A linear programming model for selection of sparse high-dimensional multiperiod portfolios

Pun, Chi Seng; Wong, Hoi Ying

2019

Pun, C. S., & Wong, H. Y. (2019). A linear programming model for selection of sparse high-dimensional multiperiod portfolios. *European Journal of Operational Research*, 273(2), 754-771. doi:10.1016/j.ejor.2018.08.025

<https://hdl.handle.net/10356/87897>

<https://doi.org/10.1016/j.ejor.2018.08.025>

© 2019 Elsevier. This is the author created version of a work that has been peer reviewed and accepted for publication by *European Journal of Operational Research*, Elsevier. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1016/j.ejor.2018.08.025>].

Downloaded on 26 Apr 2023 17:43:50 SGT

A Linear Programming Model for Selection of Sparse High-Dimensional Multiperiod Portfolios¹

Chi Seng Pun^{a,*}, Hoi Ying Wong^b

^a*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore*

^b*Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*

Abstract

This paper studies the mean-variance (MV) portfolio problems under static and dynamic settings, particularly for the case in which the number of assets (p) is larger than the number of observations (n). We prove that the classical plug-in estimation seriously distorts the optimal MV portfolio in the sense that the probability of the plug-in portfolio outperforming the bank deposit tends to 50% for $p \gg n$ and a large n . We investigate a constrained ℓ_1 minimization approach to directly estimate effective parameters that appear in the optimal portfolio solution. The proposed estimator is implemented efficiently with linear programming, and the resulting portfolio is called the linear programming optimal (LPO) portfolio. We derive the consistency and the rate of convergence for LPO portfolios. The LPO procedure essentially filters out unfavorable assets based on the MV criterion, resulting in a sparse portfolio. The advantages of the LPO portfolio include its computational superiority and its applicability for dynamic settings and non-Gaussian distributions of asset returns. Simulation studies validate the theory and illustrate its finite-sample properties. Empirical studies show that the LPO portfolios outperform the equally weighted portfolio and the estimated optimal portfolios using shrinkage and other competitive estimators.

Keywords: Investment analysis; High-dimensional portfolio selection; Dynamic mean-variance portfolio; ℓ_1 minimization; Sparse portfolio

¹This work was supported by the Nanyang Technological University [grant number M4081840] and the Research Grant Council of Hong Kong [grant number 14303915]. The authors are also grateful to the seminar participants at the Chinese University of Hong Kong, University of California Los Angeles, 2014 Quantitative Methods in Finance, the third Asian Quantitative Finance Conference, and 2015 INFORMS Annual Meeting.

*Corresponding author

Email addresses: cspun@ntu.edu.sg (Chi Seng Pun), hywong@cuhk.edu.hk (Hoi Ying Wong)

1. Introduction

Portfolio theory is central to financial economics and is the building block of the capital asset pricing model. Portfolio theory is based on the assumption that the parameters of the underlying stochastic model are known and contain no estimation errors. For example, the mean-variance (MV) portfolio optimization model assumes that the mean vector and the covariance matrix of the asset returns are known (see Markowitz (1989)). Extensions of the MV problem to dynamic settings (multiperiod and continuous-time models) maintain the same assumption, as in Li and Ng (2000); Zhou and Li (2000); Basak and Chabakauri (2010). In this paper, we adopt the term “dynamic” in a discrete multiperiod sense, in contrast to the static single-period case.

When a trading strategy is applied to real data, point estimators are often plugged into the theoretical solution. For instance, empirical studies of MV portfolios often substitute the sample mean vector and the sample variance-covariance matrix into the optimal trading strategy. Chopra and Ziemba (1993); Michaud (1989) find that the estimation error is usually so large that the optimized strategy is no longer optimal.

For portfolio management, incorporation with parameter uncertainty and aversion to ambiguity, such as those used in Garlappi et al. (2007), are plausible approaches to address the issue of estimation error in the mean. Unfortunately, the empirical performance of optimized strategies is disappointing, even after incorporation of a Bayesian framework. DeMiguel et al. (2009) empirically examine various types of theoretical strategies and, surprisingly, find that the equally weighted (EW) strategy of investing equal amounts of money in each available asset outperforms the sample-based (plug-in) strategies across many datasets. Moreover, neither a Bayesian approach nor portfolio constraints improve the performance significantly. The minimum-variance strategy, which does not involve mean estimates, generally performs the best among the various types of theoretical strategies, but it is still unable to beat the EW strategy, which is free of estimation error.

The estimation error problem is more pronounced for a large portfolio because it requires a huge number of observations. Although the number of observations can be drawn from a long estimation window, Broadie (1993) notes that this increases the possibility that the estimated parameters are not stationary. In other words, one fundamental challenge to obtaining reliable estimates is the curse of dimensionality, because the number of securities (p) is greater than the number of observations (n). For a market with p securities, the number of parameters for an MV portfolio equals $p(p + 1)/2 + p$. According to the World Bank data from 2012, the total number of listed domestic companies in the world was 47,520, of which 4102 were in the

United States. Given 260 trading days per year, 15 years of daily data are the minimal charge for estimation of the covariance matrix for the US stock market, and 150 years of daily data are the minimal charge for the world. Nevertheless, practitioners tend to use datasets from a few recent years. Therefore, the estimation issue for MV portfolios is typically a high-dimensional statistical problem, as $p \gg n$. We are interested in this high-dimensional portfolio selection problem in this paper.

Factor-modelbased methods and shrinkage methods for estimation of the covariance matrix are possible remedies to high-dimensional portfolio problems. Fan et al. (2008) estimate the covariance matrix by assuming a known factor model for the asset returns. However, the usefulness of this technique depends on the agents' ability to determine the factors, which requires knowledge of the assets' dependence structure. Ledoit and Wolf (2004) shrink the sample covariance matrix before substituting it into the optimal trading strategy. Although these covariance estimates converge to the true covariance matrix under some assumptions, it is unclear whether the resulting trading strategy tends to the true trading strategy asymptotically. In addition, these approaches suggest investing in all assets available in the market. However, Boyle et al. (2012) find that ambiguity-averse investors tend to invest in a few risky assets when the correlations are high. Fouque et al. (2016) show mathematically that the robust optimal strategy with respect to assets' correlations may suggest investing in a few assets instead of all assets available in the market. Therefore, robustness to uncertain correlation may lead to a sparse optimal portfolio.

An alternative approach imposes portfolio constraints. Jagannathan and Ma (2003) find that the imposition of a no-short-selling constraint reduces the risk in estimated optimal portfolios. DeMiguel et al. (2009) generalize this result to norm-constrained portfolios. Fan et al. (2012) study global minimum-variance portfolio with ℓ_1 norm constraints and its statistical properties. However, the construction of norm-constrained portfolio may be dominated by a "tight" constraint, and as a result, the constrained portfolio cannot reflect the investor's views. Generally speaking, the literature solves the problem due to high dimensionality by reformulating it to the framework of the least absolute shrinkage and selection operator (LASSO) method that encourages a sparse portfolio. Brodie et al. (2009) discuss the stabilization of sparse portfolios; however, these results are specific to a static setting that cannot capture the dynamic hedging demands. Our goal is to devise a general approach to both static and dynamic strategies with the original objectives.

The parameter estimation for $p > n$ belongs to the field of high-dimensional statistics that has recently generated rich innovative results. In particular, the high-dimensional sample covariance matrix is known to be singular and unstable. Consistent estimators for the covariance matrix (Σ) often require some special

structures on the data, such as tapering; Cai et al. (2010) and references therein. The implementation of MV portfolios is even more complicated because it essentially requires the precision matrix ($\Omega := \Sigma^{-1}$). Penalized likelihood methods are used to estimate sparse precision matrices in Friedman et al. (2008). Cai and Liu (2011) proposes a constrained ℓ_1 minimization to estimate Fisher's linear discriminant function that depends on the precision matrix.

Inspired by these statistical advances, we propose and investigate a novel constrained ℓ_1 minimization approach for high-dimensional portfolios that directly estimates the effective parameters that appear in the optimal trading strategies. The estimator is formulated in a manner similar to the Dantzig selector introduced in Candès and Tao (2007) and can be efficiently implemented using linear programming. We call the new estimator the linear programming optimal (LPO) estimator and its corresponding estimated optimal portfolio the LPO portfolio. The proposed approach imposes no constraints upon the original problem of interest and can be applied to heavy-tailed asset returns. Therefore, its applicability is not limited to the portfolio problems that can be reformulated as a LASSO-type problem and includes the optimal dynamic trading strategies in particular. The LPO procedure essentially seeks the sparsest solution that best approximates the theoretical optimal portfolio. As with the Dantzig selector, the LPO procedure is to trade off a small increase in reward for a large reduction in portfolio risk. We stress that the proposed approach is applicable not only for the tricky high-dimensional case but also for the normal low-dimensional case.

To highlight the advantages of LPO portfolios, we first show the pitfalls associated with the traditional plug-in approach, which substitutes sample estimates directly into the optimal portfolio strategy. Although the literature already shows that the problem of the estimation error in the sample mean is serious (see Gandy and Veraart (2013)), the estimation error in the covariance matrix is also important and worthy of investigation in its own right because the global minimum-variance (GMV) portfolio does not involve the mean vector. We use the techniques in random matrix theory to prove that the plug-in high-dimensional MV portfolio degenerates to the random stock picking strategy, such that the probability of beating the bank deposit tends to 50%.

In addition to the asymptotic result of the degeneracy of the plug-in estimator and the remedial proposal of the LPO estimator for optimal portfolios, this paper's mathematical contribution is to show the oracle property, namely, our estimator is asymptotically equivalent to the oracle (ideal) estimator that uses background parameters under dynamic settings. Our simulations verify the derived theoretical results and imply

the superiority of the LPO portfolios in finite-sample numerical studies.

This paper also contributes to the literature by showing the advantages of the LPO portfolio and addressing the open problem posed by DeMiguel et al. (2009). Our empirical study uses datasets from the S&P 500 and Russell 2000 component stocks to show that the LPO portfolio consistently outperforms the EW portfolio. DeMiguel et al. (2009) find that the EW's performance is so good that optimal portfolios hardly beat it empirically. However, the manner in which the number of stocks and the selection of stocks to the EW portfolio are determined remains unclear. They conjecture that a larger portfolio size leads to better performance from the EW portfolio. Our empirical results show that the EW portfolio on stocks selected by the LPO procedure performs much better, in terms of the Sharpe ratio and certainty equivalent rate, than the EW portfolio on all available assets. Moreover, the portfolio performance can be further enhanced by additional statistical treatments on stocks selected by the LPO procedure. Therefore, the LPO approach has advantages not only in managing (ultra-) high-dimensional portfolios but also in filtering out unfavorable stocks.

The remainder of this paper is organized as follows. Section 2 reviews the classical results of MVP theory and highlights key observations that underlie the proposed approach. Section 3 proves mathematically the pitfalls of the traditional plug-in MV portfolio. Section 4 introduces the LPO approach and proves that it is asymptotically optimal. The connections between the LPO and ℓ_1 norm-constrained approaches are discussed in Appendix A. All proofs of the statements are collected in the Appendix B. We then conduct numerical and empirical studies to verify the aforementioned assertions in Section 5. Some supplementary numerical results are given in Appendix C and Appendix D, and Section 6 concludes the paper.

2. The Multiperiod Mean-Variance Model

Consider a financial market with p risky assets (stocks) and one risk-free asset (bank deposit or bond). Denote $S_i(t)$ as the i th stock price at time t for $i = 1, \dots, p$ and $S_0(t)$ as the risk-free asset price at time t . Define the asset return as $R_i(s, t) = S_i(t)/S_i(s) - 1$, for $s < t$, $i = 0, 1, \dots, p$. We consider a self-financing investment strategy $\{u_i(t)\}_{i=0}^p$ whose total wealth at time t is denoted by $X^u(t)$, where $u_i(t)$ is the cash amount invested in the i th asset at time t . A portfolio is characterized by $u(t) := (u_1(t), \dots, u_p(t))'$ as $u_0(t) \equiv X(t) - \sum_{i=1}^p u_i(t)$, where we denote by v' the transposition of a matrix/vector v .

Suppose that an MV investor wishes to invest his or her initial wealth x_0 with the target reward level z over a fixed investment horizon $T > 0$. In this paper, we focus on multiperiod MV strategies and the special

case of single-period strategies. The multiperiod MV model allows investors to rebalance the portfolio at finite discrete times $\{t_j\}_{j=0}^{N-1}$, where $t_0 < t_1 < \dots < t_{N-1} < t_N = T$ is the partition of $[0, T]$. We set $\Delta_j = t_{j+1} - t_j$. The risk-free asset return is deterministic: $R_0(t_j, t_{j+1}) = r_f \Delta_j$, where r_f is the annual risk-free interest rate. Note that the multiperiod strategy is generally a set of feedback controls: $u_m := \{u_m(t_j)\}_{j=0}^{N-1}$ that depend on the state variables and moments of returns. The wealth of the portfolio that adopts the strategy u_m is given by

$$X^{u_m}(t_{j+1}) = X^{u_m}(t_j)(1 + r_f \Delta_j) + u_m(t_j)'(R(t_j, t_{j+1}) - r_f \Delta_j \mathbf{1}_p), \quad (1)$$

where $\mathbf{1}_p = (1, \dots, 1)' \in \mathbb{R}^p$ and $R(t_j, t_{j+1}) = (R_1(t_j, t_{j+1}), \dots, R_p(t_j, t_{j+1}))'$ is the risky asset return vector for the j th period, whose mean vector (appreciation rate vector) and covariance matrix are denoted by $\mu_j \Delta_j$ and $\Sigma_j \Delta_j$, respectively. Notice that μ_j and Σ_j are annualized parameters for the period $[t_j, t_{j+1}]$.

The objective of the multiperiod MV strategy u_m is to minimize the portfolio variance, subject to a target expected portfolio return. Mathematically, the MV problem is stated as follows.

$$\min_u \quad \text{Var } X^u(T) \quad \text{subject to} \quad \mathbb{E}[X^u(T)] \geq z \text{ and } (X^u(\cdot), u(\cdot)) \text{ satisfies (1)}. \quad (2)$$

Throughout this paper, we assume that the target reward level $z > x_0 \prod_{j=0}^{N-1} (1 + r_f \Delta_j)$. The optimal u for the above problem, if it exists, is called the multiperiod MV strategy and is denoted by u_m^* . Notice that the multiperiod strategy $u_m(t_j)$ often depends on $X(t_j)$ and is hence a feedback control. The precommitment multiperiod MV solution is solved by Li and Ng (2000), whose results are summarized as follows.

Proposition 2.1. The multiperiod MV portfolio $u_m^* = \{u_m^*(t_j)\}_{j=0}^{N-1}$ of the problem (2) is given by

$$u_m^*(t_j) = -\frac{\Sigma_j^{-1} \beta_j}{1 + \Delta_j \beta_j' \Sigma_j^{-1} \beta_j} \left[X^{u_m^*}(t_j)(1 + r_f \Delta_j) - \gamma_m \prod_{k=j+1}^{N-1} (1 + r_f \Delta_k)^{-1} \right], \quad (3)$$

for $j = 0, 1, \dots, N-1$, where $\gamma_m = \frac{z - x_0 \prod_{j=0}^{N-1} (1 + r_f \Delta_j)(1 + \Delta_j \beta_j' \Sigma_j^{-1} \beta_j)^{-1}}{1 - \prod_{j=0}^{N-1} (1 + \Delta_j \beta_j' \Sigma_j^{-1} \beta_j)^{-1}}$ and $\beta_j = \mu_j - r_f \mathbf{1}_p$ is the excess return vector. In particular, when $N = 1$, we have the single-period (static) MV portfolio u_s^* :

$$u_s^* := u_s^*(t_0) = \frac{(z - x_0(1 + r_f T))}{T} \frac{\Sigma_0^{-1} \beta_0}{\beta_0' \Sigma_0^{-1} \beta_0}, \quad (4)$$

Remark: we use the convention $\prod_1^0(\cdot) = \prod_N^{N-1}(\cdot) = 1$.

The reason for our interest in the single-period MV strategy is that the analysis based on (4) is simpler and more interpretable and yields stronger results, as in Theorem 3.1. Moreover, Pun (2018) shows that

under the high-dimensional setting, the time-consistent strategies (as formulated in Basak and Chabakauri (2010)) are equivalent to myopic strategies that ignore future optimality and adopt single-period strategies in each period. The investigation of single-period strategies is also valid for time-consistent strategies. Because many empirical studies in the literature, such as DeMiguel et al. (2009), are based on a single-period setting, we use the single-period strategy to make comparisons with these existing approaches.

To implement MV strategies (3)-(4), we must estimate the parameter sets $\{(\mu_j, \Sigma_j)\}_{j=0}^{N-1}$ at time t_0 . The market practice usually estimates $\{(\mu_j, \Sigma_j)\}_{j=0}^{N-1}$ from the (projected) mean and the covariance matrix from daily returns. To simplify matters, we assume that the return vector in each period has the same first two moments and our presentation uses $(\mu, \Sigma) := (\mu_0, \Sigma_0) = \dots = (\mu_{N-1}, \Sigma_{N-1})$. However, our proposed approach is applicable for time-varying parameters, but tedious adjustments are needed. We remark here that this assumption does not affect the nature of dynamic optimality in a multiperiod setting and that the multiperiod optimal portfolio is theoretically superior to the myopic portfolio, which repeats the static portfolio in each period. Therefore, it is interesting to investigate this dynamic strategy. The approximation error in the projection from daily data is minimal for a short-term (e.g., 1-month) investment.

The MV strategies depend on the precision matrix $\Omega := \Sigma^{-1}$ and the excess return vector $\beta := \mu - r_f \mathbf{1}_p$. Traditional implementation directly substitutes the sample covariance matrix (Gram matrix) and the sample mean estimated from daily returns into the optimal strategies in Proposition 2.1. It also involves estimating the precision matrix Ω or taking the (generalized) inverse of the Gram matrix. However, we show in the next section that this leads to unexpectedly poor performance for a high-dimensional portfolio.

We notice a key observation that the strategies in (3) and (4) are proportional to $\Omega\beta$. We then allow $\eta := \Omega\beta$ to be the effective parameter for implementation of the single-period and multiperiod MV portfolios and establish an estimation for it. To fix the notation, $\hat{\eta}$ and $\hat{\beta}$ denote estimates of η and β , respectively. Consequently, $\Theta_p := \beta'\Omega\beta$ is estimated as $\hat{\beta}'\hat{\eta}$. Hence, we define the following notations. For $j = 0, 1, \dots, N-1$,

$$\hat{u}_s := u_s^* \Big|_{\beta=\hat{\beta}}^{\eta=\hat{\eta}} = \frac{(z - x_0(1 + r_f T))}{T} \frac{\hat{\eta}}{\hat{\beta}'\hat{\eta}}, \quad \hat{\gamma}_m := \frac{z - x_0 \prod_{j=0}^{N-1} (1 + r_f \Delta_j) (1 + \Delta_j \hat{\beta}'\hat{\eta})^{-1}}{1 - \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'\hat{\eta})^{-1}}, \quad (5)$$

$$\hat{u}_m(t_j) := u_m^*(t_j) \Big|_{\beta=\hat{\beta}}^{\eta=\hat{\eta}} = \frac{-\hat{\eta}}{1 + \Delta_j \hat{\beta}'\hat{\eta}} \left[X^{\hat{u}_m}(t_j) (1 + r_f \Delta_j) - \frac{\hat{\gamma}_m}{\prod_{k=j+1}^{N-1} (1 + r_f \Delta_k)} \right], \quad (6)$$

3. Pitfalls of the sample covariance matrix in optimal portfolios

This section investigates the problems raised by the use of sample estimates. The empirical investigation of the related issue can be found in DeMiguel et al. (2009); Garlappi et al. (2007). We naturally ascribe the unsatisfactory performance of the optimal portfolio to the estimation errors in both the mean and covariance matrix; however, we focus on the estimation error of the precision matrix by inverting the Gram matrix. Our aim is to explain the poor empirical performance of the minimum-variance portfolio, which does not depend on the mean estimate. In other words, our message is that regardless of the amount of effort exerted to reduce the effects of the estimation error in the mean, the high-dimensional portfolio will still perform poorly because of the difficulty of precise estimation of the precision matrix.

Let $\{S^{(l)} = (S_1(t_l), \dots, S_p(t_l)), l = 0, 1, \dots, n\}$ be the collection of $n + 1$ observations of the asset price vectors of size p , which can be converted into n observations in the return vectors. We assume a fixed sampling interval of δt . The traditional approach estimates Σ using the projected Gram matrix:

$$\hat{\Sigma}_n = \frac{1}{n\delta t} \sum_{l=1}^n (r^{(l)} - \bar{r})(r^{(l)} - \bar{r})', \quad (7)$$

where $r^{(l)} = (r_1^{(l)}, \dots, r_p^{(l)})$ is the sample return vector of the p assets, $r_i^{(l)} = S_i(t_l)/S_i(t_{l-1}) - 1$ and $\bar{r} = \sum_{l=1}^n r^{(l)}/n$. When $p > n$, $\hat{\Sigma}_n$ is singular and an unstable estimate of Σ in general. The problem becomes more serious for estimation of the precision matrix, which is the Moore-Penrose (generalized) inverse ($\hat{\Sigma}_n^-$) for a singular matrix. Therefore, the plug-in approach uses $\hat{\eta}_n := \hat{\Sigma}_n^- \beta$ for the MV portfolios.

We denote by $\mathbb{E}_D$ and $\mathbb{P}_D$ the conditional expectation and probability given the historical data (σ -field $D = \sigma(\{S^{(l)}\}_{l=0}^n)$). Our analyses of the estimated strategies are based on the conditional expectations with respect to the historical data, which account for the variation of the data. We then analyze the asymptotic properties of the data-driven strategies when the number of assets increases with the sample size and both tend to infinity. To this end, we must specify the distribution of the return vector.

We first render the return vector as $R(t_j, t_{j+1}) = \mu\Delta_j + \sigma\sqrt{\Delta_j}\xi_j$ for $j = 0, 1, \dots, N - 1$, where σ is the volatility matrix of the asset returns such that $\Sigma = \sigma\sigma'$ and $\{\xi_j\}_{j=0}^{N-1}$ is a sequence of independent p -dimensional random vectors with a zero mean and identity covariance matrix. We assume the normality² for $\{\xi_j\}_{j=0}^{N-1}$, which is limited to this section only. We intend to show that the portfolio performance remains poor even given the most favorable normality assumption.

² An alternative model in the literature assumes log-normality for the simple return or equivalently normality for the log-return $(R_i^{\log}(t_j, t_{j+1}) := \log(S_i(t_{j+1})/S_i(t_j)) = \log(1 + R_i(t_j, t_{j+1})))$, especially in the continuous-time setting. However, the log-normal model introduces difficulty into the probabilistic calculation of multiperiod portfolios. Although the normal model

3.1. Probability of beating the bank deposit

Why are investors interested in an optimal portfolio? Investors believe that an optimal portfolio outperforms a bank deposit with a reasonably high probability. Under the normality assumption, $\mathbb{P}(X^{u_s}(T) > x_0(1 + r_f T)) > 1/2$ if the parameters β and Σ are known. In layman's terms, it is not by chance that the optimal MV portfolio beats the bank deposit. The same can be proven for the multiperiod MV problem. In fact, the multiperiod strategy is theoretically better than the single-period strategy because it has a larger opportunity set. However, the following theorem shows that both single-period and multiperiod high-dimensional MV portfolios implemented with the plug-in approach turn out to perform very poorly in an asymptotic manner.

Theorem 3.1. If $\{\xi_j\}_{j=0}^{N-1}$ are normally distributed and $\Theta_p = o(p/n)$, then as $p/n \rightarrow \infty$ and $n \rightarrow \infty$

1. $\mathbb{P}_D(X^{\hat{u}_s}(T) > x_0(1 + r_f T)) \xrightarrow{\mathbb{P}} \frac{1}{2};$
2. $\mathbb{P}_D\left(X^{\hat{u}_m}(T) > x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k)\right) \leq U_{n,p} \xrightarrow{\mathbb{P}} \frac{1}{2},$

where we recall that $\Theta_p = \beta' \Omega \beta$.

The condition $\Theta_p = o(p/n)$ in Theorem 3.1 has a financial interpretation. Note that $\Theta_p = |\Sigma^{-1/2} \beta|^2$ is the sum of the squared Sharpe ratios. Because the theorem considers the situation of $p/n \rightarrow \infty$, it is reasonable to require that the squared Sharpe ratio is much smaller than the large value of p/n . Otherwise, the market trend becomes too predictable for investors to generate a positive Sharpe ratio with a high probability.

Before presenting an interpretation of Theorem 3.1, we wish to highlight the difficulty of the proof. It may seem that the single-period result is a direct consequence of Bickel and Levina (2004)'s linear discriminant analysis, because they prove that the probability of discriminating two different classes of data

may generate a negative stock price (in the case of $R_i(t_j, t_{j+1}) < -1$) mathematically, it rarely occurs in practice or in our simulation studies because the mean of $R_i(t_j, t_{j+1})$ is nonnegative and because its variance reads $v^2 \Delta_j$ with $v < 1$ such that the occurring probability of negative stock price is at most $\Phi(-(v\sqrt{\Delta_j})^{-1}) \approx 0$ (noting that $\Delta_j = 1/252$ in our simulations). Moreover, the normal and the log-normal models are nearly identical practically by noting that $|R_i(t_j, t_{j+1}) - R_i^{log}(t_j, t_{j+1})| \leq \frac{1+R_i(t_j, t_{j+1})}{2} \left(R_i^{log}(t_j, t_{j+1})\right)^2$, which is negligible when Δ_j is small. The normality assumption for simple returns may be oversimplified, but it does not affect the insights of our results. It is noteworthy that the distributional assumption only occurs when calculating the probability of the wealth event (Section 3) and that the applicability of the proposed approach is beyond the normal simple returns.

tends to 0.5 under the worst-case scenario. However, the first probability in Theorem 3.1 tends to 0.5 for all scenarios. The techniques involved are different.

Our result indicates the seriousness of the problem caused by the plug-in approach and why it rarely gives a satisfactory empirical performance for the sample-based optimal MV portfolio. Theorem 3.1 reveals that the plug-in MV portfolio beats the bank deposit solely by pure luck for $p \gg n$. The probability of offering a positive excess return tends to 50% for a large portfolio. This performance is similar to the random stock picking or blindfolded monkey picking strategy. The multiperiod MV portfolio can have even worse results because the upper bound of its probability of beating the bank deposit tends to 50% for a large portfolio. Therefore, the more optimization we do, the poorer the sample-based optimal portfolio results, because although the optimization takes full advantage of the underlying stochastic model, including the model parameters, it also amplifies the effects of the estimation error.

4. A constrained ℓ_1 minimization approach

Rather than norm-constraining the portfolio as in DeMiguel et al. (2009); Fan et al. (2012), we construct efficient and accurate estimates for the effective parameters that appear in the optimal control. The intent is to preserve the desirable optimality posited by the theory and obtain a better out-of-sample performance. For MV portfolios, the effective parameters are (β, η) . We remark that if $(\mu_0, \Sigma_0), \dots, (\mu_{N-1}, \Sigma_{N-1})$ are different, then the effective parameters become $\{(\beta_j := \mu_j - r_f \mathbf{1}_p, \eta_j := \Sigma_j^{-1} \beta_j)\}_{j=0}^{N-1}$ and estimating them is similar to estimating (β, η) for N times.

This section proposes a constrained ℓ_1 minimization approach for direct estimation of η (while bypassing the estimation of Ω). We call it the linear programming optimal (LPO) approach because it is easily implemented using linear programming. This newly proposed approach leads to a direct estimation of the optimal control. It can be recognized from (3) and (4) that the MV strategies depend on Ω only through the term $\eta = \Omega \beta$. Inspired by Cai and Liu (2011), we estimate η as

$$\tilde{\eta} \in \arg \min_{\eta \in \mathbb{R}^p} \left\{ |\eta|_1 \text{ subject to } |\hat{\Sigma}_n \eta - \hat{\beta}_n|_\infty \leq \lambda_n \right\}, \quad (8)$$

where $\hat{\Sigma}_n$ is given in (7), $\hat{\beta}_n = \frac{1}{n\delta t} \sum_{l=1}^n r^{(l)} - r_f \mathbf{1}_p$, λ_n is a tuning parameter, $|a|_1 = \sum_{i=1}^p |a_i|$, and $|a|_\infty = \sup_i |a_i|$ for $a \in \mathbb{R}^p$. We refer readers to Candès and Tao (2007), James et al. (2009) and Cai and Liu (2011) for more details on the ℓ_1 minimization method. The LPO approach is also applicable for estimation of $\Omega \mathbf{1}_p$, the effective parameters in the minimum-variance strategy, by setting $\beta = \mathbf{1}_p$. Although

the sample estimates of Σ and β are used in (8) for a simple implementation, we stress that other improved estimators are also possible candidates for (8). From this viewpoint, we can assume that the historical data are correlated in time and make use of time series estimation. We denote the estimated strategies (5) and (6) with the LPO estimate (8) as $\tilde{u}_s$ and $\tilde{u}_m$, respectively, and call them LPO strategies (or portfolios) in general.

The LPO estimators and LPO strategies are characterized by λ_n , which is determined by the cross-validation approach (detailed in Section 5). Due to the nature of the LPO approach, when $\lambda_n > |\hat{\beta}_n|_\infty$, $\tilde{\eta} = 0$, which suggests avoidance of risky assets. We are thus interested in $\lambda_n \in [0, |\hat{\beta}_n|_\infty]$. If $n > p$ and $\hat{\Sigma}_n$ is nonsingular, $\hat{\Sigma}_n^{-1}\hat{\beta}_n$ is the candidate of the feasible set with any nonnegative λ_n , and the LPO estimator is degenerated to the plug-in sample estimator $\tilde{\eta} = \hat{\Sigma}_n^{-1}\hat{\beta}_n$ when $\lambda_n = 0$. Because λ_n is tuned based on cross-validated experiments of the LPO strategies with different λ_n values over $[0, |\hat{\beta}_n|_\infty]$, the LPO strategy with the fine-tuned λ_n performs as well as or better than the plug-in strategy. In the high-dimensional case of $p > n$, the feasible set for the minimization problem (8) may be empty for a small λ_n . The admissible set can be determined by testing the feasibility of (8) by gradually decreasing the value of λ_n from $|\hat{\beta}_n|_\infty$ to the smallest positive value. Details can be found in Section 5.

Heuristically, the idea behind this approach is to view η as the solution to the equation $\Sigma\eta - \beta = 0$. When Σ and β are unknown, we replace them with the sample estimates $\hat{\Sigma}_n$ and $\hat{\beta}_n$ (or other improved estimates). However, the estimation errors of $\hat{\Sigma}_n$ and $\hat{\beta}_n$ are so large that they may make the true η outside the set $\{\eta : \hat{\Sigma}_n\eta - \hat{\beta}_n = 0\}$. Therefore, we consider the feasible set $\{\eta : |\hat{\Sigma}_n\eta - \hat{\beta}_n|_\infty \leq \lambda_n\}$ to account for the variability in $\hat{\Sigma}_n$ and $\hat{\beta}_n$ and seek the most asymptotically sparse solution within this set that has the lowest regulation cost (or transaction cost for our application) to recover the true η . We show in the Appendix B that with the appropriate choice of λ_n , the true η belongs to the feasible set with high probability.

Notice that the MV strategies (3) and (4) are proportional to η . The sparsity of η is exactly the cardinality of the MV portfolios. A sparse portfolio is particularly welcomed by ambiguity-averse investors who wish to invest in only a few assets in the market instead of all available assets. The sparsity reduces the variability of the portfolio and hence enhances its stability. This property holds true under both low- and high-dimensional settings, as discussed by Brodie et al. (2009). Using the ℓ_1 minimization approach, we trade off the unfavorable and volatile assets. The ℓ_1 minimization is also mathematically attractive because it enables us to prove the desired oracle property.

The LPO approach has many noteworthy advantages:

1. Classic theoretical solutions are not ignored because they are part of the implementation procedure and the corresponding interpretations remain. Therefore, this approach works particularly for dynamic strategies problems.
2. This approach works for a singular $\hat{\Sigma}_n$ and for any other improved estimator of Σ .
3. This approach also works for fat-tailed distributions of asset returns.
4. The implementation is simple, as (8) can be implemented using linear programming.
5. The size of the estimation problem is reduced from $p(p+1)/2$ (estimating Ω with the separate estimation approach) to p (estimating η with the LPO approach). This significantly lowers the computational burden in the high dimensions. Note that under the high-dimensional setting, the estimate of Ω cannot be immediately obtained from $\hat{\Sigma}_n$.
6. The sparsity on η is a weaker condition than the conditions required for the consistent estimate of Ω .
7. Because the resulting LPO portfolio is sparse, the LPO procedure automatically filters out unfavorable stocks based on the MV criterion and stimulates low-dimensional portfolios on LPO-based assets.

Appendix A discusses the LPO portfolio's connections with the ℓ_1 norm-constrained portfolio. Under the (approximate) sparsity condition of η , we show that the LPO approach preserves the optimality of the MV strategy for $n \rightarrow \infty$ and $\log p/n \rightarrow 0$. We remark here that if $(\mu_0, \Sigma_0), \dots, (\mu_{N-1}, \Sigma_{N-1})$ are different, the LPO estimates of $\{\eta_j\}_{j=0}^{N-1}$ are $\{\tilde{\eta}_j\}_{j=0}^{N-1}$, where $\tilde{\eta}_j$ is solved from (8) with $\hat{\Sigma}_n$ and $\hat{\beta}_n$ replaced by the estimates of Σ_j and β_j , respectively. These N LPO estimates will also preserve the following oracle properties. A heuristic argument for this claim is that under the same conditions on the estimates of Σ_j and β_j , $\{\tilde{\eta}_j\}_{j=0}^{N-1}$ will behave similarly in p -asymptotic manner.

4.1. Oracle properties

We write the return vector in each observation period as $r_{\delta t} = \mu\delta t + \sigma\sqrt{\delta t}\xi =: \mu\delta t + \sqrt{\delta t}\Psi$, where $\Psi = (\Psi_1, \dots, \Psi_p)'$ is a random vector with a zero mean and a covariance matrix of Σ . Let $\bar{\Psi} = \Psi'\Omega\beta/\sqrt{\beta'\Omega\beta}$ be a standardized random variable with zero mean and unit variance. The LPO approach can be applied beyond the normality assumption once Ψ satisfies the distributional conditions that for any p -dimensional nonrandom vector l with $|l|_2 = 1$ and any $t \in \mathbb{R}$,

$$\mathbb{P}(l'\Omega^{\frac{1}{2}}\Psi \leq t) =: \Upsilon(t) \quad (9)$$

is a continuous distribution function symmetric about 0 and does not depend on l and that either one of the following two comments conditions for the return vector is satisfied:

- (M1) (*Sub-Gaussian-type tails*) Suppose that $\log p \leq n$ and there exist some constants $c_1 > 0$ and $K_1 > 0$ such that $\mathbb{E} \exp(c_1 \bar{\Psi}^2) \leq K_1$, and $\mathbb{E} \exp(c_1 \Psi_i^2 / \sigma_{ii}) \leq K_1$ for all i .
- (M2) (*Polynomial-type tails*) Suppose that for some $c_1, c_2 > 0$, $p \leq c_1 n^{c_2}$, and for some $\epsilon > 0$, $\mathbb{E} |\bar{\Psi}|^{4c_2+4+\epsilon} \leq K_1$, and $\mathbb{E} |\Psi_i / \sqrt{\sigma_{ii}}|^{4c_2+4+\epsilon} \leq K_1$ for all i .

Obviously, the well-known normal random vectors satisfy (9) and (M1). In fact, all elliptical distributions and the multivariate scale mixture of normals satisfy (9); all centered and bounded random variables satisfy (M1); and many classes of heavy-tailed distributions, such as t -distributions and Pareto distributions, have polynomial-type tails (see Novak (2012)). Empirical evidence and textbooks, such as Luenberger (1998), attest to the fact that financial data often exhibit heavy tails. The theories of the LPO approach built on top of these distributional assumptions are general and applicable.

We shall show that the LPO portfolios with $\Omega\beta$ estimated by $\tilde{\eta}$ as defined in (8) tend to the corresponding oracle strategies, where the parameters Σ and β are perfectly estimated, in probability. To this end, we introduce a condition shared by many high-dimensional statistical research problems such as that in Cai and Liu (2011):

- (C1) $\log p \leq n$, $\max_{1 \leq i \leq p} \sigma_{ii} \leq M$ for some constant $M > 0$, where we read $\Sigma = (\sigma_{ij})_{p \times p}$, and $\Theta_p \geq K$ for some $K > 0$.

This condition indeed also has a financial interpretation. The condition $\max_{1 \leq i \leq p} \sigma_{ii} \leq M$ means that all stocks in the market should have an upper bound in volatility. This is reasonable. If there exists a stock with infinite volatility, the associated listed firm should go bankrupt very soon after going public. The condition $\Theta_p \geq K > 0$ refers to a situation in which the squared Sharpe ratio of the market has a lower bound and is bounded away from zero. This is also natural because a function of the stock market is to provide investors with an opportunity to gain from economic growth. Therefore, most stocks in the market offer non-zero Sharpe ratios, individually. This makes the squared Sharpe ratio of the market strictly positive. Otherwise, all investors would prefer bank deposits and the stock market would not exist.

Theorem 4.1. Let $\lambda_n = C\sqrt{\Theta_p \log p/n}$ with $C > 0$ being a sufficiently large constant, $a_n = |\hat{\Sigma}_n - \Sigma|_\infty$ and $b_n = |\hat{\beta}_n - \beta|_\infty$. Suppose that the historical daily return data satisfy (M1) [or (M2)], condition (C1)

holds, and

$$d_n := (\lambda_n + b_n)|\eta|_1 + a_n|\eta|_1^2 = o_P(1). \quad (10)$$

Then we have as $n \rightarrow \infty$ and $\log p/n \rightarrow 0$,

1. $\mathbb{E}_D(X^{\tilde{u}_s}(T) - X^{u_s}(T))^2 \xrightarrow{\mathbb{P}} 0$, $\mathbb{E}[X^{\tilde{u}_s}(T)] \xrightarrow{\mathbb{P}} z$, $(\sigma_T^{\tilde{u}_s})^2 - (\sigma_T^{u_s})^2 \xrightarrow{\mathbb{P}} 0$;
2. $\mathbb{E}_D(X^{\tilde{u}_m}(T) - X^{u_m}(T))^2 \xrightarrow{\mathbb{P}} 0$, $\mathbb{E}[X^{\tilde{u}_m}(T)] \xrightarrow{\mathbb{P}} z$, $(\sigma_T^{\tilde{u}_m})^2 - (\sigma_T^{u_m})^2 \xrightarrow{\mathbb{P}} 0$.

To the best of our knowledge, we are the first to provide an estimated multiperiod optimal strategy with the oracle properties under the high-dimensional setting. Fan et al. (2012) prove the oracle properties of the static strategy implied by their approach only in terms of the variance measure. We, however, prove that both terminal wealth and its mean and variance tend to the oracle benchmarks for both static and dynamic trading strategies. We stress here that the oracle results above require only $\log p/n \rightarrow 0$. From this perspective, the LPO approach is applicable for a small $p < n$, large $p \approx n$, or even ultra-high $p > n$ (but $p < e^n$), which contrasts to the applicability of the traditional plug-in approach (Theorem 3.1).

Regarding the convergence rate of the Gram matrix, Bickel and Levina (2004) and Cai and Liu (2011) show that $a_n := |\hat{\Sigma}_n - \Sigma|_\infty \leq C\sqrt{\log p/n}$ for some large constant $C > 0$ with an overwhelmingly high probability that is greater than $1 - \mathcal{O}(p^{-1})$ and $1 - \mathcal{O}(p^{-1} + n^{-\epsilon/8})$ under (M1) and (M2), respectively. Therefore, the condition (10) has an overwhelmingly high probability of holding once the portfolio has sufficient sparsity that $|\eta|_1 = o(\sqrt{\log p/n})$, if the estimation error in the mean can be ignored, such as in the case of the minimum-variance portfolio.

We further investigate the condition (10) and attempt to show that it is not restrictive. Consider a condition that is stronger than (C1):

$$(C2) \quad \log p \leq n, \quad M^{-1} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq M \text{ for some constant } M > 0, \\ \text{and } \Theta_p \geq K \text{ for some } K > 0,$$

where $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$ are the smallest and largest eigenvalues of Σ , respectively. The financial interpretation of (C2) is similar to (C1) except that (C2) bounds all possible portfolio risks instead of the individual stock's volatility. By the Cauchy-Schwarz inequality, we have

$$|\eta|_1^2 \leq |\eta|_0 |\eta|_2^2 \leq M^2 |\eta|_0 \|\beta\|^2 \leq M^3 |\eta|_0 \Theta_p.$$

Therefore, Theorem 4.1 still holds true under (C2) and $|\eta|_0(\Theta_p \sqrt{\log p/n} + b_n \sqrt{\Theta_p}) = o(1)$, instead of (C1) and (10). Although (C2) is a widely adopted condition for high-dimensional covariance matrix estimation,

the preceding inequality implies that the larger the Θ_p , the slower the convergence rate of the LPO portfolio. This is reasonable because the oracle strategy with a large Θ_p , or equivalently a highly inflated market, has a probability close to 1 of beating the risk-free asset. It is difficult for any data-driven method to achieve such a high-standard performance. In (10), b_n reflects the estimation error in the mean vector. If this estimation error is large, the additional sparsity of the MV portfolio is required to ensure the convergence. This is a reasonable trade-off. The condition (10) is mostly a sparsity assumption on the background parameters. The interpretation of the sparsity assumption on η is that there exists a subset of stocks to invest that is sufficient for diversification purpose, by noting that the optimal diversified portfolios are proportional to η . This is generally acceptable because a sparse portfolio has good empirical performance; see Brodie et al. (2009).

We further establish the consistency result for the probability that the LPO portfolio will beat the bank deposit. To illustrate these ideas, we only present the single-period case. Consider

$$F_n := \mathbb{P}(X^{\tilde{u}_s}(T) > x_0(1 + r_f T)) = \Upsilon \left(\sqrt{T} \frac{\beta' \tilde{\eta}}{\sqrt{\tilde{\eta}' \Sigma \tilde{\eta}}} \right),$$

where $\tilde{\eta}$ is defined in (8), Υ is the elliptical distribution function of the single-period return vector, and the subscript “s” is omitted. The oracle benchmark is the one with the perfectly estimated Σ and β :

$$F := \mathbb{P}(X^{u_s}(T) > x_0(1 + r_f T)) = \Upsilon(\sqrt{T} \Theta_p).$$

To explicitly derive the convergence rate for F_n , we impose an additional condition on Υ :

$$(D) \quad \text{For any } x > 0 \text{ and } |\delta| \leq 1, \left| \frac{\Upsilon(x + \delta)}{\Upsilon(x)} - 1 \right| \leq c_1 |\delta| (|x| + 1) e^{c_2 |x\delta|}$$

for some positive constants c_1, c_2 , which are independent of x, δ .

Note that the distribution function $\Upsilon(x)$ with density function of the form $c_3(1 + |x|)^{-w} e^{-c_4 |x|^v}$ satisfies (D), where c_3 and c_4 are positive constants and w and v are constants with $0 < v \leq 2, w \geq 0$ or $v = 0, w > 1$. Some popular choices of the distribution of asset returns, including hyperbolic distributions and t -distributions, satisfy (D).

Theorem 4.2. If the conditions in Theorem 4.1 and condition (D) hold, then

$$\frac{F_n}{F} - 1 = \mathcal{O}_P \left(d_n e^{C d_n} \right)$$

for some positive constant C , where d_n is defined in Theorem 4.1.

Theorem 4.2 shows that the probability of the LPO strategy beating the bank deposit would deviate little from the oracle benchmark, which contrasts with the pitfalls of the plug-in strategy in Section 3.

5. Simulation and empirical studies

In this section, we verify the derived theoretical results by simulations and conduct an empirical study to show the practical performance of the LPO portfolios. All simulations and empirical studies are out-of-sample studies.

Recall that our LPO estimator $\tilde{\eta}$ is obtained by solving the minimization problem:

$$\min |\eta|_1 \quad \text{subject to: } |\hat{\Sigma}_n \eta - \hat{\beta}_n|_\infty \leq \lambda.$$

This type of convex optimization problem also appears in Cai and Liu (2011). The use of ℓ_1 norm allows us to apply the linear programming. We solve this problem by the parametric simplex method, which guarantees machine precision by a primal-dual gap. The details of this approach can be found in Vanderbei (2008). We transform our problem to fit the setup in Vanderbei (2008). By setting $\eta = \eta^+ - \eta^-$, where $\eta^\pm = \max(\pm\eta, 0)$ and one of η^+ or η^- is 0, the above minimization problem can be rewritten as

$$\max (-\mathbf{1}_p - \mathbf{1}_p) \begin{pmatrix} \eta^+ \\ \eta^- \end{pmatrix} \quad \text{subject to: } \begin{pmatrix} \hat{\Sigma}_n & -\hat{\Sigma}_n \\ -\hat{\Sigma}_n & \hat{\Sigma}_n \end{pmatrix} \begin{pmatrix} \eta^+ \\ \eta^- \end{pmatrix} \leq \begin{pmatrix} \lambda \mathbf{1}_p + \hat{\beta}_n \\ \lambda \mathbf{1}_p - \hat{\beta}_n \end{pmatrix}.$$

For a given λ , the parametric simplex method efficiently computes the optimal (η^+, η^-) so that we recover the optimal $\tilde{\eta}$ by $\tilde{\eta} = \eta^+ - \eta^-$.

The possible values of λ are chosen such that a feasible solution exists. We start with the λ equal to the maximum of the absolute values of the entries of $\hat{\beta}_n$. In this case, a feasible (optimal) solution is the zero vector. We then decrease λ gradually until it reaches zero or a level at which the corresponding feasible set is empty. Because the sample covariance matrix is possibly singular, it is likely that the feasible set is empty for a small λ . We denote the feasible set of λ as Λ .

We select the optimal tuning parameter λ among Λ using cross-validation (CV), a standard machine learning procedure. A similar practice of CV in the context of portfolio optimization can be found in Ban et al. (2018). Specifically, we orderly divide the data $\{r^{(1)}, \dots, r^{(n)}\}$ into f groups, where f is the number of folds of cross-validation ($f = \lfloor n/n.test \rfloor$ in this paper, where $n.test$ is the sample size of the test data), because the data are ordered in time. For each iteration, the groups with sizes n/f and $n - n/f$ act as test data and training data, respectively. For each λ , we run our program using the training data and then evaluate it with the test data.

To match the mean-variance criterion, we use the square error with the target mean (z) to evaluate the LPO portfolio in the k th iteration, which is defined as $\widehat{SE}_k^u = (X_k^{\hat{u}}(T) - z)^2$, where $X_k^{\hat{u}}(T)$ is the

terminal wealth by adopting the (single-period or multiperiod) LPO portfolio $\hat{u}$, computed using the test data. We then define the cross-validation estimate as the sum of square errors: $CV(\lambda^u) = \sum_{k=1}^f \widehat{SE}_k^u$. Then, the final choice of λ is $\lambda_{CV}^u = \min_{\lambda} CV(\lambda)$. The function CV approximates $\mathbb{E}[(X^{\hat{u}}(T) - z)^2] = \text{Var}(X^{\hat{u}}(T)) + (\mathbb{E}[X^{\hat{u}}(T)] - z)^2$. If the minimum is attained at several λ s, we choose the smallest one. The method of evaluating performance (the function CV) varies according to the user's objective, such as a (negative) Sharpe ratio. We use the sum of square errors because it is a common statistical measurement and because we expect that it will better capture the properties of MV portfolios.

5.1. Simulation study

Theorems 3.1, 4.1 and 4.2 provide the asymptotic results when $n, p \rightarrow \infty$. It is interesting to verify numerically the derived theoretical results and examine the finite-sample performance of the estimated strategies. We consider two commonly used artificial background covariance matrices that exhibit different sparsity conditions. We then compare the LPO portfolio with the traditional plug-in strategy and the oracle strategy for both single- and multiperiod settings. We suppose that $\hat{\beta}_n = \beta$ in our simulation to highlight the effects of an estimation error on the covariance matrix. This allows us to verify the claims in Section 3.

The simulation is designed as follows. We simulate 6-month daily prices ($n = 126$) of p stocks as training data for parameter estimation. We wish to invest the initial wealth ($x_0 = 1000$) into these stocks with the investment horizon $T = 1/12$ and hence also simulate 1-month daily prices ($n_{test} = 21$) as test data. Our target terminal wealth is set as $z = 1020$. The risk-free interest rate is set as $r_f = 0.01$. All p assets have the common initial price $s_i = 40$. The excess mean returns over the risk-free asset are set as $\beta = (0.3, \dots, 0.3, 0, \dots, 0)'$, where the number of non-zeros is fixed at $c_0 = 10$ while the p varies across simulations. Consider the covariance matrix of the following form: $\Sigma = 0.2^2 \Gamma$. Two models for Γ are studied.

1. Model 1. $\Gamma^{-1} = (B + \delta I)/(1 + \delta)$, where $B = (b_{ij})_{p \times p}$ with independent $b_{ij} = b_{ji} = 0.5 \times \text{Ber}(1, 0.2)$ for $1 \leq i \leq c_0, i < j \leq p$; $b_{ij} = b_{ji} = 0.5$ for $c_0 + 1 \leq i < j \leq p$; $b_{ii} = 1$ for $1 \leq i \leq p$. Here $\text{Ber}(1, 0.2)$ is a Bernoulli random variable that takes a value of 1 with probability 0.2 and 0 with a probability of 0.8; and $\delta = \max(-\lambda_{\min}(B), 0) + 0.05$ to ensure that Γ^{-1} is a positive definite matrix. Finally, the matrix is standardized to have unit diagonals.
2. Model 2. $\Gamma = (\gamma_{ij})_{p \times p}$ with $\gamma_{ij} = 0.8^{|i-j|}$ for $1 \leq i, j \leq p$.

In Model 1, only the first c_0 rows and columns of Ω are sparse; the rest of the matrix is not sparse, and the Σ itself is not sparse. In Model 2, Σ is approximately sparse and Ω is a 3-sparse matrix.

For simplicity and consistency with the assumption in Theorem 3.1, we simulate the daily returns from a normal distribution. It is noteworthy that even our simulation setting practically prevents the simulated daily stock prices from being negative (see footnote 2); some truncation schemes should be considered in simulations for safety. For illustrative and experimental purposes, Appendix C provides the parallel simulation results for log-normally distributed returns, which are shown to have the same interpretations for the case of normal returns in the main context.

For each model and each p , we run 1000 simulations to obtain the out-of-sample performance statistics and the bootstrap percentile confidence interval (CI) with 1000 bootstrap replicates for each statistic. To verify our theory, we use the empirical frequency as an estimate of the probability. We are interested in the following events: 1. each strategy beats the risk-free asset under single- and multiperiod settings; 2. the multiperiod strategy beats the single-period strategy for each implementation method; and 3. the LPO strategy outperforms the plug-in strategy. The multiperiod strategy uses three periods that equally partition the whole investment horizon. We also report the portfolio performance in terms of the Sharpe ratio. The numerical results are presented with the column charts in Figures 1-6. In their legends, [S] (resp. [M]) represents single- (resp. multi-) period MV strategy with the indicated estimation approach. The bootstrap percentile CIs are presented by the error bars of the columns.

Figure 1: Empirical frequency (%) of beating the bank deposit for Model 1.

Figures 1 and 2 show that in the simulations that implement the plug-in approach, the empirical frequency for beating the bank deposit decreases in both the single-period and multiperiod cases as p increases for a large p . Although the frequency does not reach 50%, as predicted by Theorem 3.1, the trend approaches that number. We stress that the simulation is also subject to simulation error and that the convergence rate

Figure 2: Empirical frequency (%) of beating the bank deposit for Model 2.

for the results in Theorem 3.1 is of the order $\sqrt{p/n}$, as shown in the proof. We attempt to further increase the p to realize the full converging trend, but the computational error for the generalized inverse is so large that an error message usually occurs when $p > 1200$. In fact, about 40 of 1000 simulations produce error messages for $p = 1200$. We regenerate the 40 simulations for the plug-in approach to produce results. When p goes beyond 1200, error messages become common. The numerical results are distorted and are thus not reported here.

Figure 3: Empirical frequencies (%) of (three columns on the left for each p) the multiperiod strategy beating the single-period strategies; and (two columns on the right for each p) the LPO strategy beating the plug-in strategy for Model 1.

In contrast, the LPO estimate always exists and converges rapidly. Figures 1 and 2 show that the performance of the LPO strategy approaches the oracle strategy for a large p . The convergence trend may have small fluctuation because the convergence is in the probability sense. Although the LPO strategy is robust to the increase in p , it is still slightly affected by the estimation error in the sample estimates, which

Figure 4: Empirical frequencies (%) of (three columns on the left for each p) the multiperiod strategy beating the single-period strategies; and (two columns on the right for each p) the LPO strategy beating the plug-in strategy for Model 2.

are inputs of the constrained ℓ_1 minimization. However, the LPO approach is consistently better than the plug-in approach. Figures 3 and 4 also demonstrate the application of the LPO approach to multiperiod strategies. The multiperiod strategy clearly outperforms the single-period strategy under the LPO approach, whereas the plug-in approach makes the supposedly better multiperiod strategy worse than the single-period strategy.

To further examine the out-of-sample performance, we also compute the Sharpe ratios for all implementation methods using the 1000 simulated terminal wealths. The numerical results for Models 1 & 2 are summarized in Figures 5 and 6, respectively. The LPO strategy clearly uniformly outperforms the plug-in strategy in terms of the Sharpe ratio.

Figure 5: Sharpe ratio of the simulation studies for Model 1.

In addition to the performance of the estimated strategies, we examine the errors of their terminal

Figure 6: Sharpe ratio of the simulation studies for Model 2.

wealths compared to the oracle ones. Table 1 shows the averaged squared errors of the terminal wealths of the estimated strategies based on the 1000 simulations and scaled by the squared initial wealth x_0 ($= 10^6$): $\frac{1}{10^3 x_0^2} \sum_{k=1}^{10^3} (X_k^{\hat{u}}(T) - X_k^u(T))^2$, where $X_k^{\hat{u}}(T)$ and $X_k^u(T)$ are the terminal wealths of the estimated and oracle strategies in the k th simulation, respectively. The LPO strategy is close to the oracle strategy in our simulation. This numerically verifies the oracle properties in Section 4.1.

Table 1: Averaged squared errors of the terminal wealths of the estimated strategies for Models 1 and 2.

	p	[S] LPO	[S] Plug-in	[M] LPO	[M] Plug-in
Model 1	100	0.013 (0.011,0.014)	0.058 (0.052,0.064)	0.010 (0.008,0.013)	2.306 (1.781,2.900)
	200	0.045 (0.041,0.050)	0.129 (0.118,0.143)	0.052 (0.040,0.066)	0.361 (0.258,0.502)
	400	0.039 (0.036,0.043)	0.141 (0.129,0.153)	0.034 (0.030,0.038)	0.131 (0.119,0.144)
	800	0.055 (0.050,0.059)	0.211 (0.193,0.228)	0.056 (0.049,0.064)	0.225 (0.205,0.246)
	1200	0.095 (0.087,0.104)	0.323 (0.296,0.352)	0.097 (0.088,0.105)	0.330 (0.302,0.360)
Model 2	100	0.023 (0.021,0.026)	0.203 (0.184,0.223)	0.034 (0.026,0.044)	20.179 (12.395,28.688)
	200	0.020 (0.018,0.023)	0.133 (0.120,0.144)	0.020 (0.016,0.025)	0.577 (0.452,0.725)
	400	0.021 (0.019,0.024)	0.100 (0.092,0.109)	0.020 (0.017,0.024)	0.114 (0.095,0.140)
	800	0.029 (0.026,0.032)	0.122 (0.111,0.132)	0.023 (0.020,0.026)	0.112 (0.101,0.123)
	1200	0.031 (0.028,0.034)	0.143 (0.132,0.155)	0.030 (0.026,0.035)	0.145 (0.132,0.158)

In parenthesis is the bootstrap percentile CI of the statistic above it.

5.2. Empirical study

Although our simulation verifies the developed theory, it must be tested with real data. An empirical analysis offers important insights into the advantages and limitations of the LPO approach. Our empirical study uses historical data from the S&P 500 and Russell 2000 components (high-dimensional) and from the six empirical datasets (low-dimensional) used by DeMiguel et al. (2009). The datasets are uploaded to

Mendeley Data (see Pun (2017)). We examine the out-of-sample performance of the minimum-variance portfolios with different estimates and formulations, including LPO, plug-in, LPO-EW, adaptive LPO, Glasso estimatebased, shrinkage estimatebased, factor modelbased portfolios, noshort-selling constrained, and gross exposureconstrained, and compare them to the benchmark: an equally weighted (EW) portfolio. The LPO-EW and the adaptive LPO portfolios will then be detailed. The Glasso estimate of the inverse covariance matrix adopts the algorithm in Friedman et al. (2008), and its incorporation into the minimum-variance portfolio follows Goto and Xu (2015). The shrinkage covariance estimation has a shrinkage target of the constant correlation model as specified in Ledoit and Wolf (2004). Our factor model assumes the widely used Fama-French three-factor model proposed in Fama and French (1993). The formulations of noshorting-selling and gross exposureconstrained portfolios are the same as in Fan et al. (2012), who suggest solving gross exposureconstrained portfolio with LARS-LASSO by tracking the noshorting-selling constrained portfolio. To eliminate this subjectivity, we solve for the gross exposureconstrained portfolio in a manner similar to that in Gandy and Veraart (2013), which is detailed in Appendix A. However, this approach is valid only when $p < n$. Hence, gross exposureconstrained portfolios are considered only in the studies of low-dimensional datasets.

We follow the method of DeMiguel et al. (2009) to evaluate the out-of-sample performance of the sample-based strategies. We only consider the allocation on the risky assets because this is the empirical setup in DeMiguel et al. (2009), especially for the EW (or $1/p$) strategy. This enables us to address the open problem they posted and to compare our results with theirs.

Let $w \in \mathbb{R}^p$ be the relative weights of the portfolio with p risky assets, and $w' \mathbf{1}_p = 1$. The empirical analysis conducted in DeMiguel et al. (2009) is actually for single-period strategies, so we also focus on the single-period strategies. In accordance with the observations of DeMiguel et al. (2009), we examine the minimum-variance portfolio rather than the mean-variance portfolio to obtain better out-of-sample performance and to eliminate the subjectivity in the selection of the target mean level. Estimation of the mean returns is much more demanding than estimation of the covariance matrix. We avoid running computationally intensive Bayesian estimations for each of the p assets. To preserve the stationarity of the asset return series, we use a short time horizon to estimate the parameters, as suggested by Broadie (1993). The issue of estimating the mean is reported by Michaud (1989) and Chopra and Ziemba (1993). It motivates the robust optimization approach with respect to the drift terms. The minimum-variance strategy is also considered by Jagannathan and Ma (2003); Garlappi et al. (2007).

The EW strategy invests equal amounts of money into p risky assets, that is, $w_i^{ew} = 1/p$ or $w^{ew} \propto \mathbf{1}_p$. This strategy is free of estimation error and outperforms all existing sample-based strategies in terms of the various out-of-sample performance measures used by DeMiguel et al. (2009). The optimal minimum-variance strategy invests in assets that have relative weights of $w^{min} \propto \Sigma^{-1} \mathbf{1}_p$. The corresponding implementation of the LPO approach in (8) only requires the replacement of $\hat{\beta}_n$ with $\mathbf{1}_p$.

DeMiguel et al. (2009) focus on the datasets of portfolios in which assets are selected in advance. They also pose an open question for the selection of assets for the EW portfolio. The LPO approach can select assets because only favorable assets are assigned nonzero weights. Hence, we introduce two adaptive strategies based on assets selected with the LPO approach. The first is the LPO-EW portfolio, which applies the EW strategy to those selected assets. The second is the adaptive LPO (AdLPO) portfolio, which re-estimates the covariance matrix of the selected stocks for the minimum-variance strategy. The motivation for the adaptive LPO approach is that it typically has fewer selected stocks than observations (n). In such a situation, the Gram matrix is a consistent estimator for many distributions.

5.2.1. Empirical design.

Our analysis relies on rolling sampling windows. Specifically, we choose an estimation window length $n.train$ data. At each test time t_i , $i = 1, \dots, n.test$, we use the data of the previous $n.train$ data (including the test data for the times before t_i) to estimate the parameters needed to implement a particular strategy. We then compute the return at time t_i using the estimated strategy implemented in $[t_{i-1}, t_i]$ and the data at time t_i . This process is continued until all $n.test$ out-of-sample returns are computed.

The out-of-sample performance is reflected by several performance measures adopted by DeMiguel et al. (2009). The Sharpe ratio is arguably the most common statistic for computation of the ratio between the sample mean $\hat{\mu}$ and the standard deviation $\hat{\sigma}$ of the out-sample returns. Specifically,

$$\widehat{SR} = \hat{\mu}/\hat{\sigma}.$$

To test whether the Sharpe ratios of each strategy and the benchmark EW strategy are statistically distinguishable, we also compute the p -value of the difference using the approach suggested by Memmel (2003).

Because we are concerned about the minimum-variance portfolios, it is also important to separately measure the sample variance of the out-sample returns:

$$\widehat{Var} = \hat{\sigma}^2.$$

The third adopted performance measure is the certainty equivalent (CEQ) return, defined as

$$\widehat{CEQ}(\gamma) = \hat{\mu} - \frac{\gamma}{2} \hat{\sigma}^2,$$

where γ is the risk aversion, which has choices of $\{1, 2, 5\}$ in our study. The certainty equivalent return is the equivalent of a “risk-free” rate for a particular trading strategy.

The last measure is the portfolio turnover:

$$\widehat{Turnover} = \frac{1}{n.test} \sum_{i=1}^{n.test} \sum_{j=1}^p \left(|\hat{w}_{j,t_{i+1}} - \hat{w}_{j,t_i^+}| \right),$$

where $\hat{w}_{j,t_{i+1}}$ is the portfolio weight in asset j at time t_{i+1} , and $\hat{w}_{j,t_i^+}$ is the portfolio weight before rebalancing at time t_{i+1} . The portfolio turnover measures the average percentage of wealth traded for a specific strategy. Therefore, the portfolio turnover is used mainly to determine the possible effects associated with transaction costs. The smaller the $\widehat{Turnover}$, the lower the costs.

5.2.2. High-dimensional data: results and interpretations.

The index component lists are current as of January 7, 2015. The daily³ financial data are downloaded from Yahoo!Finance for January 1, 2013 to December 31, 2014 (504 daily data). For simplicity, we eliminate the assets with incomplete data. The two datasets ultimately contain 486 assets for the S&P 500 dataset and 1420 assets for the Russell 2000 dataset. The empirical results for these two datasets are consistent. Hence, to save space in this paper, we present only the results for the S&P 500 dataset, whereas the results for the Russell 2000 dataset are given in Appendix D. In this subsection, we take $n.train = 252$ (days) and $n.test = 252$ (days).

Using the S&P 500 component data and the performance measures above, Table 2 shows the empirical results of the following competitive strategies: “Glasso” is the minimum-variance strategy that uses the Glasso estimate of the inverse covariance matrix, “NS” is the minimum-variance strategy with no short-selling constraint, “Shrink” is the minimum-variance strategy that uses the shrinkage covariance matrix estimate, and “FF-LPO” is the LPO strategy that uses Fama-French three-factor-modelbased covariance matrix estimate as input.

³We use daily data because they are consistent with our model setup and improve the estimation of the covariance matrix in comparison to monthly or yearly data. In practice, market practitioners analyze daily, or even higher-frequency, data. However, the extension of our approach to high-frequency data is left for a future study.

Table 2: Empirical comparisons of the competitive strategies for S&P 500 components ($p = 486$) in terms of numerous measures.

Strategy	EW	LPO	Plug-in	Glasso	NS	Shrink	LPO-EW	AdLPO	FF-LPO
$\widehat{SR}$ (e-02)	7.813	12.884	8.540	8.926	9.547	10.301	11.466	6.364	17.299
(p -value)	–	(0.112)	(0.454)	(0.100)	(0.301)	(0.350)	(0.254)	(0.402)	(0.120)
$\widehat{Var}$ (e-05)	5.576	2.957	3.671	4.554	2.774	2.603	5.468	3.059	3.319
$\widehat{CEQ} _{\gamma=1}$ (e-04)	5.555	6.859	4.991	5.796	4.889	5.125	8.205	3.367	9.800
$\widehat{CEQ} _{\gamma=2}$ (e-04)	5.277	6.711	4.807	5.568	4.751	4.995	7.931	3.214	9.634
$\widehat{CEQ} _{\gamma=5}$ (e-04)	4.440	6.267	4.257	4.885	4.335	4.605	7.111	2.755	9.136
$\widehat{Turnover}$	0.391	0.750	1.417	0.009	0.085	0.430	0.538	1.293	0.217

The LPO strategy clearly outperforms the EW, plug-in, Glasso-based, no-short-selling, and shrinkage-based strategies in terms of the Sharpe ratio and certainty equivalent. Compared to the plug-in strategy, the adaptive LPO strategy generates a lower out-of-sample variance and turnover rate but also a lower return. The LPO selection procedure stimulates improved trading strategies, such as LPO-EW, and can be incorporated with improved estimates. With the Fama-French research factors (downloaded from Prof. French's website), the covariance matrix estimate is improved, and the LPO portfolio based on it (FF-LPO) has a high Sharpe ratio. However, these factors may not be effective in all datasets, as shown in the empirical studies of low-dimensional datasets, and the determination of factors becomes crucial.

The plug-in strategy is so poor that its turnover statistics are the highest among all available strategies. It is worthy of note that the Glasso-based strategy has the lowest turnover, even if its performance is similar to that of the plug-in strategy. The turnover rate of LPO is, however, larger than the EW. However, given the market constraints, the implementation of the EW strategy could be very expensive for a high-dimension portfolio. Financial securities are purchased in a broad lot size to pay a minimal amount for each stock in the market. The minimal charge for an extensive stock could be high. The EW and many other strategies require the investment of equal amounts of money in all available stocks. With a large p , the minimum entry ticket for the EW strategy could be unaffordable, even for financial institutions. Because the LPO-EW involves only a few stocks, its implementation is less costly and likely more profitable.

DeMiguel et al. (2009) conjecture that a larger portfolio size leads to better performance from the EW

strategy. However, the plug-in approach outperforms the EW in terms of the Sharpe ratio but underperforms in terms of certainty equivalent. More importantly, the LPO-EW that consists of fewer stocks significantly outperforms the EW. In other words, an appropriate reduction of the portfolio size can improve the EW performance. The key is to select the right stocks. In DeMiguel et al. (2009), most of the selected assets are financial indices that have increased in recent decades. The LPO approach is a potential scheme for the selection of stocks and determination of the portfolio size based on the minimum-variance criterion for improving the EW strategy. We speculate that the minimum-variance criterion, together with the sparsity in LPO, will help practitioners to select stocks with the maximal hedging effect. The EW strategy can be a robust strategy. This idea is inspired by DeMiguel et al. (2009), who explore a linear combination of the minimum-variance portfolio and the EW portfolio. Our combination is, however, highly nonlinear.

5.2.3. *Low-dimensional data: results and interpretations.*

Conclusions drawn from a single dataset may not be convincing. In this section, we use the six empirical datasets in DeMiguel et al. (2009) for a thorough examination of the LPO approach. The number of assets in those datasets ranges from 3 to 24, and the test period is between July 1963 and November 2004. The data are collected monthly. We follow the evaluation method in DeMiguel et al. (2009): set $n.train = 120$ (months), and the remaining data in the dataset serve as test data. We refer the readers to DeMiguel et al. (2009) for detailed description of the datasets and the empirical performance of various strategies. Here, we report only the top two strategies in their studies: EW and Plug-in (named “1/N” and “min”, respectively, in DeMiguel et al. (2009)).

Table 3 gives the empirical results for the six empirical datasets, where “GC” is the minimum-variance strategy with a gross-exposure constraint, and “FF” is the minimum-variance strategy with the three-factor-modelbased covariance matrix estimate. We report only the Sharpe ratios with p -values and out-of-sample variances because certainty equivalent usually gives the same conclusion as the Sharpe ratio and because the implications of turnovers are similar to those in the previous section.

We first address the results for the dataset of Mkt/SMB/HML. In this very-low-dimensional setting ($p = 3$), the variable selection of the LPO approach may not be reflected, especially when all underlying assets contribute to portfolio construction (at least in the cross-validation). This is the case when we consider the three Fama-French factors as the entire universe. Hence, one can find that the LPO and Plug-in have identical results. Although this coincidence is not guaranteed because the LPO is data-driven, the probability of occurrence is high when p is small. As for the identity between Plug-in and FF, it is expected because

the FF uses the same underlying assets as factors, which will result in identical estimates.

Table 3: Empirical comparisons of the competitive strategies for six empirical datasets in terms of Sharpe ratios and variances.

		EW	LPO	Plug-in	GC	FF
S&P Sectors	$\widehat{SR}$	0.1876	0.1066 (0.07)	0.0820 (0.05)	0.0951 (0.09)	0.1451 (0.25)
$p = 11$	$\widehat{Var} (10^{-3})$	1.724	1.306	1.378	1.569	1.160
Ind. portfolios	$\widehat{SR}$	0.1353	0.1296 (0.43)	0.1554 (0.30)	0.1174 (0.33)	0.1791 (0.14)
$p = 11$	$\widehat{Var} (10^{-3})$	1.933	1.399	1.437	1.367	1.407
Inter'l portfolios	$\widehat{SR}$	0.1277	0.1495 (0.17)	0.1490 (0.21)	1.568 (0.15)	0.1143 (0.22)
$p = 9$	$\widehat{Var} (10^{-3})$	1.891	1.772	1.806	1.764	2.253
Mkt/SMB/HML	$\widehat{SR}$	0.2240	0.2493 (0.23)	0.2493 (0.23)	0.0675 (0.00)	0.2493 (0.23)
$p = 3$	$\widehat{Var} (10^{-3})$	0.337	0.257	0.257	0.805	0.257
FF-1-factor	$\widehat{SR}$	0.1623	0.2709 (0.00)	0.2778 (0.01)	0.2660 (0.01)	0.1876 (0.31)
$p = 21$	$\widehat{Var} (10^{-3})$	2.890	1.474	1.489	1.450	1.773
FF-4-factor	$\widehat{SR}$	0.1753	0.072 (0.08)	-0.018 (0.01)	0.1225 (0.20)	-0.0376 (0.00)
$p = 24$	$\widehat{Var} (10^{-3})$	2.229	0.087	0.080	0.719	0.082

In parentheses is the p -value of the difference between the Sharpe ratios of each strategy and the EW.

The LPO strategy generally improves the plug-in strategy in terms of Sharpe ratios and variances. The LPO strategy has uniformly lower variances than the EW strategy. In terms of Sharpe ratios, the LPO and EW strategies are comparable; however, it is noteworthy that the EW usually outperforms the LPO when the variance of EW is much larger. As discussed in DeMiguel et al. (2009), the plug-in strategies are comparable with the EW strategy. From our discussion in the Appendix A, it is also expected that the performance of the GC strategy is similar to that of the LPO strategy, but the GC, which is subject to a larger estimation error, shows unstable performance in some datasets. The FF strategy performs well in the S&P sectors as in the previous section. However, LPO outperforms FF in three of five datasets (excluding the draw for Mkt/SMB/HML) in terms of Sharpe ratios. Moreover, the LPO has stably lower out-of-sample variance.

5.2.4. Limitations of LPO and future research.

The LPO approach is not without limitations. All LPO-based approaches, including the LPO, LPO-EW, and adaptive LPO, suffer from a high turnover rate relative to the EW approach. Although the LPO approach significantly improves the profit-based performance measures of high-dimensional portfolios, the estimation stability remains a concern. The optimized portfolio weights change quite frequently and eventually increase the transaction costs in practice. In other words, the unstable estimates of Σ continue to have some effects on the estimated cardinality of η , and LPO-EW therefore still generates a high turnover rate. Therefore, the set of favorable stocks changes with the sampling window. Statistically, this problem is known as the false-positive identification problem. Future studies may consider additional statistical treatments to lower the false-positive rate.

We stress that the aim of the LPO strategy is to construct an estimation for the optimal strategy or for effective parameters in the optimal strategy. We use the constrained ℓ_1 minimization to realize that concept. However, it could be interesting to explore other possibilities. The high-dimensional Bayesian approach is probably an alternative. A recent study by Olivares-Nadal and DeMiguel (2018) demonstrates that the addition of a transaction cost constraint as an additional regularization term may improve the portfolio's performance.

Although we demonstrate the use of LPO in single- and multiperiod MV portfolios, the concept should be generally applicable to continuous-time models and other objective functions as in Gao et al. (2016); Yao et al. (2016); Pun and Wong (2016); Chiu et al. (2017). We have put this into our research agenda.

6. Conclusions

For those who wish to invest in a few favorable assets rather than in all available assets in the market, a sparse optimal portfolio is an attractive approach. This paper proposes the LPO approach using the constrained ℓ_1 minimization with this practical purpose in mind. Investing in an optimized portfolio is not necessarily optimal due to the difficulty of controlling estimation errors. We prove mathematically that the estimation error on the covariance matrix alone makes an optimized high-dimensional portfolio tend to perform similar to that based on a random stock picking strategy. The portfolio optimized with the LPO estimator, however, asymptotically converges to the true optimal portfolio in probability. The LPO approach not only improves the performance of static MV portfolios but is also useful for dynamic MV portfolios. The derived theory is verified with several simulation studies. Empirical analysis that contrasts the LPO

approach with other competitive implementations further confirms the advantages of the LPO approach and its generalizations to LPO-EW and adaptive LPO portfolios.

The analysis of this paper is based on the assumption of unconstrained markets. In a market with portfolio constraints, for instance, a no-shorting constraint, the optimal strategy does not admit a closed-form expression and relies on some numerical schemes. We refer to Cui et al. (2014) for a detailed investigation of mean-variance policy in no-short-selling and cone-constrained markets. An interesting topic for future study is how to incorporate the related numerical schemes with the LPO approach.

References

- Ban, G.-Y., N. El Karoui, A.E.B. Lim. 2018. Machine Learning and Portfolio Optimization. *Management Science*. **64** 1136-1154.
- Basak, S., G. Chabakauri. 2010. Dynamic Mean-Variance Asset Allocation. *The Review of Financial Studies*. **23** 2970-3016.
- Bickel, P.J., E. Levina. 2004. Some Theory for Fisher's Linear Discriminant Function, 'Naive Bayes', and Some Alternatives When There Are Many More Variables Than Observations. *Bernoulli*. **10** 989-1010.
- Bickel, P.J., Y. Ritov, A.B. Tsybakov. 2009. Simultaneous Analysis of LASSO and Dantzig Selector. *The Annals of Statistics*. **37** 1705-1732.
- Bielecki, T., H. Jin, S. Pliska, X. Zhou. 2005. Continuous-Time Mean-Variance Portfolio Selection with Bankruptcy Prohibition. *Mathematical Finance*. **15** 213-244.
- Boyle, P., L. Garlappi, R. Uppal, T. Wang. 2012. Keynes Meets Markowitz: The Trade-off Between Familiarity and Diversification. *Management Science*. **58** 253-272.
- Broadie, M. 1993. Computing Efficient Frontiers using Estimated Parameters. *Annals of Operations Research*. **45** 21-58.
- Brodie, J., I. Daubechies, C. De Mol, D. Giannone, I. Loris. 2009. Sparse and Stable Markowitz Portfolios. *Proceedings of the National Academy of Sciences of the United States of America*. **106** 12267-12272.
- Cai, T.T., C.H. Zhang, H.H. Zhou. 2010. Optimal Rates of Convergence for Covariance Matrix Estimation. *The Annals of Statistics*. **38** 2118-2144.
- Cai, T.T., W.D. Liu. 2011. A Direct Estimation Approach to Sparse Linear Discriminant Analysis. *Journal of the American Statistical Association*. **106** 1566-1577.
- Candès, E., T. Tao. 2007. The Dantzig Selector: Statistical Estimation When p Is Much Larger Than n . *The Annals of Statistics*. **35** 2313-2351.
- Chiu, M.C., C.S. Pun, H.Y. Wong. 2017. Big Data Challenges of High-Dimensional Continuous-Time Mean-Variance Portfolio Selection and a Remedy. *Risk Analysis*. **37**, 1532-1549.
- Chopra, V.K., W.T. Ziemba. 1993. The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management*. **19** 6-11.
- Cui, X., J. Gao, X. Li, D. Li. 2014. Optimal Multi-Period Mean-Variance Policy under No-Shorting Constraint. *European Journal of Operational Research*. **234** 459-468.
- DeMiguel, V., L. Garlappi, R. Uppal. 2009. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*. **22** 1915-1953.

- DeMiguel, V., L. Garlappi, F.J. Nogales, R. Uppal. 2009. A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management Science*. **55** 798-812.
- Fama, E., K. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*. **33** 3-56.
- Fan, J., Y. Fan, J. Lv. 2008. High-dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics*. **147** 186-197.
- Fan, J., J. Zhang, K. Yu. 2012. Vast Portfolio Selection with Gross-Exposure Constraints. *Journal of the American Statistical Association*. **107** 592-606.
- Fouque, J.-P., C.S. Pun, H.Y. Wong. 2016. Portfolio Optimization with Ambiguous Correlation and Stochastic Volatilities. *SIAM Journal on Control and Optimization*. **54** 2309-2338.
- Friedman, J., T. Hastie, R. Tibshirani. 2008. Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics*. **9** 432-441.
- Gandy, A., L.A.M. Veraart. 2013. The Effect of Estimation in High-Dimensional Portfolios. *Mathematical Finance*. **23** 531-559.
- Gao, J., Y. Xiong, D. Li. 2016. Dynamic Mean-Risk Portfolio Selection with Multiple Risk Measures in Continuous-Time. *European Journal of Operational Research*. **249** 647-656.
- Garlappi, L., R. Uppal, T. Wang. 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *The Review of Financial Studies*. **20** 41-81.
- Goto S., Y. Xu. 2015. Improving Mean Variance Optimization through Sparse Hedging Restrictions. *Journal of Financial and Quantitative Analysis*. **50** 1415-1441.
- Jagannathan, R., T. Ma. 2003. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*. **58** 1651-1684.
- James, G.M., P. Radchenko, J. Lv. 2009. DASSO: Connections Between the Dantzig Selector and LASSO. *Journal of the Royal Statistical Society, Series B*. **71** 127-142.
- Ledoit, O., M. Wolf. 2004. Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management*. **30** 110-119.
- Lee, S., F. Zou, F.A. Wright. 2014. Convergence of Sample Eigenvalues, Eigenvectors, and Principal Component Scores for Ultra-high-dimensional Data. *Biometrika*. **101** 484-490.
- Li, D., W.L. Ng. 2000. Optimal Dynamic Portfolio Selection: Multiperiod Mean-Variance Formulation. *Math. Finan.* **10** 387-406.
- Luenberger, D.G. 1998. *Investment Science*. Wiley, New York.
- Markowitz, H.M. 1989. *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Cambridge, MA: Basil Blackwell.
- Memmel, C. 2003. Performance Hypothesis Testing with the Sharpe Ratio. *Finance Letters*. **1** 21-23.
- Michaud, R.O. 1989. The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal*. **45** 31-42.
- Novak, S.Y. 2012. *Extreme Value Methods with Applications to Finance*, CRC Press.
- Olivares-Nadal, A.V., V. DeMiguel. 2018. Technical Note—A Robust Perspective on Transaction Costs in Portfolio Optimization. *Operations Research*. **66** 733-739.
- Pun, C.S. 2017. *Low- and High-Dimensional Stock Price Data*. Mendeley Data, v2. <http://dx.doi.org/10.17632/ndxfrshm74.2>.
- Pun, C.S. 2018. Time-Consistent Mean-Variance Portfolio Selection with Only Risky Assets. *Economic Modelling*.
- Pun, C.S., H.Y. Wong. 2016. Resolution of Degeneracy in Merton's Portfolio Problem. *SIAM Journal on Financial Mathematics*. **7**, 786-811.

- Vanderbei, R. 2008. *Linear Programming, Foundations and Extensions*. Springer.
- Yao, H., Z. Li, D. Li. 2016. Multi-Period Mean-Variance Portfolio Selection with Stochastic Interest Rate and Uncontrollable Liability. *European Journal of Operational Research*. **252** 837-851.
- Zhou, X.Y., D. Li. 2000. Continuous-Time Mean-Variance Portfolio Selection: A Stochastic LQ Framework. *Applied Mathematics & Optimization*. **42** 19-33.

Appendix A. Connections between LPO and ℓ_1 norm constrained portfolios

In this section, we investigate the relationship between the LPO portfolio and the ℓ_1 norm-constrained portfolio. These two portfolios are constructed based on the ℓ_1 minimization (Dantzig Selector) approach and the least absolute shrinkage and selection operator (LASSO) approach, whose relationship is discussed by Bickel, Ritov, and Tsybakov (2009) and James et al. (2009).

The minimization problem in (8) shares many similarities with the LASSO optimization problem:

$$\min |\pi|_1 \quad \text{subject to } \|\hat{\sigma}'_n \pi - \hat{\sigma}_n^{-1} \hat{\beta}_n\|_2^2 \leq \delta_n \quad (\text{A.1})$$

for some nonnegative δ_n , where $\hat{\sigma}_n$ is the empirical volatility matrix such that $\hat{\sigma}_n \hat{\sigma}'_n = \hat{\Sigma}_n$. We call the solution to the LASSO-type problem (A.1) the LASSO solution. As noted by James et al. (2009), when $p = 2$, these two solutions are identical. However, the equivalence of these two solutions in general does not hold in $p \geq 3$ dimensions. When the tuning parameters δ_n and λ_n are the same, the LASSO solution is always a feasible solution of (8). Hence, when these two solutions are not identical, the LPO estimator is sparser than the LASSO solution in terms of ℓ_1 norm. The conditions for their equivalence are provided in James et al. (2009).

Under mild conditions, the LASSO solution to (A.1) is equal to the solution to

$$\min \|\hat{\sigma}'_n \pi - \hat{\sigma}_n^{-1} \hat{\beta}_n\|_2^2 \quad \text{subject to } |\pi|_1 \leq c \quad (\text{A.2})$$

for some nonnegative c . The solution to (A.2) is ℓ_1 norm-constrained portfolios considered in Gandy and Veraart (2013). By the analyses by Bickel, Ritov, and Tsybakov (2009) and James et al. (2009), the LPO portfolio and the ℓ_1 norm constrained portfolio often behave the same when $n > p$. However, (A.2) requires $\hat{\sigma}_n^{-1}$ as input that essentially assumes that the Gram matrix $\hat{\Sigma}_n$ is invertible. In the high-dimensional case of $p > n$, $\hat{\Sigma}_n$ is singular, and substitution of the generalized inverse of $\hat{\sigma}_n$ into (A.2) would lead to highly biased results. Compared to the LASSO formulation (ℓ_1 minimization with quadratic constraints), our proposed LPO approach has the distinctive advantage of applicability for a singular $\hat{\Sigma}_n$ and any other improved estimator of Σ .

Another competitive portfolio is the ℓ_1 norm-constrained minimum-variance portfolio proposed by Fan et al. (2012), which is the solution to the risk minimization problem

$$\min w' \hat{\Sigma}_n w \quad \text{subject to } w' \mathbf{1} = 1, |w|_1 \leq c. \quad (\text{A.3})$$

Its implementation relies on a least-angle regression (LARS)-LASSO algorithm to approximately locate the solution path, and it depends on the choice of the tracking portfolio. Although such an approximation is proven useful in the static portfolio selection, it is limited to single-period (quadratic objective) problems. By recognizing the connection between LASSO and Dantzig Selector, the proposed LPO approach aims to extend the high-dimensional portfolio concept to dynamic portfolio management. Specifically, the LPO minimum-variance portfolio can be easily implemented with the following two steps:

1. find $\tilde{\eta} \in \arg \min_{\eta \in \mathbb{R}^p} \left\{ |\eta|_1 \text{ subject to } |\hat{\Sigma}_n \eta - \mathbf{1}|_\infty \leq \lambda_n \right\}$;
2. normalize it: $\tilde{w} := \tilde{\eta} / \mathbf{1}' \tilde{\eta}$.

Moreover, it can be proved that the LPO minimum-variance portfolios have the desired oracle properties.

Appendix B. All proofs of propositions and theorems

Proof of Proposition 2.1. Both (3) and (4) are classic equations in the literature. We restate (3) from Li and Ng (2000) into an alternative, but equivalent, form to facilitate the use of the notation in this paper. Let $P_j = R(t_j, t_{j+1}) - r_f \Delta_j \mathbf{1}_p$ and read (1) as $X^{u_m}(t_j)(1 + r_f \Delta_j) + u_m(t_j)' P_j$. Given the results stated in Li and Ng (2000), our result follows a simple calculation, noting that

$$\mathbb{E}[P_j] = \Delta_j \beta_j, \quad \mathbb{E}[P_j P_j'] = \Delta_j \Sigma_j + \Delta_j^2 \beta_j \beta_j', \quad (\mathbb{E}[P_j P_j'])^{-1} \mathbb{E}[P_j] = \frac{\Sigma_j^{-1} \beta_j}{1 + \Delta_j \beta_j' \Sigma_j^{-1} \beta_j}.$$

□

To prove Theorem 3.1, we require the following two lemmas and one proposition.

Lemma Appendix B.1. For $\epsilon_{n,p} > 0$ (possibly depends on n, p),

$$\mathbb{P}_D \left(\left| \prod_{j=0}^{N-1} [1 - \hat{\eta}_n' \sigma \sqrt{\Delta_j} \xi_j] - \left(1 - \hat{\eta}_n' \sigma \sum_{j=0}^{N-1} \sqrt{\Delta_j} \xi_j \right) \right| > \epsilon_{n,p} \right) \leq \mathcal{O} \left(\left[\frac{\hat{\eta}_n' \Sigma \hat{\eta}_n}{\epsilon_{n,p}} \right]^2 \right).$$

Proof of Lemma Appendix B.1. Recall that $\hat{\eta}_n$ is adapted to D and $\xi_j, j = 0, \dots, N-1$ are independent random vectors. We first note that

$$\begin{aligned} \mathbb{E}_D \left[\prod_{j=0}^{N-1} (1 - \hat{\eta}_n' \sigma \sqrt{\Delta_j} \xi_j)^2 \right] &= \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\eta}_n' \Sigma \hat{\eta}_n), \\ \mathbb{E}_D \left[\left(1 - \hat{\eta}_n' \sigma \sum_{j=0}^{N-1} \sqrt{\Delta_j} \xi_j \right)^2 \right] &= 1 + T \hat{\eta}_n' \Sigma \hat{\eta}_n, \\ \mathbb{E}_D \left[\prod_{j=0}^{N-1} (1 - \hat{\eta}_n' \sigma \sqrt{\Delta_j} \xi_j) (-\hat{\eta}_n' \sigma \sqrt{\Delta_l} \xi_l) \right] &= \Delta_l \hat{\eta}_n' \Sigma \hat{\eta}_n, \end{aligned}$$

for $l = 0, 1, \dots, N-1$, and thus

$$\begin{aligned} & \mathbb{E}_D \left[\left| \prod_{j=0}^{N-1} [1 - \hat{\eta}'_n \sigma \sqrt{\Delta_j} \xi_j] - (1 - \hat{\eta}'_n \sigma \sum_{j=0}^{N-1} \sqrt{\Delta_j} \xi_j) \right|^2 \right] \\ &= \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\eta}'_n \Sigma \hat{\eta}_n) - (1 + T \hat{\eta}'_n \Sigma \hat{\eta}_n) = \mathcal{O}([\hat{\eta}'_n \Sigma \hat{\eta}_n]^2). \end{aligned}$$

The Markov-type inequality in Lemma Appendix B.1 is proved as

$$\begin{aligned} & \mathbb{P}_D \left(\left| \prod_{j=0}^{N-1} [1 - \hat{\eta}'_n \sigma \sqrt{\Delta_j} \xi_j] - \left(1 - \hat{\eta}'_n \sigma \sum_{j=0}^{N-1} \sqrt{\Delta_j} \xi_j \right) \right| > \epsilon_{n,p} \right) \\ & \leq \mathbb{E}_D \left[\left| \prod_{j=0}^{N-1} [1 - \hat{\eta}'_n \sigma \sqrt{\Delta_j} \xi_j] - (1 - \hat{\eta}'_n \sigma \sum_{j=0}^{N-1} \sqrt{\Delta_j} \xi_j) \right|^2 \right] / \epsilon_{n,p}^2 = \mathcal{O} \left(\left[\frac{\hat{\eta}'_n \Sigma \hat{\eta}_n}{\epsilon_{n,p}} \right]^2 \right). \end{aligned}$$

□

Lemma Appendix B.2. Let $\hat{\eta}_n = \Sigma_n^- \beta$. Then,

1. $\hat{\eta}'_n \Sigma \hat{\eta}_n / \beta' \hat{\eta}_n \xrightarrow{\mathbb{P}} 0$ as $p/n \rightarrow \infty$;
2. If $\beta' \Omega \beta = o\left(\frac{p}{n}\right)$, then $\beta' \hat{\eta}_n / \sqrt{\hat{\eta}'_n \Sigma \hat{\eta}_n} \xrightarrow{\mathbb{P}} 0$ (and in $\mathcal{L}^2$) as $p/n \rightarrow \infty$; and
3. If $\beta' \Omega \beta = o\left(\frac{p}{n}\right)$, then $\sqrt{\hat{\eta}'_n \Sigma \hat{\eta}_n} \xrightarrow{\mathbb{P}} 0$ and $\beta' \hat{\eta}_n \xrightarrow{\mathbb{P}} 0$ as $p/n \rightarrow \infty$.

Proof of Lemma Appendix B.2. The third assertion is a direct consequence of the first and second assertions according to Slutsky's theorem. Hence, we only prove the first and second assertions.

Consider the identity covariance matrix $\Sigma = I$. The eigenvectors of $\hat{\Sigma}_n$, denoted as $\hat{\xi}_1, \dots, \hat{\xi}_n$, which correspond to non-zero eigenvalues $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n$ of $\hat{\Sigma}_n$, span a subspace of $\mathbb{R}^p$. Moreover, $(\hat{\lambda}_1, \dots, \hat{\lambda}_n)$ and $(\hat{\xi}_1, \dots, \hat{\xi}_n)$ are independent, whereas $\hat{\xi}_j$ are identically distributed uniformly on the unit p -sphere. The Moore-Penrose inverse of $\hat{\Sigma}_n$ can then be expressed as

$$\hat{\Sigma}_n^- = \sum_{i=1}^n \frac{1}{\hat{\lambda}_i} \hat{\xi}_i \hat{\xi}_i'.$$

1. As proven in the Theorem 1(ii) in Lee et al. (2014), $\hat{\lambda}_i = \mathcal{O}_P\left(\frac{p}{n}\right)$ for all i as $\frac{p}{n} \rightarrow \infty$. We recognize that

$$\frac{\hat{\eta}'_n \hat{\eta}_n}{\beta' \hat{\eta}_n} = \frac{\sum_{i=1}^n \frac{1}{\hat{\lambda}_i^2} (\beta' \hat{\xi}_i)^2}{\sum_{i=1}^n \frac{1}{\hat{\lambda}_i} (\beta' \hat{\xi}_i)^2} \leq \frac{\hat{\lambda}_1}{\hat{\lambda}_n^2} = \mathcal{O}_P\left(\frac{n}{p}\right).$$

Hence, $\hat{\eta}'_n \hat{\eta}_n / \beta' \hat{\eta}_n$ converges to 0 in probability as $p/n \rightarrow \infty$.

2. According to the Cauchy-Schwarz inequality,

$$\left[\frac{\beta' \hat{\eta}_n}{\sqrt{\hat{\eta}_n' \hat{\eta}_n}} \right]^2 = \frac{[\sum_{i=1}^n \frac{1}{\hat{\lambda}_i} (\beta' \hat{\xi}_i)^2]^2}{\sum_{i=1}^n \frac{1}{\hat{\lambda}_i^2} (\beta' \hat{\xi}_i)^2} \leq \sum_{i=1}^n (\beta' \hat{\xi}_i)^2.$$

The expectation of the right-hand side in the above inequality is equal to $(n/p) \|\beta\|^2$. Hence, $\beta' \hat{\eta}_n / \sqrt{\hat{\eta}_n' \hat{\eta}_n}$ converges to 0 in $\mathcal{L}^2$ once $\|\beta\|^2 = o(\frac{p}{n})$.

Yet, consider the general positive definite Σ . We take $\beta^* = \Sigma^{-\frac{1}{2}} \beta$, $\hat{\Sigma}_n^* = \Sigma^{-\frac{1}{2}} \hat{\Sigma}_n \Sigma^{-\frac{1}{2}}$ and $\hat{\eta}_n^* = \hat{\Sigma}_n^{*-1} \beta^* = \Sigma^{\frac{1}{2}} \hat{\eta}_n$, where $\hat{\Sigma}_n^*$ is a Gram matrix of pseudo-samples $\{r^{*(l)}\}_{l=1}^n$ and $r^{*(l)} = (\delta t \Sigma)^{-\frac{1}{2}} r^{(l)}$ has the identity covariance matrix. Notice that $\frac{\hat{\eta}_n' \Sigma \hat{\eta}_n}{\beta' \hat{\eta}_n} = \frac{\hat{\eta}_n^{*'} \hat{\eta}_n^*}{\beta^{*'} \hat{\eta}_n^*}$ and $\frac{\beta' \hat{\eta}_n}{\sqrt{\hat{\eta}_n' \Sigma \hat{\eta}_n}} = \frac{\beta^{*'} \hat{\eta}_n^*}{\sqrt{\hat{\eta}_n^{*'} \hat{\eta}_n^*}}$. According to the results of the case of the identity covariance matrix, if the condition for $\|\beta^*\|^2 = \Theta_p$ is satisfied, then $\hat{\eta}_n' \Sigma \hat{\eta}_n / \beta' \hat{\eta}_n$ and $\beta' \hat{\eta}_n / \sqrt{\hat{\eta}_n' \Sigma \hat{\eta}_n}$ converges to 0 in probability as $p/n \rightarrow \infty$.

□

For any estimators $(\hat{\eta}, \hat{\beta})$, we can solve for $X^{\hat{u}_s}(T)$ and $X^{\hat{u}_m}(T)$ and their means and variances.

Proposition Appendix B.1. Given the estimators for $(\hat{\eta}_s, \hat{\beta}_s)$ and $(\hat{\eta}_m, \hat{\beta}_m)$, the terminal wealths become

$$X^{\hat{u}_s}(T) = z \frac{\beta' \hat{\eta}}{\hat{\beta}' \hat{\eta}} + x_0(1 + r_f T) \left[1 - \frac{\beta' \hat{\eta}}{\hat{\beta}' \hat{\eta}} \right] + \frac{z - x_0(1 + r_f T)}{\sqrt{T} \hat{\beta}' \hat{\eta}} \hat{\eta}' \sigma \xi_0, \quad (\text{B.1})$$

$$X^{\hat{u}_m}(T) = \hat{\gamma}_m + \left(x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k) - \hat{\gamma}_m \right) \prod_{j=0}^{N-1} \left(\frac{1 - \Delta_j (\beta - \hat{\beta})' \hat{\eta} - \sqrt{\Delta_j} \hat{\eta}' \sigma \xi_j}{1 + \Delta_j \hat{\beta}' \hat{\eta}} \right), \quad (\text{B.2})$$

where $\hat{\gamma}_m$ is defined in (5). Moreover, the portfolio means read

$$\mathbb{E}[X^{\hat{u}_s}(T)] = z + (z - x_0(1 + r_f T)) \left[\frac{\beta' \hat{\eta}}{\hat{\beta}' \hat{\eta}} - 1 \right], \quad (\text{B.3})$$

$$\mathbb{E}[X^{\hat{u}_m}(T)] = z + \frac{z - x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k)}{\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}' \hat{\eta}) - 1} \left[1 - \prod_{j=0}^{N-1} (1 - \Delta_j (\beta - \hat{\beta})' \hat{\eta}) \right], \quad (\text{B.4})$$

and variances read

$$(\sigma_T^{\hat{u}_s})^2 = (z - x_0(1 + r_f T))^2 \frac{\hat{\eta}' \Sigma \hat{\eta}}{T (\hat{\beta}' \hat{\eta})^2}, \quad (\text{B.5})$$

$$\begin{aligned} (\sigma_T^{\hat{u}_m})^2 &= \frac{\left(z - x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k) \right)^2}{\left(\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}' \hat{\eta}) - 1 \right)^2} \left[\prod_{j=0}^{N-1} [(1 - \Delta_j (\beta - \hat{\beta})' \hat{\eta})^2 + \Delta_j \hat{\eta}' \Sigma \hat{\eta}] \right. \\ &\quad \left. - \prod_{j=0}^{N-1} (1 - \Delta_j (\beta - \hat{\beta})' \hat{\eta})^2 \right]. \end{aligned} \quad (\text{B.6})$$

Proof of Proposition Appendix B.1. The single-period result is the special case of the multiperiod result with $N = 1$. For the multiperiod case, we use the recursive relationship (1):

$$\begin{aligned}
X^{\hat{u}_m}(t_{j+1}) &= X^{\hat{u}_m}(t_j)(1 + r_f \Delta_j) + \hat{u}_m(t_j)' P_j \\
&= (1 + r_f \Delta_j) \left[1 - \frac{\hat{\eta}' P_j}{1 + \Delta_j \hat{\beta}' \hat{\eta}} \right] X^{\hat{u}_m}(t_j) + \frac{\gamma_m \prod_{k=j+1}^{N-1} (1 + r_f \Delta_k)^{-1}}{1 + \Delta_j \hat{\beta}' \hat{\eta}} \hat{\eta}' P_j \\
&=: a_j X^{\hat{u}_m}(t_j) + b_j,
\end{aligned}$$

where γ_m is defined in (5). Then $X^{\hat{u}_m}(T) = \prod_{j=0}^{N-1} a_j x_0 + b_{N-1} + \sum_{j=0}^{N-2} (b_j \prod_{k=j+1}^{N-1} a_k)$. Some simple but tedious calculations bring about the desirable expression of $X^{\hat{u}_m}(T)$. We then compute its first two moments:

$$\begin{aligned}
\mathbb{E}[X^{\hat{u}_m}(T)] &= \gamma_m + \left(x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k) - \gamma_m \right) \prod_{j=0}^{N-1} \frac{1 - \Delta_j (\beta - \hat{\beta})' \hat{\eta}}{1 + \Delta_j \hat{\beta}' \hat{\eta}}, \\
\mathbb{E}[(X^{\hat{u}_m}(T))^2] &= \gamma_m^2 + 2\gamma_m \left(x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k) - \gamma_m \right) \prod_{j=0}^{N-1} \frac{1 - \Delta_j (\beta - \hat{\beta})' \hat{\eta}}{1 + \Delta_j \hat{\beta}' \hat{\eta}} \\
&\quad + \left(x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k) - \gamma_m \right)^2 \prod_{j=0}^{N-1} \frac{(1 - \Delta_j (\beta - \hat{\beta})' \hat{\eta})^2 + \Delta_j \hat{\eta}' \Sigma \hat{\eta}}{(1 + \Delta_j \hat{\beta}' \hat{\eta})^2}.
\end{aligned}$$

The expressions of the mean and the variance then follow by substitution of γ_m in (5) and simplification. $\square$

Proof of Theorem 3.1. Using (B.1) with $\hat{\beta} = \beta$, it is easy to have

$$\mathbb{P}_D(X^{\hat{u}_s}(T) > x_0(1 + r_f T)) = \Phi \left(\sqrt{T} \frac{\beta' \hat{\eta}_n}{\sqrt{\hat{\eta}_n' \Sigma \hat{\eta}_n}} \right) \xrightarrow{P} \frac{1}{2},$$

where the Lemma Appendix B.2 and the continuous mapping theorem are used.

For the multiperiod case, we have

$$\begin{aligned}
&\mathbb{P}_D \left(X^{\hat{u}_m}(T) > x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k) \right) \\
&= \mathbb{P} \left(1 - \prod_{j=0}^{N-1} (1 + \Delta_j \beta' \hat{\eta}_n)^{-1} \prod_{j=0}^{N-1} [1 - \sqrt{\Delta_j} \hat{\eta}_n' \sigma \xi_j] > 0 \right) \\
&\leq \mathbb{P}(1 - \hat{\eta}_n' \sigma \sum_{j=0}^{N-1} \sqrt{\Delta_j} \xi_j - \epsilon_{n,p} < 1 + \beta' \hat{\eta}_n T + \mathcal{O}((\beta' \hat{\eta}_n)^2)) + \mathcal{O} \left(\left[\frac{\hat{\eta}_n' \Sigma \hat{\eta}_n}{\epsilon_{n,p}} \right]^2 \right) \\
&= \Phi \left(\frac{\beta' \hat{\eta}_n T + \mathcal{O}((\beta' \hat{\eta}_n)^2) + \epsilon_{n,p}}{\sqrt{\hat{\eta}_n' \Sigma \hat{\eta}_n T}} \right) + \mathcal{O} \left(\left[\frac{\hat{\eta}_n' \Sigma \hat{\eta}_n}{\epsilon_{n,p}} \right]^2 \right).
\end{aligned}$$

By taking $\epsilon_{n,p} = (\hat{\eta}'_n \Sigma \hat{\eta}_n)^{\frac{3}{4}}$ and using the Lemma Appendix B.2, the upper bound of $\mathbb{P}(X^{\hat{u}_m}(T) > x_0(1 + r_f T))$ converges to 1/2 in probability. $\square$

To prove Theorems 4.1 and 4.2 in Section 4.1, we require Lemma 2 in Cai and Liu (2011), which is restated using our notation as follows.

Lemma Appendix B.3. Under (M1) and (C1), we have a probability greater than $1 - \mathcal{O}(p^{-1})$ that

$$|\hat{\Sigma}_n \eta - \hat{\beta}_n|_{\infty} \leq \lambda_n. \quad (\text{B.7})$$

Under (M2) and (C1), (B.7) holds with a probability greater than $1 - \mathcal{O}(p^{-1} + n^{-\epsilon/8})$.

Remark: By the definition of $\tilde{\eta}$, we have a high probability that $|\tilde{\eta}|_1 \leq |\eta|_1$.

Proof of Lemma Appendix B.3. In Lemma 2 in Cai and Liu (2011), we let $X = \beta \delta t + \sqrt{\delta t} \Psi$ and $Y = 0$. We then have $\delta t |\hat{\Sigma}_n \eta - \hat{\beta}_n|_{\infty} \leq C \sqrt{\Theta_p \log p/n}$, and thus the result follows. $\square$

Proof of Theorem 4.1. By the definition of $\tilde{\eta}$, we have

$$|(\eta)' \hat{\Sigma}_n \tilde{\eta} - (\eta)' \beta| \leq (\lambda_n + |\hat{\beta}_n - \beta|_{\infty}) |\eta|_1 \leq (\lambda_n + b_n) |\eta|_1.$$

Using (B.7) (with a different moment condition), we also have

$$|(\eta)' \hat{\Sigma}_n \tilde{\eta} - \beta' \tilde{\eta}| \leq (\lambda_n + |\hat{\beta}_n - \beta|_{\infty}) |\tilde{\eta}|_1 \leq (\lambda_n + b_n) |\eta|_1.$$

Combining the above two inequalities yields

$$|\beta' \tilde{\eta} - \Theta_p| \leq 2(\lambda_n + b_n) |\eta|_1 \leq C(\lambda_n + b_n) |\eta|_1, \quad (\text{B.8})$$

$$\text{and } |\hat{\beta}'_n \tilde{\eta} - \Theta_p| \leq (2\lambda_n + 3b_n) |\eta|_1 \leq C(\lambda_n + b_n) |\eta|_1. \quad (\text{B.9})$$

Next, we notice that $|\Sigma \tilde{\eta} - \hat{\beta}_n|_{\infty} \leq |\Sigma - \hat{\Sigma}_n|_{\infty} |\tilde{\eta}|_1 + \lambda_n \leq a_n |\eta|_1 + \lambda_n$, and thus

$$|\tilde{\eta}' \Sigma \tilde{\eta} - \hat{\beta}'_n \tilde{\eta}| \leq a_n |\eta|_1^2 + \lambda_n |\eta|_1 \leq d_n, \quad (\text{B.10})$$

$$|\tilde{\eta}' \Sigma \tilde{\eta} - \beta' \tilde{\eta}| \leq a_n |\eta|_1^2 + (\lambda_n + b_n) |\eta|_1 = d_n. \quad (\text{B.11})$$

Now we investigate the oracle properties of the LPO strategies under the single-period and multiperiod cases.

1. For the single-period strategy given by (B.1), we have

$$X^{\tilde{u}_s}(T) - X^{u_s}(T) = (z - x_0(1 + r_f T)) \left[\frac{\beta'_n \tilde{\eta}}{\hat{\beta}'_n \tilde{\eta}} - 1 + \frac{1}{\sqrt{T}} \left(\frac{\tilde{\eta}' \sigma \xi}{\hat{\beta}'_n \tilde{\eta}} - \frac{\eta' \sigma \xi}{\Theta_p} \right) \right].$$

Notice that (B.9) implies,

$$K \leq \Theta_p \leq \hat{\beta}'_n \tilde{\eta} + |\hat{\beta}'_n \tilde{\eta} - \Theta_p| \leq \hat{\beta}'_n \tilde{\eta} + C(\lambda_n + b_n)|\eta|_1 \Rightarrow \hat{\beta}'_n \tilde{\eta} \geq K - o_P(1). \quad (\text{B.12})$$

It is also clear that conditional on D , $\frac{\beta'_n \tilde{\eta}}{\hat{\beta}'_n \tilde{\eta}} - 1$ and $\frac{\tilde{\eta}' \sigma \xi}{\hat{\beta}'_n \tilde{\eta}} - \frac{\eta' \sigma \xi}{\Theta_p}$ are orthogonal. Taking conditional expectation with respect to the historical data, we have

$$\begin{aligned} \mathbb{E}_D(X^{\tilde{u}_s}(T) - X^{u_s}(T))^2 &= (z - x_0(1 + r_f T))^2 \left[\left(\frac{\beta'_n \tilde{\eta}}{\hat{\beta}'_n \tilde{\eta}} - 1 \right)^2 + \frac{1}{T} \left(\frac{\tilde{\eta}' \Sigma \tilde{\eta}}{(\hat{\beta}'_n \tilde{\eta})^2} - \frac{2}{\hat{\beta}'_n \tilde{\eta}} + \frac{1}{\Theta_p} \right) \right] \\ &= (z - x_0(1 + r_f T))^2 \left[\left(\frac{\beta'_n \tilde{\eta}}{\hat{\beta}'_n \tilde{\eta}} - 1 \right)^2 + \frac{1}{T} \left(\frac{\tilde{\eta}' \Sigma \tilde{\eta} - \hat{\beta}'_n \tilde{\eta}}{(\hat{\beta}'_n \tilde{\eta})^2} + \frac{\hat{\beta}'_n \tilde{\eta} - \Theta_p}{\hat{\beta}'_n \tilde{\eta} \Theta_p} \right) \right] \\ &\leq \frac{C}{(K - o_P(1))^2} \left[(\beta'_n \tilde{\eta} - \hat{\beta}'_n \tilde{\eta})^2 + |\tilde{\eta}' \Sigma \tilde{\eta} - \hat{\beta}'_n \tilde{\eta}| + |\hat{\beta}'_n \tilde{\eta} - \Theta_p| \right] \\ &\leq \frac{C}{(K - o_P(1))^2} d_n \rightarrow 0 \text{ in probability,} \end{aligned}$$

where (B.9) and (B.10) are used. Using (B.3), we have

$$\begin{aligned} |\mathbb{E}[X^{\tilde{u}_s}(T)] - z| &= (z - x_0(1 + r_f T)) \left| \frac{\beta'_n \tilde{\eta}}{\hat{\beta}'_n \tilde{\eta}} - 1 \right| \leq \frac{C}{K - o_P(1)} |(\beta - \hat{\beta}_n)' \tilde{\eta}| \\ &\leq \frac{C}{K - o_P(1)} b_n |\eta|_1 \leq \frac{C}{K - o_P(1)} d_n \rightarrow 0 \text{ in probability.} \end{aligned}$$

Similarly, using (B.5), we have

$$\begin{aligned} |(\sigma_T^{\tilde{u}_s})^2 - (\sigma_T^{u_s})^2| &= \frac{(z - x_0(1 + r_f T))^2}{T} \left| \frac{\tilde{\eta}' \Sigma \tilde{\eta}}{(\hat{\beta}'_n \tilde{\eta})^2} - \frac{1}{\Theta_p} \right| \\ &\leq \frac{C}{(K - o_P(1))^2} \left(|\tilde{\eta}' \Sigma \tilde{\eta} - \hat{\beta}'_n \tilde{\eta}| + |\hat{\beta}'_n \tilde{\eta} - \Theta_p| \right) \\ &\leq \frac{C}{(K - o_P(1))^2} d_n \rightarrow 0 \text{ in probability.} \end{aligned}$$

2. For the multiperiod strategy given in (B.2), we have

$$\begin{aligned} &X^{\tilde{u}_m}(T) - X^{u_m}(T) \\ &= \left(z - x_0 \prod_{j=0}^{N-1} (1 + r_f \Delta_j) \right) \left[\left(1 - \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1} \right)^{-1} - \left(1 - \prod_{j=0}^{N-1} (1 + \Delta_j \Theta_p)^{-1} \right)^{-1} \right. \\ &\quad \left. - \prod_{j=0}^{N-1} \frac{1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta} - \sqrt{\Delta_j} \tilde{\eta}' \sigma \xi_j}{1 + \Delta_j \hat{\beta}'_n \tilde{\eta}} + \prod_{j=0}^{N-1} \frac{1 - \sqrt{\Delta_j} \eta' \sigma \xi_j}{1 + \Delta_j \Theta_p} \right] \end{aligned}$$

Taking conditional expectation with respect to the historical data, we have

$$\begin{aligned}
& \sqrt{\mathbb{E}_D(X^{\tilde{u}_m}(T) - X^{u_m}(T))^2} \\
& \leq C \left\{ \left| \left(1 - \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1} \right)^{-1} - \left(1 - \prod_{j=0}^{N-1} (1 + \Delta_j \Theta_p)^{-1} \right)^{-1} \right| \right. \\
& \quad \left. + \left[\mathbb{E}_D \left(\prod_{j=0}^{N-1} \frac{1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta} - \sqrt{\Delta_j} \tilde{\eta}' \sigma \xi_j}{1 + \Delta_j \hat{\beta}'_n \tilde{\eta}} + \prod_{j=0}^{N-1} \frac{1 - \sqrt{\Delta_j} \eta' \sigma \xi_j}{1 + \Delta_j \Theta_p} \right)^2 \right]^{\frac{1}{2}} \right\} \\
& =: C(|I_1| + (I_2)^{\frac{1}{2}}).
\end{aligned}$$

Using (B.12), we have

$$\begin{aligned}
|I_1| & \leq \left(1 - \prod_{j=0}^{N-1} (1 + \Delta_j (K - o_P(1)))^{-1} \right)^{-2} \left| \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1} - \prod_{j=0}^{N-1} (1 + \Delta_j \Theta_p)^{-1} \right| \\
& \leq C \mathcal{O}_P(1) \sum_{j=0}^{N-1} \left| (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1} - (1 + \Delta_j \Theta_p)^{-1} \right| \\
& \leq C \mathcal{O}_P(1) \sum_{j=0}^{N-1} (1 + \Delta_j (K - o_P(1)))^{-2} \Delta_j |\hat{\beta}'_n \tilde{\eta} - \Theta_p| \\
& \leq C \mathcal{O}_P(1) (\lambda_n + b_n) |\eta|_1 \rightarrow 0 \text{ in probability,}
\end{aligned}$$

where we use a simple lemma: if complex numbers $z_1, \dots, z_n$ and $w_1, \dots, w_n$ are of modulus $\leq \theta$, then $|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i| \leq \theta^{n-1} \sum_{i=1}^n |z_i - w_i|$, in which $\theta = 1$ for our application. However, a simple calculation yields

$$I_2 = \prod_{j=0}^{N-1} \frac{(1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})^2 + \Delta_j \tilde{\eta}' \Sigma \tilde{\eta}}{(1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2} - \prod_{j=0}^{N-1} \frac{1}{1 + \Delta_j \Theta_p}.$$

Notice that

$$\left| \frac{(1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})^2 + \Delta_j \tilde{\eta}' \Sigma \tilde{\eta}}{(1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2} \right| \leq \frac{1 + \Delta_j \hat{\beta}_n \tilde{\eta} + o_P(1)}{(1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2} \leq \frac{1}{1 + \Delta_j \hat{\beta}'_n \tilde{\eta}} + o_P(1) \leq 1.$$

Hence, we have

$$\begin{aligned}
I_2 & \leq \sum_{j=0}^{N-1} \left| \frac{(1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})^2 + \Delta_j \tilde{\eta}' \Sigma \tilde{\eta}}{(1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2} - \frac{1}{1 + \Delta_j \Theta_p} \right| \\
& \leq \sum_{j=0}^{N-1} \left[\left| \frac{(1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})^2 - 1 + \Delta_j (\tilde{\eta}' \Sigma \tilde{\eta} - \hat{\beta}'_n \tilde{\eta})}{(1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2} \right| + \left| \frac{1}{1 + \Delta_j \hat{\beta}'_n \tilde{\eta}} - \frac{1}{1 + \Delta_j \Theta_p} \right| \right] \\
& \leq C \mathcal{O}_P(1) d_n \rightarrow 0 \text{ in probability.}
\end{aligned}$$

Therefore, $\mathbb{E}_D(X^{\tilde{u}_m}(T) - X^{u_m}(T))^2 \xrightarrow{\mathbb{P}} 0$. Using (B.4), we have

$$\begin{aligned} |\mathbb{E}[X^{\tilde{u}_m}(T)] - z| &= \frac{z - x_0 \prod_{k=0}^{N-1} (1 + r_f \Delta_k)}{\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta}) - 1} \left| 1 - \prod_{j=0}^{N-1} (1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta}) \right| \\ &\leq \frac{C}{T \hat{\beta}'_n \tilde{\eta}} \sum_{j=0}^{N-1} |\Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta}| \leq \frac{C}{K - o_P(1)} b_n |\eta|_1 \rightarrow 0 \text{ in probability.} \end{aligned}$$

Similarly, using (B.6), we have

$$\begin{aligned} &|(\sigma_T^{\tilde{u}_m})^2 - (\sigma_T^{u_m})^2| \\ &\leq C \left[\left| \frac{1}{\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta}) - 1} - \frac{1}{\prod_{j=0}^{N-1} (1 + \Delta_j \Theta_p) - 1} \right| + \right. \\ &\quad \left| \frac{\prod_{j=0}^{N-1} [(1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})^2 + \Delta_j \tilde{\eta}' \Sigma \tilde{\eta}] - \prod_{j=0}^{N-1} (1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})}{(\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta}) - 1)^2} - \frac{1}{\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta}) - 1} \right| \\ &=: C[|I_3| + |I_4|]. \end{aligned}$$

Notice that

$$\begin{aligned} |I_3| &\leq \left| \frac{\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1}}{1 - \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1}} - \frac{\prod_{j=0}^{N-1} (1 + \Delta_j \Theta_p)^{-1}}{1 - \prod_{j=0}^{N-1} (1 + \Delta_j \Theta_p)^{-1}} \right| \\ &\leq C \mathcal{O}_P(1) \left| \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1} - \prod_{j=0}^{N-1} (1 + \Delta_j \Theta_p)^{-1} \right| \\ &\leq C \mathcal{O}_P(1) \sum_{j=0}^{N-1} \left| (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1} - (1 + \Delta_j \Theta_p)^{-1} \right| \\ &\leq C \mathcal{O}_P(1) (\lambda_n + b_n) |\eta|_1 \rightarrow 0 \text{ in probability.} \end{aligned}$$

However,

$$\begin{aligned} &|I_4| \\ &= \left| \frac{\prod_{j=0}^{N-1} [(1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})^2 + \Delta_j \tilde{\eta}' \Sigma \tilde{\eta}] - \prod_{j=0}^{N-1} (1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta}) - (\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta}) - 1)}{\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2 (1 - \prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^{-1})^2} \right| \\ &\leq C \mathcal{O}_P(1) \left| \prod_{j=0}^{N-1} \frac{(1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})^2 + \Delta_j \tilde{\eta}' \Sigma \tilde{\eta}}{(1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2} - \prod_{j=0}^{N-1} \frac{1}{1 + \Delta_j \hat{\beta}'_n \tilde{\eta}} + \frac{1 - \prod_{j=0}^{N-1} (1 - \Delta_j (\beta - \hat{\beta}_n)' \tilde{\eta})}{\prod_{j=0}^{N-1} (1 + \Delta_j \hat{\beta}'_n \tilde{\eta})^2} \right| \\ &\leq C \mathcal{O}_P(1) (d_n + \mathcal{O}_P(1) b_n |\eta|_1) \rightarrow 0 \text{ in probability.} \end{aligned}$$

Therefore, $(\sigma_T^{\tilde{u}_m})^2 - (\sigma_T^{u_m})^2 \xrightarrow{\mathbb{P}} 0$.

□

Proof of Theorem 4.2. We can similarly deduce that

$$|\tilde{\eta}'\Sigma\tilde{\eta} - \Theta_p| \leq Cd_n \text{ and } \tilde{\eta}'\Sigma\tilde{\eta} \geq \Theta_p + o_P(1)$$

by (B.8) and (B.11). Then we have

$$\begin{aligned} \left| \frac{\beta'\tilde{\eta}}{\sqrt{\tilde{\eta}'\Sigma\tilde{\eta}}} - \sqrt{\Theta_p} \right| &\leq \left| \frac{\beta'\tilde{\eta} - \tilde{\eta}'\Sigma\tilde{\eta}}{\sqrt{\tilde{\eta}'\Sigma\tilde{\eta}}} \right| + \left| \sqrt{\tilde{\eta}'\Sigma\tilde{\eta}} - \sqrt{\Theta_p} \right| \\ &\leq \frac{d_n}{\sqrt{\Theta_p + o_P(1)}} + \frac{d_n}{\sqrt{\Theta_p + o_P(1)} + \sqrt{\Theta_p}} \leq C\mathcal{O}_P(1) \frac{d_n}{\sqrt{\Theta_p}}. \end{aligned}$$

With the condition (D), we have

$$\left| \frac{F_n}{F} - 1 \right| \leq C \frac{d_n}{\sqrt{\Theta_p}} (\sqrt{\Theta_p} + 1) e^{Cd_n} \leq Cd_n e^{Cd_n}.$$

□

Appendix C. Simulation results for log-normally distributed returns

In this section, we consider the same setting of the simulation study in Section 5.1, except that the returns have a log-normal distribution. Because this study is for illustrative propose, we consider p only up to 800 and conduct 100 simulations for every setting. Because the results and interpretations are similar to what we had in Section 5.1, we present only the simulation study for the covariance matrix of Model 2 to save space.

Figure C.7: Empirical frequency (%) of beating the bank deposit.

Figure C.8: Empirical frequencies (%) of (three columns on the left for each p) the multiperiod strategy beating the single-period strategies; and (two columns on the right for each p) the LPO strategy beating the plug-in strategy.

Figure C.9: Sharpe ratio of the simulation studies with log-normally distributed returns.

Table C.4: Averaged squared errors of the terminal wealths of the estimated strategies.

p	[S] LPO	[S] Plug-in	[M] LPO	[M] Plug-in
100	0.040	0.144	0.073	9.875
200	0.040	0.135	0.068	0.471
400	0.009	0.044	0.009	0.036
800	0.011	0.064	0.010	0.066

From the figures and the table above, it is easy to draw the same conclusions as in Section 5.1. The LPO strategy is better than the plug-in strategy in terms of the probability of outperforming the bank deposit, the Sharpe ratio, and the distance to the oracle. Figure C.7 also shows that the probability of outperforming the

bank deposit decreases when $p > n$, which verifies Theorem 3.1. Moreover, Table C.4 verifies Theorem 4.1.

Appendix D. Empirical results for the Russell 2000 dataset

We extend our analysis to a different, larger dataset. In the data preprocessing procedure, we discard Russell 2000 index components whose volatility is abnormal (i.e., the standard deviation of the asset returns is greater than 0.4). This threshold for volatility is chosen as the maximum value of realized volatilities of the S&P 500 components. Finally, there are 1420 stocks ($p = 1420$).

Table D.5 displays the empirical result for the Russell 2000 components. When the plug-in approach is implemented, the generalized inverses are often computationally infeasible for a large p , such as the value in this dataset. We therefore discard the plug-in result to avoid misleading results. If we consider the investment of all assets in an economy with no prior knowledge of favorable assets, the traditional plug-in approach usually halts and malfunctions during implementation. The LPO approach still works.

Table D.5: Empirical comparisons of the competitive strategies for Russell 2000 components in terms of numerous measures.

Strategy	EW	LPO	LPO-EW	AdLPO	Shrink
$\widehat{SR}$ (e-02)	2.695	5.074	9.028	16.781	12.194
(p -value)	–	(0.295)	(0.009)	(0.003)	(0.100)
$\widehat{Var}$ (e-05)	10.821	31.083	13.424	7.982	1.678
$\widehat{CEQ} _{\gamma=1}$ (e-04)	2.263	7.391	9.788	14.593	4.912
$\widehat{CEQ} _{\gamma=2}$ (e-04)	1.722	5.837	9.117	14.194	4.828
$\widehat{CEQ} _{\gamma=5}$ (e-04)	0.986	1.175	7.104	12.997	4.576
$\widehat{Turnover}$	0.011	0.445	0.017	2.226	0.225

The implications of the analysis of the Russell 2000 component data are similar to those discussed in the S&P 500 component data. The EW strategy further deteriorates to produce low Sharpe ratios and certainty equivalent values for a large p . The LPO approach significantly outperforms the EW. The out-of-sample variance of the EW is smaller than that of the LPO. It is expected because when p increases, all data-driven methods will increase their variance due to the accumulation of estimation error, and there are inactive stocks in this dataset that will not be picked by the LPO strategy. The shrinkage approach is better than the

LPO in terms of the Sharpe ratio but worse in terms of the certainty equivalent. LPO-EW is a significant improvement over LPO, and the adaptive LPO is the best of the strategies for this dataset. The turnover of LPO-EW is minimal and very similar to that of the EW approach.