k largest eigenvalues of Σ_g, and $\{\xi_1, \dots, \xi_k\}$ the corresponding eigenvectors. It is proven that leading eigenvectors of Σ_g are approximately equal to normalized columns of B as $p \to \infty$, that is, $\tilde{b}_j \approx \sqrt{\lambda_j}\xi_j$. Moreover, the ith LF f_{it}^u is approximately equal to $\xi_i' g_t/\sqrt{\lambda_i}$, that is, $f_{it}^u \approx \xi_i' g_t/\sqrt{\lambda_i}$, which is the principal component of

g_t apart from a constant multiplier. Furthermore, the covariance structure of the LF loadings is approximated by using

$$BB' = \Sigma_B = \sum_{i=1}^{k} \tilde{b}_j \tilde{b}_j' = \sum_{i=1}^{k} \lambda_i \xi_i \xi_i'. \tag{3}$$

With all above analyses pieced together, the covariance matrix Σ of asset returns y_t is decomposed into the following three parts:

$$\Sigma = \Sigma_A + \Sigma_B + \Sigma_u, \tag{4}$$

where the first part Σ_A and the second part Σ_B are, respectively, the observed and unobservable common components of Σ, and the third part Σ_u is the idiosyncratic components of Σ.

1.2 Selection of k

Given the number m of observed factors that are set at $m = \{1, 3, 5\}$ in existing FF models, our EF model involves the selection of k LFs approximated by principal components. The selection of k can be accomplished through the procedures proposed by Bai and Ng (2002) and Alessi, Barigozzi, and Capasso (2010). The work of Alessi, Barigozzi, and Capasso (2010) is a variant of Bai and Ng (2002) aiming for a reliable estimate of k. An additional tuning constant is introduced in the penalty function by Alessi, Barigozzi, and Capasso (2010) as in Hallin and Liška (2007). Since a more complex procedure from Alessi, Barigozzi, and Capasso (2010) cannot improve portfolio performance, we will adopt the procedure of Bai and Ng (2002) to select k LFs.

Specifically, denote by $\widehat{G} = (\widehat{g}_1, \ldots, \widehat{g}_n)'$ the $n \times p$ sample matrix of residual returns, with n denoting the number of observations. Define $\widehat{F}_k^u = (\widehat{f}_1^u, \ldots, \widehat{f}_n^u)'$ as an $n \times k$ matrix estimate for the LFs and $\widehat{B}$ as the $p \times k$ matrix estimate of unobservable factor loadings. The loss function based on a certain information criterion is specified as in Bai and Ng (2002)

$$\mathrm{IC}(k) = \log\left(\frac{1}{pn} ||\widehat{G} - \widehat{F}_k^u \widehat{B}'||_F^2\right) + kf(n, p),$$

where $||\widehat{G} - \widehat{F}_k^u \widehat{B}'||_F^2$ is the sum of squared disturbances from regressing $\widehat{G}$ on $\widehat{F}_k^u$, and

$$f(n, p) = \frac{p+n}{pn} \log\left(\frac{pn}{p+n}\right)$$

is the penalty function for overfitting. The optimal value of k is determined by minimizing the loss function

$$k^* = \arg\min_{k \geq 0} \mathrm{IC}(k).$$

1.3 Estimation of the EF Model

The matrix of observed factor loadings is estimated with the OLS method from a sample of n observations as follows:

$$\widehat{A} = [(F^{o\prime} F^o)^{-1} F^{o\prime} Y]',$$

where $Y = (y_1, \ldots, y_n)'$ is an $n \times p$ data matrix of asset returns y_t and $F^o = (f_1^o, \ldots, f_n^o)'$ is an $n \times m$ data matrix of observed factors f_t^o. The covariance matrix of Af_t^o is then estimated by

$$\widehat{\mathrm{Cov}}(Af_t^o) = \widehat{A}\widehat{\Sigma}_{f^o}\widehat{A}', \tag{5}$$

where $\widehat{\Sigma}_{f^o} = \frac{1}{n}\sum_{t=1}^n (f_t^o - \overline{f}^o)(f_t^o - \overline{f}^o)'$ is the sample covariance matrix of observed factors f_t^o whose sample mean is defined as $\overline{f}^o = \frac{1}{n}\sum_{t=1}^n f_t^o$.

Denote by $\widehat{g}_t = y_t - \widehat{A}f_t^o$ the estimated residuals from regressing y_t on f_t^o. The sample covariance matrix of $\widehat{g}_t$ has the eigen-decomposition of $S_g = \sum_{i=1}^p \widehat{\lambda}_i \widehat{\xi}_i \widehat{\xi}_i'$, where $\widehat{\lambda}_i$ and $\widehat{\xi}_i$ are, respectively, the ith largest eigenvalue and corresponding eigenvector of S_g. Denote the estimate of the first k LFs by $\widehat{f}_t^u = [\widehat{\xi}_{1t}'\widehat{g}_t/\sqrt{\widehat{\lambda}_1}, \ldots, \widehat{\xi}_{kt}'\widehat{g}_t/\sqrt{\widehat{\lambda}_k}]'$. Regressing $\widehat{g}_t$ on the first k LFs under $k < p$, one obtains the OLS estimate $\widehat{B}$ and the covariance matrix estimate of Bf_t^u as follows:

$$\widehat{\mathrm{Cov}}(Bf_t^u) = \widehat{B}\widehat{\Sigma}_{f^u}\widehat{B}' = \sum_{i=1}^k \widehat{\lambda}_i \widehat{\xi}_i \widehat{\xi}_i'. \tag{6}$$

With covariance matrices of the observable and unobservable factors estimated from Equations (5) and (6), we proceed to estimate the covariance matrix Σ_u of idiosyncratic disturbances u_t. Denote the sample disturbance covariance matrix by

$$R_u = (\widehat{r}_{ij})_{p \times p} = \sum_{i=k+1}^p \widehat{\lambda}_i \widehat{\xi}_i \widehat{\xi}_i'.$$

Let $\widehat{\Sigma}_u(R_u)$ be an estimate of Σ_u based on R_u (to be discussed further in the next section). As an analog to Equation (4), the estimate for the covariance matrix Σ of asset returns y_t follows from adding up estimates of all the three components

$$\widehat{\Sigma} = \widehat{A}\widehat{\Sigma}_{f^o}\widehat{A}' + \sum_{i=1}^k \widehat{\lambda}_i \widehat{\xi}_i \widehat{\xi}_i' + \widehat{\Sigma}_u(R_u). \tag{7}$$

2 Estimation of Σ_u

2.1 Different Structural Assumptions for Σ_u

The disturbance covariance matrix Σ_u plays a vital role in estimating the covariance matrix Σ of asset returns and its inverse Σ^{-1}. Different assumptions on the structure of Σ_u have been made in the literature. We will show that the new DD structure for Σ_u is efficiently good for estimating the EF model in high dimensions.

First, a diagonal structure is assumed for Σ_u in exact factor models, where disturbances u_t are cross-sectionally uncorrelated. The estimated covariance matrix $\widehat{\Sigma}_u(R_u)$ is a diagonal one determined by diagonal elements $\widehat{r}_{ii}$ of the sample disturbance covariance matrix R_u

$$\widehat{\Sigma}_{u,\mathrm{diag}}(R_u) = \mathrm{diag}(\widehat{r}_{11}, \ldots, \widehat{r}_{pp}).$$

This assumption seems too restrictive for estimating Σ_u because some cross-sectional correlations are likely to remain among the disturbances even after taking out the effects of common factors.

Second, a sparse structure for Σ_u is specified to relax the diagonal assumption. This structure is commonly used in approximate factor models (Bai and Ng 2002). Fan, Liao, and Mincheva (2013) suggest a principal orthogonal complement thresholding (POET) estimator to estimate Σ_u. A soft thresholding rule is used to shrink off-diagonal entries of sample disturbance covariances toward zero values, generating a sparse estimator as follows:

$$\widehat{\Sigma}_{u,\mathrm{POET}}(R_u) = (\widehat{\sigma}_{ij})_{p\times p}, \quad \widehat{\sigma}_{ij} = \begin{cases} \widehat{r}_{ij}, & i=j \\ \mathrm{sign}(\widehat{r}_{ij})(|\widehat{r}_{ij}| - \tau_{ij})I(|\widehat{r}_{ij}| \geq \tau_{ij}), & i \neq j, \end{cases} \tag{8}$$

where sign$(\cdot)$ is the sign function and $I(\cdot)$ is the indicator function. To be adaptive to the variability of individual entries, τ_{ij} is selected to be $\tau_{ij} = C\omega_n\sqrt{\widehat{r}_{ii}\widehat{r}_{jj}}$, where $C>0$ and $\omega_n = 1/\sqrt{p} + \sqrt{\log p/n}$.

The selection of C is crucial for thresholding to be desirable. The thresholding sparse estimate may be a diagonal matrix if C is sufficiently large, or non-positive definite if C is relatively small. Fan, Liao, and Mincheva (2013) suggest selecting C over an interval $[C_{\min}, C_{\max}]$ by using cross-validation to ensure positive definiteness. The estimated disturbances $\{\widehat{u}_t\}_{t\leq n}$ are randomly divided into two subsets indexed by J_1 and J_2. The sampling size is $n_1 = n \times \left(1 - \frac{1}{\log(n)}\right)$ for J_1 and $n_2 = n - n_1$ for J_2. In the jth splitting, let $\widehat{\Sigma}_u^{\mathrm{Train},j}(C)$ be the thresholding estimator for the training data set $\{\widehat{u}_t\}_{t\in J_1}$, and $\widehat{\Sigma}_u^{\mathrm{Val},j}$ be the sample covariance matrix based on the validation set $\{\widehat{u}_t\}_{t\in J_2}$. The optimal C can be obtained by repeating the splitting procedure for N times and minimizing the cross-validated predictive error

$$C^* = \arg\min_{C_{\min}\leq C\leq C_{\max}} \frac{1}{N}\sum_{j=1}^{N} ||\widehat{\Sigma}_u^{\mathrm{Train},j}(C) - \widehat{\Sigma}_u^{\mathrm{Val},j}||_F^2, \tag{9}$$

where the Frobenius-norm operator is defined as $||M||_F = \{\mathrm{trace}(MM')\}^{1/2}$ for any square matrix M.

Third, the covariance matrix Σ_u of disturbances u_t is assumed in our model to have a DD structure. This covariance structure posits that each diagonal element of Σ_u is large compared with the sum of absolute values of the other entries in the same row, with the correlation matrix of u_t having uniformly small off-diagonal entries. The set of "symmetric c-DD" matrices is defined as

$$\mathrm{SDD}_c^+ = \{D = (d_{ij})_{p\times p}\,|D^T = D\,\mathrm{and}\,d_{jj} \geq c\sum_{i:i\neq j}|d_{ji}|\ \mathrm{forall}\ 1 \leq j \leq p\},$$

where $c>0$, with $c=1$ as a popular choice. Mendoza, Raydan, and Tarazaga (1998) develop an alternating projection algorithm on SDD_c^+. This algorithm is used to obtain the DD estimator of Σ_u denoted by $\widehat{\Sigma}_{u,\mathrm{DD}}$ and then to estimate Σ_g by means of $\widehat{\Sigma}_g = \sum_{i=1}^k \widehat{\lambda}_i\widehat{\xi}_i\widehat{\xi}_i' + \widehat{\Sigma}_{u,\mathrm{DD}}$. However, such a sequential method attaches no importance to the structure of the low-rank matrix $\sum_{i=1}^k \widehat{\lambda}_i\widehat{\xi}_i\widehat{\xi}_i'$ when estimating Σ_u.

To estimate jointly the low-rank matrix and the DD matrix, an iterative projection algorithm based on the DD-PCA is used as in Ke, Xue, and Yang (2020). Specifically, the large covariance matrix Σ_g in Equation (2) is decomposed into a low-rank matrix L $(= \Sigma_B)$ and a DD matrix D $(= \Sigma_u)$, that is, $\Sigma_g = L + D$. Denote the sample counterpart of Σ_g by a $p \times p$ matrix S_g. The joint estimation of L and D is carried out by solving the following problem:

$$\min_{(L,D)} ||S_g - L - D||_F, \text{ subject to } \mathrm{rank}(L) \leq k, \, L = L', \, D \in \mathrm{SDD}_c^+.$$

Denote by $\widehat{L}$ and $\widehat{D}$ the estimates of L and D from the DD-PCA, respectively. The covariance matrices Σ_u and Σ_g are estimated by

$$\widehat{\Sigma}_{u,\mathrm{DD}} = \widehat{D} \qquad \text{and} \qquad \widehat{\Sigma}_g = \widehat{L} + \widehat{D}. \tag{10}$$

The "DD-PCA algorithm" iteratively projects $(S_g - L)$ onto SDD_c^+ to update D and then proceeds to project $(S_g - D)$ onto $\mathcal{L}_\mathcal{K} = \{L : \mathrm{rank}(L) = k\}$ to update L. To reduce the cost of computation, the algorithm replaces the projection onto SDD_c^+ with a projection onto

$$\mathrm{DD}_c^+ = \left\{ D = (d_{ij})_{p\times p} \, |d_{jj} \geq c \sum_{i:i\neq j} |d_{ji}| \text{ for all } \forall j \right\}$$

and conducts a procedure of symmetrization. Given an input matrix S_g and the number k of principal components, the "DD-PCA algorithm" involves the following steps:

- Let $D^{(0)} = 0$ and initialize k and S_g.
- For $t = 1, 2, \ldots,$
  - update $L^{(t)} = \mathcal{P}_{\mathcal{L}_\mathcal{K}}(S_g - D^{(t-1)})$,
  - update $\tilde{A}^{(t)} = \mathcal{P}_{\mathcal{DD}_c^+}(S_g - L^{(t)})$,
  - update $D^{(t)} = (\tilde{A}^{(t)} + (\tilde{A}^{(t)})^T)/2$.
- Stop if the convergence criterion is met.

The calculation of $\mathcal{P}_{\mathcal{L}_\mathcal{K}}$ is to project $[S_g - D^{(t-1)}]$ onto the low-rank matrix to update L. For any symmetric matrix M with an eigenvalue decomposition $M = Q\Lambda Q^T$ where $\Lambda = \mathrm{diag}\{\eta_1, \eta_2, \ldots, \eta_p\}$ and $|\eta_1| \geq |\eta_2| \geq \cdots \geq |\eta_p|$, use the best-rank k-approximation by $\mathcal{P}_{\mathcal{L}_\mathcal{K}}(M) = Q_k \Lambda_k Q_k^T$, where Q_k contains the first k columns of Q and $\Lambda_k = \mathrm{diag}\{\eta_1, \eta_2, \ldots, \eta_k\}$. The calculation of $\mathcal{P}_{\mathrm{DD}_c^+}$ is to project $[S_g - L^{(t)}]$ onto the DD cone. The projection $\mathcal{P}_{\mathrm{DD}_c^+}$ is computed using the Mendoza–Raydan–Tarazaga (MRT) algorithm (Mendoza, Raydan, and Tarazaga 1998), which can be implemented by the function "ProjDD" in the R package "ddpca."

1.2 Stability of the DD Estimator for Σ_u

One advantage of the DD estimator is its induced stability of estimating the inverse covariance matrix. As shown in Ke, Xue, and Yang (2020), $||\widehat{D}^{-1}||$ will have a nice upper bound for $c > 1$ if $\widehat{D}$ is restricted to be symmetric c-DD. This bound prevents $||\widehat{D}^{-1}||$ from exploding and ensures that the DD estimator of Σ_u^{-1} is well conditioned, which in turn improves its stability. While the upper bound for $||\widehat{D}^{-1}||$ applies to the case with $c > 1$, numerical studies also work out well for the case of $c = 1$, as found in Ke, Xue, and Yang (2020) and confirmed in our article.

As proved by Ke, Xue, and Yang (2020), estimation errors for Σ_g are dominated by those for L rather than D; however, estimation errors for Σ_g^{-1} are dominated by those for D^{-1} rather than L^{-1}. A small estimation error for D^{-1} could lead to a large estimation error for Σ_g^{-1}. According to Equation (7), the sum of covariance estimates for the FF factors and LFs forms an estimate for the low rank matrix. Therefore, to obtain $\widehat{\Sigma}^{-1}$ with precision, it is more important to improve estimation accuracy for $\widehat{\Sigma}_u^{-1}$ than for the low rank part $(\widehat{\Sigma}_A + \widehat{\Sigma}_B)$.

The inclusion of LFs explains some variations left unexplained by FF factors and improves estimation accuracy not just for Σ_u but for $\widehat{\Sigma}_u^{-1}$ as well. Although some error may come with the LFs, the net estimation accuracy arises both for $\widehat{\Sigma}_u^{-1}$ and for $\widehat{\Sigma}^{-1}$. Hence, the EF model is expected to improve portfolio performance out of sample, when compared with classical FF models.

In the following, some simulations were performed to demonstrate the efficiency of the DD structure in reducing estimation errors, compared with other structures. Following Ke, Xue, and Yang (2020), we generate simulation data from $g_t = Bf_t^u + u_t$, where f_t^u is a 5×1 vector of five LFs, B is a $p \times 5$ matrix of factor loadings, and u_t is a $p \times 1$ vector of disturbance terms. The factors are randomly drawn from an independent and identically distributed (i.i.d.) normal distribution with mean zero and unit variance, that is, $N(0, 1)$. The factor loadings are also drawn from the i.i.d.$N(0, 1)$. The disturbance vectors $(u_1, \ldots, u_n)$ are generated from a multivariate normal distribution with zero mean vector and covariance matrix Σ_u, where $\Sigma_u = (\mathrm{diag}(H))^{1/2}\Theta(\mathrm{diag}(H))^{1/2}$ and $H = (h_1, \ldots, h_p)$. The elements $(h_1, \ldots, h_p)$ are generated from a uniform distribution $U[0.5, 1.5]$. The correlation matrix $\Theta = (\theta_{ij})_{p\times p}$ is an exponential decay matrix with $\theta_{ij} = \rho^{|i-j|}$ for $i \neq j$ and $\theta_{ij} = 1$ otherwise, where the decay base is set at $\rho = 0.5$.

For each simulation data set of n observations, the LF model is used to estimate the covariance matrix of g_t expressed as follows:

$$\widehat{\Sigma}_g = \sum_{i=1}^{k} \widehat{\lambda}_i \widehat{\xi}_i \widehat{\xi}_i' + \widehat{\Sigma}_u,$$

where the true covariance matrix of g_t is $\Sigma_g = B\Sigma_{f^u}B' + \Theta$. The difference $||\widehat{\Sigma}_g - \Sigma_g||_F$ between the true covariance matrix and its estimate is calculated to quantify estimation errors of Σ_g. Errors in estimating Σ_g^{-1}, Σ_u, and Σ_u^{-1} are measured in the same way.

Table 1 displays the estimation errors for Σ_u, Σ_u^{-1}, Σ_g, and Σ_g^{-1} across 1000 simulations under three different structures of Σ_u, with k principal components applied to the cases of $p = (50, 100)$ for $n = 60$. As seen from Table 1, the DD estimator of disturbance covariances has smaller errors in estimating Σ_u^{-1} and Σ_g^{-1} than the diagonal and sparse estimators in almost all cases with the only exception for $k = 5$ and $p = 100$. As expected, the differences in estimation error across three structures of Σ_u are very slim for Σ_g but a bit large for Σ_g^{-1}. This arises because the estimation error of Σ_g is dominated by the error of a lower-rank matrix that remains the same for the three structures, whereas the estimation error of Σ_g^{-1} is affected by the errors of Σ_u^{-1} that are quite different among these structures.

The condition number of a matrix, defined as the ratio of its maximum to minimum eigenvalues, indicates the numerical stability of its inverse. A matrix and its inverse have

Table 1. Estimation errors of Σ_g, Σ_g^{-1}, Σ_u, and Σ_u^{-1}

	$p=50$				$p=100$			
Structure of Σ_u	Σ_g	Σ_g^{-1}	Σ_u	Σ_u^{-1}	Σ_g	Σ_g^{-1}	Σ_u	Σ_u^{-1}
				$k=5$				
Diagonal	38.51	7.08	6.25	8.11	82.34	11.40	8.30	12.10
	(5.89)	(0.13)	(0.06)	(0.15)	(13.18)	(0.16)	(0.04)	(0.18)
Sparse	*38.24*	6.55	4.87	8.41	*82.16*	*8.79*	6.20	*9.58*
	(5.87)	(1.17)	(0.18)	(1.95)	(13.19)	(0.76)	(0.19)	(0.91)
DD	38.29	*6.03*	*4.69*	*7.02*	*82.16*	10.03	*6.17*	10.69
	(5.92)	(0.21)	(0.16)	(0.24)	(13.22)	(0.22)	(0.14)	(0.24)
				$k=7$				
Diagonal	38.60	6.64	6.66	7.81	82.55	10.75	8.58	11.47
	(5.84)	(0.11)	(0.07)	(0.17)	(13.13)	(0.13)	(0.06)	(0.15)
Sparse	*38.39*	7.72	6.05	11.19	*82.42*	9.61	7.16	10.68
	(5.82)	(1.05)	(0.14)	(1.88)	(13.13)	(0.90)	(0.18)	(1.12)
DD	38.43	*5.68*	*5.71*	*6.95*	*82.42*	*9.41*	*6.82*	*10.05*
	(5.86)	(0.17)	(0.14)	(0.23)	(13.16)	(0.20)	(0.15)	(0.22)
				$k=10$				
Diagonal	38.57	7.11	7.31	9.06	82.71	10.33	9.14	11.24
	(5.81)	(0.28)	(0.07)	(0.45)	(13.08)	(0.12)	(0.08)	(0.17)
Sparse	*38.42*	10.77	7.25	19.24	*82.61*	12.73	8.42	15.06
	(5.80)	(1.41)	(0.10)	(3.35)	(13.08)	(1.58)	(0.14)	(2.15)
DD	38.46	*6.12*	*6.87*	*8.75*	*82.61*	*9.10*	*7.92*	*9.93*
	(5.82)	(0.26)	(0.10)	(0.45)	(13.10)	(0.18)	(0.14)	(0.22)

Notes: This table displays means and standard deviations (values in parentheses) of estimating errors for Σ_g, Σ_g^{-1}, Σ_u, and Σ_u^{-1} across 1000 simulations under $k = (5, 7, 10)$, $n=60$, $p = (50, 100)$, and $\rho = 0.5$. The estimators are based on different (diagonal, sparse, and DD) structural assumptions of covariance matrix Σ_u.

the same condition number. A matrix will be ill-conditioned and nearly singular if its condition number is extremely high. A matrix that is not invertible has a condition number going to infinity (Won et al. 2013). On these occasions, the computation of its inverse is prone to large numerical errors. Condition numbers of a matrix can be compared between its estimators to shed light on their stability. Condition numbers for $\widehat{\Sigma}_u$ and $\widehat{\Sigma}_g$ are presented in Table 2, which are computed from the above simulation data. Clearly, the mean and standard deviation of condition numbers of $\widehat{\Sigma}_g$ and $\widehat{\Sigma}_u$ are much smaller for the DD and diagonal structures than for the sparse structure. The DD and diagonal structures are then expected to have more stable estimation than the sparse structure.

3 Performance Evaluation: Setup

3.1 Data Sets and Performance Metrics

The data sets used in our work for performance evaluation/comparison are publicly available from the recent literature. As listed in Table 3, our data sets include: (i) 100 portfolios

Table 2. Condition numbers of $\widehat{\Sigma}_u$ and $\widehat{\Sigma}_g$

Structure of Σ_u	$p=50$				$p=100$			
	Σ_g		Σ_u		Σ_g		Σ_u	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
				$k=5$				
Diagonal	194.71	35.16	4.32	0.79	462.42	81.20	5.05	0.77
Sparse	428.76	109.29	18.66	7.19	896.99	182.52	15.41	3.47
DD	205.82	34.45	7.27	1.07	453.01	72.95	7.74	0.92
				$k=7$				
Diagonal	227.14	40.37	3.94	0.72	518.10	92.51	4.82	0.76
Sparse	468.93	101.18	15.90	4.59	967.26	195.59	13.51	2.82
DD	235.52	38.75	6.33	0.91	502.29	81.95	7.10	0.84
				$k=10$				
Diagonal	287.23	49.84	3.61	0.66	613.63	110.30	4.63	0.77
Sparse	570.80	120.54	16.38	5.34	1193.94	256.62	13.87	3.53
DD	292.44	47.36	5.64	0.72	588.26	95.15	6.56	0.80

Notes: This table displays means and standard deviations of the condition numbers for $\widehat{\Sigma}_u$ and $\widehat{\Sigma}_g$ across 100 simulations under $k = (5, 7, 10)$, $n=60$, $p = (50, 100)$, and $\rho = 0.5$. The estimators are based on different (diagonal, sparse, and DD) structural assumptions of Σ_u.

formed on size and investment (100IN), (ii) 100 FF portfolios (100FF), (iii) 100 portfolios formed on size and operating profitability (100OP), (iv) 100 common stocks selected randomly from the S&P 500 index (100SP), (v) the combination of the 100OP and 100SP (200OS), (vi) the combination of the 100FF and 100SP portfolios (200FS), (vii) 235 stocks selected from the New York Stock Exchange (235NY), and (viii) 500 stocks selected from the Center for Research in Security Prices (CRSP) Stock Database (500CP). The return data in 100SP, 235NY, and 500CP are downloaded from the CRSP. The remaining data are extracted from the website of Ken French (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Missing data are imputed using the predictive mean matching method.

The out-of-sample performance of asset allocation is often assessed using three criteria: (i) out-of-sample portfolio variance (VR), (ii) out-of-sample portfolio Sharpe ratio (SR), and (iii) portfolio turnover (TO) in terms of trading volume. The VR measures the risk of portfolio formation, the SR indicates the risk-adjusted return on portfolio investment, and the TO signifies the economical stability of portfolio selection since lower TO for rebalancing portfolios tends to imply a lower cost of transaction.

Following DeMiguel et al. (2009), we use a "rolling-window" procedure to compare portfolio performance among different estimators. The length n of window is chosen to span $n = 120$ consecutive data points and this window corresponds to a 10-year period for monthly data. Various portfolio weights are computed using the return data over the window. Computation is repeated for the next month by including data for the next month but dropping data for the earliest month. Such window rolling across periods is continued until

Table 3. Data description

Data set	p	Time period	Abbreviation	Maximum correlation	Mean absolute correlation
100 portfolios formed on size and investment	100	January 1977–December 2021	100IN	0.94	0.75
100 FF portfolios	100	January 1977–December 2021	100FF	0.94	*0.71*
100 portfolios formed on size and operating profitability	100	January 1977–December 2021	100OP	0.94	0.75
100 random stocks from the S&P 500 universe	100	January 1977–December 2021	100SP	0.71	*0.28*
The combination of 100OP and 100SP	200	January 1977–December 2021	200OS	0.94	0.46
The combination of 100FF and 100SP	200	January 1977–December 2021	200FS	0.94	0.45
235 stocks traded in theNew York Stock Exchange	235	January 1977–December 2021	235NY	0.77	0.26
500 stocks downloaded from the CRSP	500	January 1977–December 2021	500CP	0.95	0.23

Notes: This table lists various data sets as named in Column 1. Column 2 reports the number of assets in each portfolio, Column 3 is its sample period, and Column 4 is its abbreviation. Columns 5 and 6 present the maximum correlation of asset returns and the average of absolute-value correlation for each data set, respectively. The 100SP, 235NY, and 500CP data sets are downloaded from the CRSP. The remaining data sets are extracted from the website of Ken French at: (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

the end of the data set is reached. In the end, $T-n$ vectors w_t of portfolio weights are generated for each out-of-sample period $t = \{n, n+1, \dots, T-1\}$.

Holding a portfolio with weights w_t for one period yields the out-of-sample return r_{t+1} at time $t+1$, where $r_{t+1} = w_t' y_{t+1}$ and y_{t+1} is a $p \times 1$ vector of asset returns at time $t+1$. For each portfolio, its VR, SR, and TO out of sample are computed below

$$\widehat{\mathrm{VR}} = \frac{1}{T-n-1}\sum\nolimits_{t=n}^{T-1} (r_{t+1} - \widehat{\mu})^2, \quad \text{with} \quad \widehat{\mu} = \frac{1}{T-n}\sum\nolimits_{t=n}^{T-1} r_{t+1},$$
$$\widehat{\mathrm{SR}} = \frac{\widehat{\mu}}{\widehat{\mathrm{VR}}}, \quad \text{and} \tag{11}$$
$$\widehat{\mathrm{TO}} = \frac{1}{T-n-1}\sum_{t=n}^{T-1}\sum_{j=1}^{p}(|w_{j,t+1} - w_{j,t^+}|),$$

where $w_{j,t}$ is the portfolio weight in asset j at time t, w_{j,t^+} is the *before-rebalancing* portfolio weight at time $t+1$, and $w_{j,t+1}$ is the *after-rebalancing* desired portfolio weight at time $t+1$. Portfolio TO is measured as the simple average of rebalancing trades across the p available assets and over the $T-n-1$ trading periods.

A desirable portfolio yields high returns adjusted to risk with low TO, yet a trade-off needs to be made between risk and TO. Risk reduction often requires a portfolio to be rebalanced, thus leading to high transaction costs. It is then interesting to work with the net SR (NSR) adjusted for such costs. This new ratio is computed according to Equation (11) but uses a different mean return that is adjusted by the transaction cost. Following DeMiguel, Martin-Utrera, and Nogales (2013), this kind of return is computed as

$$r_{t+1}^{\text{adj}-} = (1 + w_t' y_{t+1})(1 - \kappa \sum_{j=1}^{p} |w_{j,t+1} - w_{j,t^+}|) - 1, \tag{12}$$

where the transaction cost κ is often set at 50 basis points per trade.

3.2 The Proposed Estimator and Alternative Strategies

Given an inverse covariance estimator $\widehat{\Sigma}^{-1}$, the MVP weight vector is

$$w_{\text{MVP}} = \frac{\widehat{\Sigma}^{-1} e}{e' \widehat{\Sigma}^{-1} e},$$

where $e = (1, 1, \ldots, 1)'$. Replacing $\widehat{\Sigma}^{-1}$ with alternative estimators leads to different portfolios. Each structural assumption for Σ_u results in a different estimator $\widehat{\Sigma}$ and its inverse counterpart $\widehat{\Sigma}^{-1}$. The MVP resulting from an EF model with m FF factors and k LFs of PCA is denoted by FFm-PCAk-Z, where Z = {D, S, DD} represents a particular covariance structure (diagonal, sparse, DD) for Σ_u. If $k = X$, the number of LFs is *adaptively* chosen as in Bai and Ng (2002), which varies over the rolling windows.

The proposed portfolio, FFm-PCAX-DD, will be compared with a variety of existing portfolios in the finance literature, including both factor-free and factor-based portfolios. The factor-free portfolios include: (i) the MVP based on the sample covariance matrix S, denoted as MVP-S; (ii) the MVP constructed from the shrinkage estimator and proposed as $\widehat{\Sigma}_{\text{LW}} = (1 - \alpha)S + \alpha \overline{\sigma}^2 I$ by Ledoit and Wolf (2004), denoted as LW-I; (iii) the MVP formed via the graphical lasso proposed by Goto and Xu (2015), denoted as Glasso; and (iv) the MVP resulting from the NLS estimator of Ledoit and Wolf (2017) and applied directly to asset returns, denoted as NLS. The Glasso and NLS are state-of-the-art methods but work in a fundamentally different way. The structure-based Glasso aims to shrink an inverse covariance matrix to a sparse one while the structure-free NLS may not necessarily result in a sparse estimator. In general, the structure-based approach tends to perform well when the actual structure is close to the hypothesized one, but the structure-free method can do well otherwise as evidenced in Shi et al. (2020).

The factor-based portfolios are used for comparison, including: (v) the MVP derived from approximate factor models with the residual covariances Σ_g estimated by the NLS as in De Nard, Ledoit, and Wolf (2021), where one-factor and five-factor FF models are considered and denoted as FF1-NLS and FF5-NLS, respectively; and (vi) the MVP obtained from the FF models with a particular residual covariance structure Z = {D, S, DD}, denoted by FFm-Z. All the above MVPs are listed in Table 4 for comparison.

Table 4. Portfolio strategies

Strategy	Abbreviation
MVP with the sample covariance matrix	MPV-S
MVP with the Ledoit and Wolf (2004) shrinkage covariance matrix	LW-I
MVP based on the graphical lasso inverse covariance matrix estimator	Glasso
MVP based on the NLS estimator (Ledoit and Wolf 2017)	NLS
MVP with the NLS estimator applied to residual covariance matrices in FF models	FFm-NLS
MVP with a particular residual covariance structure $Z = \{D, S, \mathrm{DD}\}$ in FF models	FFm-Z
MVP with a particular residual covariance structure $Z = \{D, S, \mathrm{DD}\}$ in EF models	FFm-PCAX-Z

Notes: This table lists all the portfolio strategies to be evaluated and compared.

Table 5. Monthly out-of-sample portfolio variances based on factor models with different structures of residual/disturbance covariances

	$m = 1$		$m = 3$	
Method	100FF	100SP	100FF	100SP
Without LFs				
FFm-D	31.82	13.33	13.76	12.26
	(0.00)	(0.01)	(0.10)	(0.00)
FFm-S	*20.36*	12.07	15.73	10.96
	(0.00)	(0.01)	(0.00)	(0.09)
FFm-DD	22.80	*11.09*	*13.08*	*10.75*
	(0.00)	(0.62)	(0.84)	(0.33)
With LFs				
FFm-PCAX-D	14.42	12.35	13.56	11.33
	(0.01)	(0.00)	(0.08)	(0.01)
FFm-PCAX-S	13.78	10.82	13.09	10.58
	(0.11)	(0.65)	(0.73)	(0.42)
FFm-PCAX-DD	***13.34***	***10.72***	***13.01***	***10.40***
	(1.00)	(1.00)	(1.00)	(1.00)

Notes: This table reports the effects of LFs and covariance structures on out-of-sample portfolio variances for data sets 100FF and 100SP. The variances are measured in $\%^2$. The values in parentheses are the significance level of testing the null hypothesis of no difference in the out-of-sample variance between FFm-PCAX-DD and other portfolios. The p-values (in parentheses) are computed using the stationary bootstrap technique with bootstrap samples of $B = 1000$ and block size $b = 5$ (Ledoit and Wolf 2011). In each column, the minimum out-of-sample variance is recorded in italic for each type of model (with or without LFs) and in bold for each data set across the two types of models.

4 Empirical Analysis

4.1 Benefits of LFs and DD Structure in the EF Model

To illustrate the benefits brought by the DD structure and LFs to the EF model, Table 5 displays the out-of-sample variance (VR, i.e., risk) for portfolios with and without LFs under different covariance structures Z = {D, S, DD}. Our EF model, FFm-PCAX-DD, is compared with other models with D and S structures for $m = \{1, 3\}$. For FFm-S and FFm-PCAX-S, we conduct sparse thresholding computation and choose the tuning parameter in the way described earlier. This computing method is applied to FFm-S under soft-thresholding of residual covariances for $k = 0$ and to estimated residual returns $\{\hat{g}_t\}$ in FFm-PCAX-S for $k = X$. For other EF models, the number of LFs is also selected in an adaptive manner.

In Table 5, the minimum VR in each data set is first indicated in italics separately for FF and EF models. The smaller of the two minima is further highlighted in bold type for each data set. To examine risk differences between various models, we compute the p-values using the stationary bootstrap method from Ledoit and Wolf (2011), with the bootstrap sample chosen at $B = 1000$ and the block size at $b = 5$. The values in parentheses denote the p-values to test the null hypothesis of no difference in portfolio risk between FFm-PCAX-DD and other models.

Table 5 shows that the use of a DD structure for Σ_u is more advantageous for risk reduction than the diagonal and sparse structures. FFm-PCAX-DD creates uniformly smaller VR than FFm-PCAX-S and FFm-PCAX-D for all cases considered. The risk reduction from FFm-PCAX-DD relative to FFm-PCAX-D is substantial and significant at the 1% confidence level for most cases. This is also the case when comparison is made between FFm-DD and FFm-D (or FFm-S).

Table 5 also shows that the inclusion of LFs to explain residuals of an FF model helps reduce portfolio risk. For example, FFm-PCAX-DD attains smaller VR than FFm-DD for all cases. This is also the case when comparison is made between FFm-PCAX-D and FFm-D. This outcome is not surprising because adding LFs to the FF model tends to improve the accuracy of asset allocation and the stability of portfolio weights so as to reduce estimation error (for Σ_u^{-1}) and portfolio risk.

To illustrate the insufficiency of using FF factors only, Figure 1 plots the number of LFs selected in FF1-PCAX-DD for data sets (100FF, 100SP). This number varies over the rolling windows and implies the necessity for using LFs to improve model accuracy. More LFs are selected in EF models for 100FF than for 100SP because assets are more strongly correlated in 100FF than in 100SP as shown in Table 3.

Table 6 records condition numbers of $\widehat{\Sigma}$ for comparison between different models to display the implication of their covariance structures for estimation stability. The DD structure results in a smaller condition number than the diagonal structure, and a much smaller number than the sparse structure. The stability of covariance estimates due to the DD structure is helpful for risk reduction as shown already in Table 5.

4.2 Out-of-Sample Portfolio Performance

There are no guidelines for the choice of m FF factors in the finance literature that often struggles with what value is taken from $m = \{1, 3, 5\}$ (FF 2018). A value of m may be appropriate for some data sets but not for others, so it has to be determined in a data-driven manner. The K-fold cross-validation is suggested for this purpose by minimizing the out-of-sample variance (Shi et al. 2020). The return data set Y is split up into $K(= 5)$ roughly equal groups. The ith group $Y^{(i)}$ is treated as validation data while the other $K{-}1$ groups $Y^{(-i)}$ serve as training data. Denote by $VR_m^{(i)}$ the out-of-sample portfolio variance of FFm-PCAX-DD, with m FF factors applied to the validation data $Y^{(i)}$. The optimal number of m is obtained by minimizing the cross-validated out-of-sample variance over $m \in \{1, 3, 5\}$, that is,

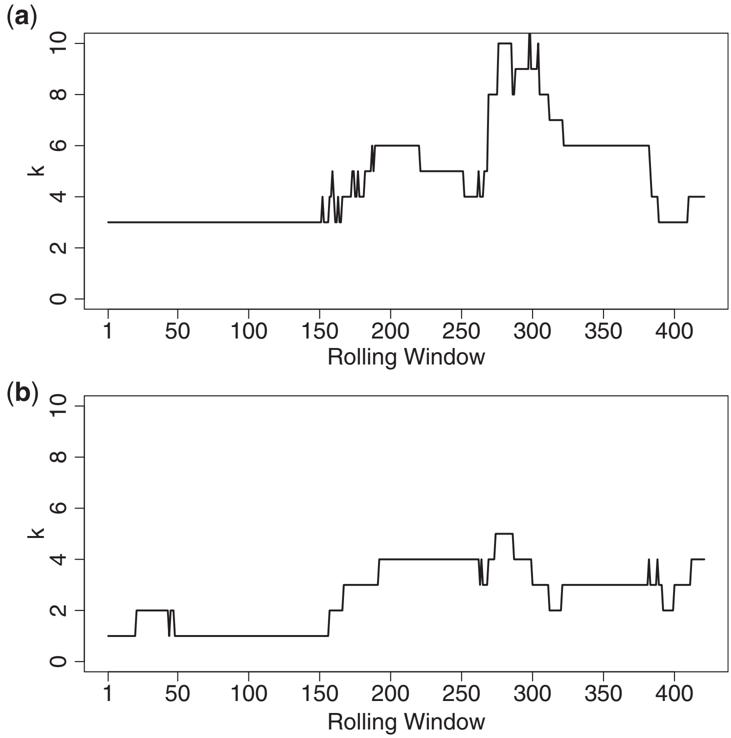

Figure 1. The number k of LFs selected over the rolling windows in FF1-PCAX-DD for (a) the 100FF data set and (b) the 100SP data set.

$$m^* = \min_m \left\{ \sum_{i=1}^{K} VR_m^{(i)} | m \in \{1, 3, 5\} \right\}. \tag{13}$$

Given the rolling-window approach for evaluating the out-of-sample portfolio performance, we can optimize m window by window. However, this would greatly increase computational burdens. To avoid this problem, we follow Goto and Xu (2015) and optimize m just for the first window and keep it unchanged over other windows. As such, the value of m in FFm-PCAX-DD for data sets (100IN, 100FF, 100OP, 100SP, 200OS, 200FS, 235NY, and 500CP) is selected to be (1, 3, and 5) in Tables 7–10.

Table 7 reports the out-of-sample variance (VR, i.e., risk) of various portfolios for different data sets, where the lowest VR of all portfolios is highlighted in bold for each data set. As expected, FFm-PCAX-DD achieves smaller VR than all factor-based portfolios (FF1-NLS and FF5-NLS). The risk reduction from it is more substantial for larger data sets, which is especially significant at the 1% confidence level for data sets (235NY and 500CP). MVP-S is the simplest and also least efficient of factor-free portfolios due to its highest VR, which is not defined for large data sets (200OS, 200FS, 235NY, and 500CP) due to the singularity of its sample covariance matrix. LW-I improves over MVP-S, albeit performing worse than Glasso and NLS. Compared with NLS, FFm-PCAX-DD substantially reduces portfolio risk for large data sets (200OS, 200FS, 235NY, and

Table 6. Condition numbers of various covariance matrix estimates based on factor models with different covariance structures

	$m=1$		$m=3$	
Method	100FF	100SP	100FF	100SP
Without LFs				
FFm-D	743.5	142.2	1258.6	164.3
	(289.8)	(65.9)	(332.5)	(66.3)
FFm-S	1252.1	763.1	5089.6	437.1
	(586.8)	(824.0)	(5299.4)	(207.3)
FFm-DD	*613.2*	*137.5*	*1122.9*	*159.0*
	(239.7)	(59.8)	(296.9)	(55.3)
With LFs				
FFm-PCAX-D	1562.9	197.6	1559.0	210.8
	(381.2)	(77.8)	(404.4)	(84.9)
FFm-PCAX-S	2532.1	396.1	2479.9	389.5
	(889.5)	(191.7)	(879.5)	(187.9)
FFm-PCAX-DD	*1359.8*	*166.0*	*1325.7*	*180.7*
	(351.8)	(63.7)	(328.8)	(75.7)

Notes: This table reports means and standard deviations (in parentheses) of condition numbers of the covariance matrices based on various factor models with different covariance structures.

Table 7. Monthly out-of-sample portfolio variances estimated from factor-free and factor-based models

Method	100IN	100FF	100OP	100SP	200OS	200FS	235NY	500CP
MVP-S	68.75	57.52	46.16	41.61	–	–	–	–
	(0.00)	(0.00)	(0.00)	(0.00)	(–)	(–)	(–)	(–)
LW-I	14.01	14.54	11.95	12.08	10.74	11.18	7.54	6.26
	(0.00)	(0.02)	(0.59)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
NLS	**11.74**	13.61	11.58	10.80	9.48	9.64	6.92	5.89
	(0.87)	(0.09)	(0.89)	(0.26)	(0.04)	(0.12)	(0.00)	(0.01)
Glasso	11.80	13.78	**11.12**	10.41	**8.46**	9.29	5.49	5.23
	(1.00)	(0.02)	(0.24)	(0.97)	(0.94)	(0.33)	(0.15)	(0.36)
FF1-NLS	12.15	13.84	11.83	10.86	9.55	9.69	7.00	5.92
	(0.33)	(0.01)	(0.71)	(0.11)	(0.03)	(0.10)	(0.00)	(0.01)
FF5-NLS	12.21	14.11	12.00	10.94	9.66	9.82	7.01	5.95
	(0.16)	(0.00)	(0.50)	(0.09)	(0.02)	(0.06)	(0.00)	(0.01)
FFm-PCAX-DD	11.81	**13.01**	11.65	**10.40**	8.49	**8.88**	**5.08**	**5.00**
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)

Notes: This table reports monthly out-of-sample variances of the studied portfolios for various data sets. The variances are measured in $\%^2$. The p-values in parentheses are used to test the null hypothesis of no difference in the out-of-sample variance between FFm-PCAX-DD and other portfolios. These p-values are computed using the stationary bootstrap technique with bootstrap sampling of $B = 1000$ and block size $b = 5$ (Ledoit and Wolf 2011). "–" indicates that MVP-S cannot be constructed due to the singularity of its sample covariance matrix. In each column, the minimum out-of-sample variance is highlighted in bold for each data set across all the portfolio strategies.

Table 8. Monthly out-of-sample portfolio SRs estimated from factor-free and factor-based models

Method	100IN	100FF	100OP	100SP	200OS	200FS	235NY	500CP
MVP-S	0.073	0.215	0.216	0.157	–	–	–	–
	(0.00)	(0.00)	(0.07)	(0.05)	(–)	(–)	(–)	(–)
LW-I	0.285	0.402	**0.343**	0.247	0.235	0.313	0.281	0.304
	(0.03)	(0.85)	(0.38)	(0.35)	(0.04)	(0.07)	(0.74)	(0.67)
NLS	0.324	0.396	0.323	0.268	0.256	0.314	0.281	0.320
	(0.91)	(0.45)	(0.97)	(0.87)	(0.07)	(0.04)	(0.71)	(0.98)
Glasso	0.320	0.389	0.331	**0.269**	0.281	0.322	**0.303**	**0.324**
	(0.67)	(0.26)	(0.69)	(0.78)	(0.24)	(0.04)	(0.50)	(0.83)
FF1-NLS	0.313	0.394	0.316	0.258	0.255	0.315	0.278	0.322
	(0.39)	(0.38)	(0.64)	(0.64)	(0.07)	(0.03)	(0.65)	(0.91)
FF5-NLS	0.307	0.392	0.318	0.264	0.250	0.317	0.276	0.314
	(0.10)	(0.32)	(0.78)	(0.94)	(0.06)	(0.06)	(0.59)	(0.89)
FFm-PCAX-DD	**0.325**	**0.405**	0.323	0.265	**0.307**	**0.374**	0.290	0.319
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)

Notes: This table reports monthly out-of-sample SRs of the studied portfolios for various data sets. The p values in parentheses are used to test the null hypothesis of no difference in the SR between FFm-PCAX-DD and other portfolios. These p-values are computed using the stationary bootstrap technique with bootstrap sampling of $B = 1000$ and block size $b = 5$ (Ledoit and Wolf 2008). "–" indicates that MVP-S cannot be constructed due to the singularity of its sample covariance matrix. In each column, the maximum out-of-sample SR is recorded in bold for each data set across all the portfolio strategies.

Table 9. Monthly out-of-sample portfolio TO indices estimated from factor-free and factor-based models

Method	100IN	100FF	100OP	100SP	200OS	200FS	235NY	500CP
MVP-S	7.625	7.119	6.698	2.655	–	–	–	–
LW-I	1.015	0.956	0.939	0.378	0.744	0.734	0.409	0.336
NLS	**0.646**	0.819	0.975	0.317	0.535	0.530	0.265	0.246
Glasso	0.684	0.715	**0.677**	**0.236**	**0.477**	0.485	**0.172**	**0.149**
FF1-NLS	0.873	0.821	0.869	0.275	0.563	0.541	0.276	0.258
FF5-NLS	0.730	0.824	0.793	0.365	0.622	0.621	0.298	0.276
FFm-PCAX-DD	0.723	**0.714**	0.719	0.266	0.501	**0.484**	0.243	0.225

Notes: This table reports the monthly TO of the studied portfolios for various data sets. "–" indicates that MVP-S cannot be constructed due to the singularity of its sample covariance matrix. In each column, the minimum TO is recorded in bold for each data set across all the portfolio strategies.

500CP). Compared with Glasso, it performs slightly better in risk reduction since it attains slightly smaller VR for five out of eight data sets and has similar VR for the remaining data sets.

Table 8 records out-of-sample SRs (i.e., risk-adjusted returns) of different portfolios for various data sets. The highest SR is acquired by LW-I for the data set 100OP, by Glasso for three data sets (100SP, 235NY, and 500CP), and by FFm-PCAX-DD for four data

Table 10. Monthly out-of-sample NSRs from factor-free and factor-based models

Method	100IN	100FF	100OP	100SP	200OS	200FS	235NY	500CP
MVP-S	−0.365	−0.249	−0.272	−0.050	–	–	–	–
	(0.00)	(0.00)	(0.00)	(0.00)	(–)	(–)	(–)	(–)
LW-I	0.146	0.271	0.203	0.193	0.120	0.202	0.205	0.235
	(0.00)	(0.09)	(0.57)	(0.14)	(0.01)	(0.01)	(0.26)	(0.37)
NLS	**0.226**	0.279	0.176	0.219	0.168	0.228	0.230	0.268
	(0.47)	(0.18)	(0.08)	(0.77)	(0.06)	(0.02)	(0.80)	(0.96)
Glasso	0.216	0.290	**0.226**	**0.232**	0.197	0.240	**0.265**	**0.291**
	(0.99)	(0.33)	(0.49)	(0.55)	(0.30)	(0.04)	(0.13)	(0.31)
FF1-NLS	0.182	0.280	0.186	0.216	0.163	0.227	0.225	0.268
	(0.13)	(0.14)	(0.13)	(0.64)	(0.06)	(0.02)	(0.67)	(0.96)
FF5-NLS	0.199	0.279	0.200	0.208	0.148	0.216	0.219	0.257
	(0.12)	(0.06)	(0.44)	(0.32)	(0.01)	(0.01)	(0.50)	(0.78)
FFm-PCAX-DD	0.216	**0.303**	0.215	0.224	**0.218**	**0.290**	0.235	0.266
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)

Notes: This table reports monthly out-of-sample NSRs of the studied portfolios for various data sets after adjusting transaction costs (at 50 basis points per trade). The values in parentheses are used to test the null hypothesis of no difference in the NSR between FFm-PCAX-DD and other portfolios. These p-values are computed using the stationary bootstrap technique with bootstrap sampling of $B = 1000$ and block size $b = 5$ (Ledoit and Wolf 2008). "–" indicates that MVP-S cannot be constructed due to the singularity of its sample covariance matrix. In each column, the maximum out-of-sample NSR is recorded in bold for each data set across all the portfolio strategies.

sets (100IN, 100FF, 200OS, and 200FS). Although Glasso attains a higher SR than FFm-PCAX-DD for three data sets, the difference is not significant. Additionally, FFm-PCAX-DD outperforms Glasso for the 200FS data set at the 5% significance level.

The SR considers both the mean returns and portfolio risk. In principle, reduction in the portfolio risk or increase in the mean returns can improve the SR. To provide some visual insights into how the out-of-sample portfolio returns spread out, Figure 2 displays the box plots for the distribution of out-of-sample portfolio returns from FFm-PCAX-DD under various data sets. To conserve space, only the factor-free approaches are included for comparisons. Note that MVP-S is not workable for the last four data sets and thus it is not included in the corresponding box plots. As expected, the range of portfolio returns from MVP-S over the rolling windows is the largest among all the methods for the first four data sets. Generally, the portfolio returns of FFm-PCAX-DD vary over a smaller range, compared with other approaches. This implies its good ability to reduce the variance of portfolio returns (or portfolio risks). Moreover, FFm-PCAX-DD yields a slightly higher mean return than Glasso and other approaches for the 200OS and 200FS data sets, which is another factor leading to higher SR for these two data sets.

Table 9 provides the monthly TO (i.e., rebalancing frequency) of different portfolios for all data sets. NLS, Glasso, and FFm-PCAX-DD are associated with, respectively, the smallest TO for the data set 100IN, five data sets (100OP, 100SP, 200OS, 235NY, and 500CP), and two data sets (100FF, 200FS). FFm-PCAX-DD produces the second smallest TO for five data sets (100OP, 100SP, 200OS, 235NY, and 500CP).

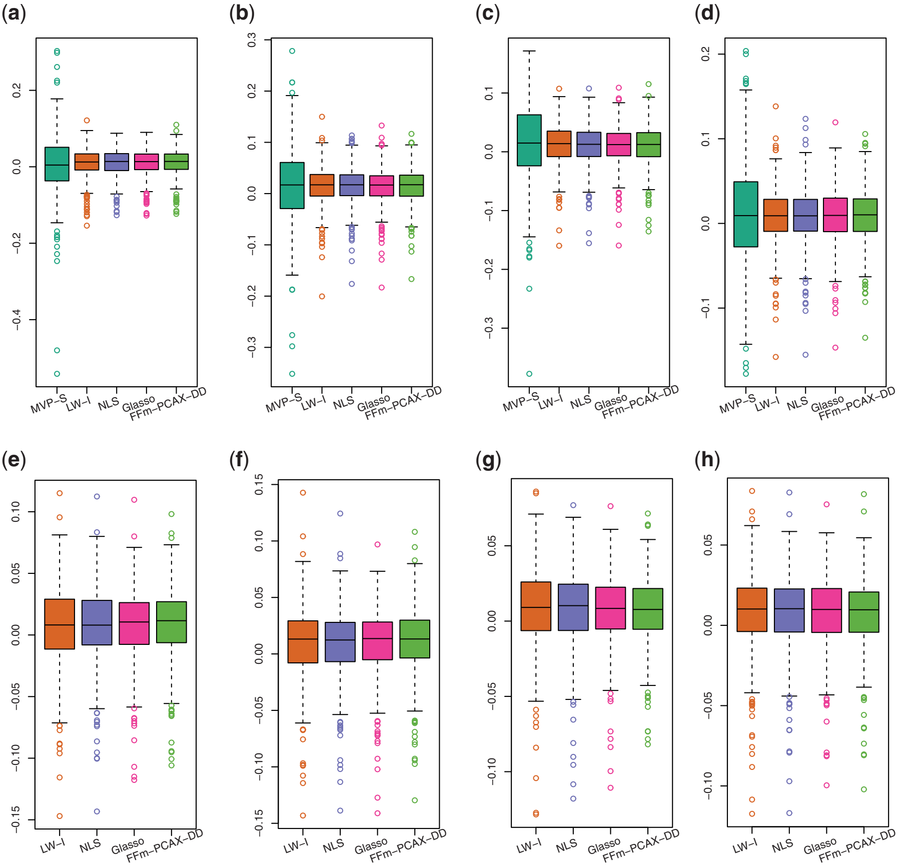

Figure 2. The distribution of out-of-sample portfolio returns for the various data sets. (a) 100IN. (b) 100FF. (c) 100OP. (d) 100SP. (e) 200OS. (f) 200FS. (g) 235NY. (h) 500CP.

Table 10 compares monthly out-of-sample NSRs (i.e., adjusted for portfolio risk and transaction cost) among different portfolios for various data sets. It is observed that FFm-PCAX-DD is capable of attaining the highest NSRs for three out of eight data sets and the second highest for four data sets (100IN, 100OP, 100SP, and 235NY). Moreover, its improvement over Glasso, NLS, and all other portfolios is statistically significant at the 5% confidence level for the 200FS data set.

The above comparisons indicate that FFm-PCAX-DD improves over most of the above approaches and is competitive to Glasso in terms of lowering out-of-sample risk and improving risk/cost-adjusted returns. However, it is important to note that FFm-PCAX-DD is much more computationally efficient than Glasso. The computation load is often a practical concern in portfolio selection, especially when there is a large number of assets. To illustrate the computational efficiency of FFm-PCAX-DD, Figure 3 displays the computation time of different methods to estimate the covariance matrix of asset returns for the first window under the 100IN and 500CP data sets. The experiment is carried out using a personal computer with a 2.3-GHz Quad-Core Intel Core i5 processor. To simplify the

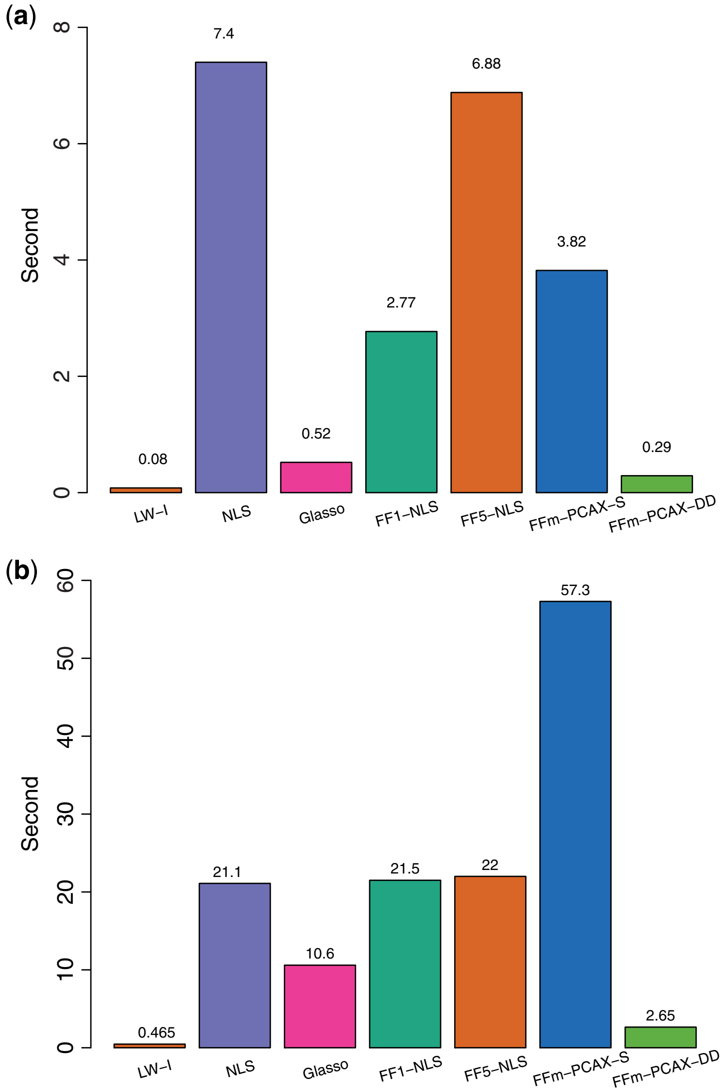

Figure 3. The computation time of different methods for the first window under the (a) 100IN and (b) 500CP data sets.

comparison, the time spent for selecting the tuning parameter of Glasso is not counted. As can be seen from Figure 3, under the 500CP data set, it takes 10.6 s for Glasso to compute the covariance estimate but around 2.65 s for FFm-PCAX-DD to do it. Clearly, FFm-PCAX-DD can greatly reduce computation time compared with Glasso, especially when the number of assets is large.

Relative performance in portfolio comparison depends on data sets and estimation windows as well as evaluation measures and testing methods (Goto and Xu 2015). Although it is difficult to display uniformly better performance of FFm-PCAX-DD for each data set, the above empirical work does show that FFm-PCAX-DD performs overall well in terms of low risk and high return (adjusted for risk and cost) across various data sets. If also taking into account its computation efficiency, FFm-PCAX-DD can be counted as a favorable strategy for portfolio selection, especially when a large number of assets are involved.

4.3 Comparison under Random Sampling

Each data set used in above comparison is confined to a single-fixed sample as in most of the extant studies. To extract more information on portfolio performance from a given data set, we follow Goto and Xu (2015) and compare different estimators under randomized samples. We use the 200OS data universe as an example and randomly choose 100 assets from it for 50 times. We then carry out a comprehensive performance comparison

Table 11. Comparison between the FFm-PCAX-DD and other competing strategies under random sampling

Method	Mean	SD	Max	Min	FFm-PCAX-DD > ALT	FFm-PCAX-DD < ALT
			Panel A: Out-of-sample variance			
FFm-PCAX-DD	9.69	0.46	10.93	8.94		
MVP-S	45.39	6.58	79.74	36.64	0	50
LW-I	11.32	0.65	12.90	9.90	0	50
NLS	11.18	0.73	12.97	9.93	0	50
Glasso	9.88	0.51	11.02	9.04	11	39
FF1-NLS	11.00	0.64	12.33	9.67	0	50
FF5-NLS	11.16	0.68	12.86	9.78	0	50
			Panel B: Out-of-sample SR			
FFm-PCAX-DD	0.288	0.024	0.328	0.239		
MVP-S	0.134	0.049	0.233	0.022	50	0
LW-I	0.256	0.028	0.320	0.206	47	3
NLS	0.260	0.029	0.336	0.206	45	5
Glasso	0.282	0.026	0.339	0.230	31	19
FF1-NLS	0.261	0.027	0.322	0.203	46	4
FF5-NLS	0.256	0.028	0.314	0.202	48	2
			Panel C: TO			
FFm-PCAX-DD	0.443	0.026	0.509	0.390		
MVP-S	4.445	0.282	4.930	3.868	0	50
LW-I	0.519	0.025	0.567	0.471	0	50
NLS	0.986	0.178	1.506	0.657	0	50
Glasso	0.415	0.020	0.450	0.371	48	2
FF1-NLS	0.671	0.070	0.847	0.543	0	50
FF5-NLS	0.766	0.100	1.071	0.584	0	50

Notes: This table reports the results from comparison between all portfolios formed via FFm-PCAX-DD and other competing strategies. The comparison is conducted in terms of distributions of out-of-sample portfolio variances ($\%^2$) in Panel A, of out-of-sample SRs in Penal B, and of TO indexes in Panel C. Those distributions are computed across 50 randomized samples, each of which consists of 100 assets randomly selected from the 200OS data set.

across the resulting 50 samples. Such random sampling makes it statistically more reliable to assess the relative ability of each portfolio strategy to perform in three aspects (VR, SR, and TO) out of sample.

Table 11 reports summary statistics of estimation results in terms of SR, VR, and TO (i.e., their monthly means, standard deviations, and minimum and maximum values). "FFm-PCAX-DD > ALT" is used to count the number of times that FFm-PCAX-DD achieves a realized monthly value of VR, SR, or TO higher than alternative portfolios (MVP-S, LW-I, NLS, Glasso, FF1-NLS, and FF5-NLS) in the 50 randomized samples, with m selected in accordance with Equation (13).

Panel A of Table 11 displays that FFm-PCAX-DD achieves lower portfolio risk than Glasso in 39 out of all the 50 runs and than the other portfolio strategies in all the 50 runs. Such striking role of the EF model for risk reduction enables it to achieve superior performance on risk-adjusted returns as well. Panel B reveals that FFm-PCAX-DD acquires a higher SR than all the alternatives except Glasso in almost all the 50 runs. Panel C also exhibits that FFm-PCAX-DD has lower TO than almost all alternative portfolios except Glasso at all times and this good performance in portfolio rebalancing occurs under random sampling.

5 Conclusion

This article develops an EF model for portfolio selection in large dimensions. The intended enhancements are acquired by introducing LFs to extract more information from residual returns of an FF model and by introducing a new structure, called the DD, to improve the stability of covariance estimates for idiosyncratic returns. Our EF model is general as it nests as special cases the FF model and LF model. Such modeling generalization can not only retain the economic interpretability of FF models but improve their estimation precision via LFs as well.

We find from the empirical comparison that the proposed approach performs much better than FF models in reducing portfolio variances and enhancing risk-adjusted returns. This outcome arises because FF factors are complemented by LFs while the DD-structure improves over other disturbance structures. Our approach also compares competitively with the Glasso as a state-of-the-art portfolio strategy. Moreover, the EF model enjoys computational simplicity when compared with the Glasso and becomes useful for portfolio formation with a large number of assets.

This work is limited to monthly asset returns and does not take into account ARCH/GARCH effects. Future studies can be carried out to examine the performance of this proposed method under time dependence. Additionally, this work discusses the stability of DD structure via simulation for multivariate normal distributions. It will be worth investigating this stability under non-normal or even unknown distributions.

Funding

The work of Dr. Shu was funded by the Science and Technology Development Fund of Macau SAR (FDCT/0033/2020/A1), the Department of Science and Technology of Guangdong Province (EF020/FBA-SLJ/2022/GDSTC), and the University of Macau Research Committee

(MYRG2020-00081-FBA). The work of Dr. Shi was supported by the National Natural Science Foundation of China (No. 72101119).

References

Aharoni, G., B. Grundy, and Q. Zeng. 2013. Stock Returns and the Miller Modigliani Valuation Formula: Revisiting the Fama French Analysis. *Journal of Financial Economics* 110: 347–357.

Alessi, L., M. Barigozzi, and M. Capasso. 2010. Improved Penalization for Determining the Number of Factors in Approximate Factor Models. *Statistics & Probability Letters* 80: 1806–1813.

Bai, J. 2009. Panel Data Models with Interactive Fixed Effects. *Econometrica* 77: 1229–1279.

Bai, J., and Y. Liao. 2017. Inferences in Panel Data with Interactive Effects Using Large Covariance Matrices. *Journal of Econometrics* 200: 59–78.

Bai, J., and S. Ng. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70: 191–221.

Bai, J., and S. Ng. 2006. Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica* 74: 1133–1150.

Bai, J., and S. Shi. 2011. Estimating High Dimensional Covariance Matrices and Its Applications. *Annals of Economics and Finance* 12: 199–215.

Bender, J., T. Blackburn, and X. Sun. 2019. Clash of the Titans: Factor Portfolios versus Alternative Weighting Schemes. *The Journal of Portfolio Management* 45: 38–49.

Callot, L. A., A. B. Kock, and M. C. Medeiros. 2017. Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice. *Journal of Applied Econometrics* 32: 140–158.

De Nard, G., O. Ledoit, and M. Wolf. 2021. Factor Models for Portfolio Selection in Large Dimensions: The Good, the Better and the Ugly. *Journal of Financial Econometrics* 19: 236–257.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal. 2009. A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management Science* 55: 798–812.

DeMiguel, V., A. Martin-Utrera, and F. J. Nogales. 2013. Size Matters: Optimal Calibration of Shrinkage Estimators for Portfolio Selection. *Journal of Banking & Finance* 37: 3018–3034.

Ding, Y., Y. Li, and X. Zheng. 2021. High Dimensional Minimum Variance Portfolio Estimation under Statistical Factor Models. *Journal of Econometrics* 222: 502–515.

Fama, E. F., and K. R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33: 3–56.

Fama, E. F., and K. R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1–22.

Fama, E. F., and K. R. French. 2018. Choosing Factors. *Journal of Financial Economics* 128: 234–252.

Fan, J., Y. Liao, and H. Liu. 2016. An Overview of the Estimation of Large Covariance and Precision Matrices. *The Econometrics Journal* 19: C1–C32.

Fan, J., Y. Liao, and M. Mincheva. 2013. Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75: 603–680.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance* 75: 1327–1370.

Giglio, S., Y. Liao, and D. Xiu. 2021. Thousands of Alpha Tests. *The Review of Financial Studies* 34: 3456–3496.

Goto, S., and Y. Xu. 2015. Improving Mean Variance Optimization through Sparse Hedging Restrictions. *Journal of Financial and Quantitative Analysis* 50: 1415–1441.

Hallin, M., and R. Liška. 2007. Determining the Number of Factors in the General Dynamic Factor Model. *Journal of the American Statistical Association* 102: 603–617.

Han, Y. 2006. Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model. *Review of Financial Studies* 19: 237–271.

Horn, R. A., and C. R. Johnson. 2012. *Matrix Analysis*. Cambridge University Press.

Jagannathan, R., and T. Ma. 2003. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance* 58: 1651–1683.

Kan, R., and G. Zhou. 2007. Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis* 42: 621–656.

Ke, Z. T., L. Xue, and F. Yang. 2020. Diagonally-Dominant Principal Component Analysis. *Journal of Computational and Graphical Statistics* 29: 592–607.

Kolm, P. N., R. Tütüncü, and F. J. Fabozzi. 2014. 60 Years of Portfolio Optimization: Practical Challenges and Current Trends. *European Journal of Operational Research* 234: 356–371.

Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10: 603–621.

Ledoit, O., and M. Wolf. 2004. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* 88: 365–411.

Ledoit, O., and M. Wolf. 2008. Robust Performance Hypothesis Testing with the Sharpe Ratio. *Journal of Empirical Finance* 15: 850–859.

Ledoit, O., and M. Wolf. 2011. Robust Performances Hypothesis Testing with the Variance. *Wilmott* 2011: 86–89.

Ledoit, O., and M. Wolf. 2017. Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *The Review of Financial Studies* 30: 4349–4388.

Li, G., L. Huang, J. Yang, and W. Zhang. 2021. A Synthetic Regression Model for Large Portfolio Allocation. *Journal of Business & Economic Statistics* 1–13. Doi: 10.1080/07350015.2021.1961787.

Mendoza, M., M. Raydan, and P. Tarazaga. 1998. Computing the Nearest Diagonally Dominant Matrix. *Numerical Linear Algebra with Applications* 5: 461–474.

Merton, R. C. 1980. "*On Estimating the Expected Return on the Market: An Exploratory Investigation*." Working paper, National Bureau of Economic Research.

Page, S., and M. Taborsky. 2011. The Myth of Diversification: Risk Factors vs. asset Classes. *The Journal of Portfolio Management* 37: 1–2.

Rajaratnam, B., H. Massam, C. M. Carvalho. 2008. Flexible Covariance Estimation in Graphical Gaussian Models. *Annals of Statistics* 36: 2818–2849.

Shi, F., L. Shu, A. Yang, and F. He. 2020. Improving Minimum Variance Portfolios by Alleviating Over-Dispersion of Eigenvalues. *Journal of Financial and Quantitative Analysis* 55: 2700–2731.

Won, J.-H., J. Lim, S.-J. Kim, and B. Rajaratnam. 2013. Condition-Number-Regularized Covariance Estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75: 427–450.