Correcting In-Sample Optimism Bias: Realized Volatility of Large Optimal Portfolios*

Jianqing Fan† Princeton University

Alex Furger‡ Princeton University

Dacheng Xiu§ University of Chicago

This is a Preliminary Version: January 19, 2015

Abstract

Using high-frequency data, we estimate the risk of a large portfolio with weights being the solution of an optimization problem subject to some linear inequality constraints. We propose a fully nonparametric approach as a benchmark, as well as a factor-based semiparametric approach with observable factors to attack the curse of dimensionality. We provide in-fill asymptotic distributions of the realized volatility estimators of the optimal portfolio, while taking into account the estimation error in the optimal portfolio weights as a result of the covariance matrix estimation. Our theoretical findings suggest that ignoring such an error leads to a first-order asymptotic bias which undermines the statistical inference. Such a bias is related to in-sample optimism in portfolio allocation. Our simulation results suggest satisfactory finite sample performance after bias correction, and that the factor-based approach becomes increasingly superior with a growing cross-sectional dimension. Empirically, using a large cross-section of high-frequency stock returns, we find our estimator successfully addresses the issue of in-sample optimism.

Keywords: quadratic programing, in-sample optimism, exposure constraint, big data.

1 Introduction

Optimal portfolio allocation and its risk assessment have taking an important role in the modern financial industry, since the seminal work by Markowitz (1952). As the universe of financial instruments for trading and hedging is expanding and the trading speed and frequency are rising, there

*We thank Christian Hansen and Jean Jacod for their helpful comment. Please find the most recent version on the web: http://faculty.chicagobooth.edu/dacheng.xiu/research/RVOP.pdf.

†Department of Operations Research & Financial Engineering, Princeton University. Address: 26 Prospect Avenue, Princeton, NJ 08544, USA. Email address: jqfan@princeton.edu.

‡Department of Operations Research & Financial Engineering, Princeton University. Address: 26 Prospect Avenue, Princeton, NJ 08544, USA. E-mail address: afurger@princeton.edu.

§Booth School of Business, University of Chicago. Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: dacheng.xiu@chicagobooth.edu.

is an urgent demand for statistical tools that could address the increasing difficulty in portfolio allocation.

The covariance matrix is a key input to this problem, hence its estimation has attracted much attention recently. It is however by no means a solved issue, as the curse of dimensionality is a tremendous hurdle to its statistical modeling and inference. The fundamental obstacle is the lack of a sufficient number of observations over time, which are required to accommodate the dimensionality of the universe of assets.

The large amount of intraday data sprung from high-frequency trading meet the data requirements for statistical inference. On a typical 6.5-hour trading day, the number of transactions for a liquid stock exceeds 10,000, providing up-to-date and accurate information about the stock's own fluctuation and its co-movement with others. Motivated by this opportunity, we conduct econometric analysis of the risk of a large portfolio with intraday data.

The use of high-frequency data not only enlarges the effective number of observations, it facilities the asymptotic inference to the extent that asymptotic distributions with a fixed cross-sectional dimension could provide a very decent characterization of the finite sample performance of an estimator. Therefore, the usual asymptotic inference can be carried out. This is in sharp contrast to the typical high-dimensional asymptotic analysis, which yields at best an optimal convergence rate rather than a central limit result. In addition, the in-fill asymptotic technique is often used to study high-frequency data, allowing general data generating processes, such as Itô-semimartingales.

A particular aspect of our study that differs from the existing literature is the focus on understanding the impact of the first-step covariance estimation, an input to the portfolio optimization, on the optimal allocations. The estimation error thereby influences the portfolio weights, and it is then propagated into the uncertainty of the risk of the optimal portfolio. Surprisingly, there is hardly any work in the literature that addresses this question in theory, despite well-designed out-of-sample tests which may take this into account in the empirical comparison. We target this issue with a fairly general semimartingale model. We find that by coincidence the first-step estimation does not affect the asymptotic variance of the realized volatility of a portfolio, but instead imposes a first-order asymptotic bias. Without correcting the bias, the central limit result breaks down.

We propose a nonparametric estimator based on the sample covariance and a factor-model-based semiparametric estimator to estimate the risk of the portfolio. Once the biases are corrected, we find that the latter estimator is more efficient in theory and performs better in finite sample studies, in particular when the portfolio weights are unknown and the number of assets under management is large. The reason for that, as revealed by our study, is the superior performance of the estimated inverse covariance matrix when an additional factor structure is imposed. While there is a clear trade-off between the efficiency gain and robustness loss to model misspecification risk, in our view the factor-based approach is an ideal alternative that is worth implementing in practice.

A classical pitfall of using optimized portfolios to forecast future portfolio volatility is *in-sample optimism*. This is the tendency for the in-sample volatility of an optimized portfolio to largely underestimate the population volatility such a portfolio would have experienced, even in-sample.

Several classical references demonstrate this empirically, including Jobson and Korkie (1981), Frost and Savarino (1988), and Michaud (1989). More recently, Basak, Jagannathan, and Ma (2009) propose an exact correction to produce an unbiased estimate of out-of-sample volatility based on Gaussian assumptions, and also describe a jackknife estimator which corrects the in-sample optimism in their setting. We point out that such a bias arises naturally out of the first-stage covariance matrix estimation in our general continuous-time setting, to the extent that it dominates the asymptotic distribution of the estimator.

A growing number of methods have been proposed to overcome the curse of dimensionality inherent to the covariance estimation in the presence of a large universe of assets. Two approaches stand out. The first class of estimators are shrinkage estimators, Ledoit and Wolf (2004a), Ledoit and Wolf (2004b), Ledoit and Wolf (2012), which effectively shrink the eigenvalues of the sample covariance matrix to towards some fixed target. The other class of estimators use explicit factor models on the data generating process to convert the high-dimensional covariance matrix into one which is low-rank plus sparse. These factor models may use either observed factors, Fan, Fan, and Lv (2008), Fan, Liao, and Mincheva (2011) or unobserved factors, Fan, Liao, and Mincheva (2013). All of the above approaches have been shown to improve the quality of optimized portfolios over the sample covariance estimator. In the extreme, but common, case, when $p \gg n$, these methods also make the portfolio selection feasible. However, no limiting distribution is provided for the risk of the portfolio, as is typical to high-dimensional methods. An exception is Fan, Liao, and Shi (2013), which derives an upper confidence bound on the volatility of the portfolio, while explicitly allowing dimension to grow with the number of observations. However, their portfolio weights are deterministic. By contrast, we adopt high-frequency data to cope with the lack of data in terms of time series observations, to the extent that our in-fill asymptotic approximation delivers a useful limiting distribution for inference, even in the case with estimated portfolio weights.

Several authors have suggested that instead of, or in concert with, reducing the intrinsic dimensionality of the covariance, a modification can be made to the portfolio allocation optimization itself. A prominent method that is suggested in the literature is the so-called *exposure constraint*, see e.g. Jagannathan and Ma (2003) and Fan, Zhang, and Yu (2012). The exposure constraint formulation elicits superior statistical performance, even when the population optimal portfolio may not satisfy the given exposure constraint. We impose more general constraints on the portfolio optimization, which nest the exposure constraint as a special case.

Other authors have studied the use of intraday data for large covariance matrix estimation, see, e.g., Wang and Zou (2010), Fan, Li, and Yu (2012), Tao, Wang, and Zhou (2013), Hautsch, Kyj, and Malec (2013), and Lunde, Shephard, and Sheppard (2014). These papers focus on the sample covariance based estimator, or a noise robust version of it, whereas Fan, Furger, and Xiu (2014) propose an approximate-factor-model-based estimator, which takes advantage of the Global Industrial Classification Standard (GICS) to form blocks of stocks and then threshold the off-block-diagonal residual covariances. Their asymptotic analysis is based on the marriage of infill and diverging dimension asymptotics. All these papers ignore the impact of the first-stage estimation error on the portfolio weight, which then propagates into the evaluation of the portfolio risk. We fill in this gap by providing a feasible asymptotic theory that keeps track of this error, which turns to a bias that is related to the in-sample optimism. We also provide a factor-model based approach, which has superior performance over the sample covariance based method when as many as 100 assets are considered for allocation.

Our paper is closely related to a growing amount of literature on the spot-covariance-based estimation with high-frequency data. Jacod and Rosenbaum (2013) propose to estimate a general integrated function of spot-covariance by aggregating an increasing number of local estimates. They apply this strategy to integrated quarticity estimation. Mykland and Zhang (2009) propose a similar idea, although their asymptotic theory is based on a finite number of local windows. Aït-Sahalia and Xiu (2014) extend the principal component analysis to high-frequency data, and advocate the use of intraday data to attack the curse of dimensionality, which is similar in spirit to our work. Aït-Sahalia, Kalnina, and Xiu (2014) develop a continuous-time factor model to estimate the idiosyncratic volatility. They also build the Fama-French factors at 5-minute intervals, which we use in the empirical work. See also Li and Xiu (2013), Kalnina and Xiu (2014), and Li, Todorov, and Tauchen (2013) for other applications. Earlier work on spot-variance estimation include Foster and Nelson (1996) and Comte and Renault (1998) in a setting without jumps. Also see Renò (2008), Kristensen (2010), and references therein.

The structure of the rest of the paper is as follows. Section 2 sets up the model and provides the assumptions. Section 3 discusses a general constrained optimization problem and its solutions, which are critical for the econometric analysis, detailed in Section 4. Section 5 verifies our theory using Monte Carlo simulations. Section 6 includes an empirical study for illustration.

In terms of notations, we use $\|\cdot\|$ and $\|\cdot\|_1$ to denote the Euclidean norm and $\mathbb{L}_1$ norm, respectively, for vectors or matrices, and $|\cdot|_\infty$ for the $\mathbb{L}_\infty$ norm of a vector. $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is a filtered probability space, and $\mathcal{M}^{++}_{d\times d}$ is the set of positive-definite matrices.

2 Model Setup and Assumptions

Imagine an investor instantaneously rebalances her portfolio within a fixed window $[0, t]$. The realized risk of her portfolio is given by:

$$\mathrm{RV}_t = \int_0^t (\omega^*_{s-})^\intercal c_s \omega^*_{s-} ds,$$

where ω^* is her trading strategy,[1] and c is the instantaneous covariance matrix of asset returns under consideration. RV_t is the target of this paper.

To analyze the portfolio risk, we need trading strategies. In this paper, we consider two general class of strategies. The first strategy involves a general minimum-risk portfolio investor, who

[1]Since at each time point s, she would invest according to the allocation ω^* based on the information up to $s-$, hence we use ω^*_{s-}.

considers a constrained portfolio optimization problem prior to each time point s:

$$\min_{w} w^{\intercal}\mathbb{E}(c_s|\mathcal{F}_{s-})w, \quad \text{subject to } Aw \le b, \tag{1}$$

where $\mathcal{F}_{s-}$ is the filtration that summarizes information up to $s-$, and ω is the portfolio weight to be determined. A and b together define a general constraint the portfolio weight needs to satisfy. The solution to the optimization problem is written as ω^*, which is unique if $\mathbb{E}(c_s|\mathcal{F}_{s-})$ is positive-definite.

The second class of strategies is implemented with a deterministic ω_s, for $0 \le s \le t$. For instance, the popular equal-weighted portfolio is a special case of this class. For strategies withIn this class, we write $\omega_s^* = \omega_s$. Needless to say, these strategies are more passive as opposed to the first class.

Apparently, the role of the (instantaneous) covariance matrix is two-fold in this problem. On the one hand, for the optimal portfolio, the (expected) covariance, or more precisely, its inverse is the key input to the portfolio weight. On the other hand, once the portfolio weight is settled, whether obtained from optimization or decided in priori, the covariance is required again to evaluate the risk of a portfolio. We turn to modeling the covariance first.

We start with a nonparametric model – only the usual assumptions from the literature on the underlying processes are made, which are essentially satisfied by most asset pricing models in finance.[2]

Assumption 1. *Suppose the log asset prices, summarized in Y, follow a general d-dimensional Itô semimartingale, i.e.,*

$$Y_t = Y_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s dW_s + J_t, \quad \textit{where} \quad J_t = (\delta 1_{\{|\delta|_\infty \le 1\}}) * (\mu - \nu)_t + (\delta 1_{\{|\delta|_\infty > 1\}}) * \mu_t, \tag{2}$$

and its spot covariance, denoted as $c_t = \sigma_t \sigma_t^{\intercal}$, follows

$$vech(c_t) = vech(c_0) + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s d\tilde{W}_s + \tilde{J}_t, \quad \textit{where} \quad \tilde{J}_t = (\tilde{\delta} 1_{\{|\tilde{\delta}|_\infty \le 1\}}) * (\mu - \nu)_t + (\tilde{\delta} 1_{\{|\tilde{\delta}|_\infty > 1\}}) * \mu_t, \tag{3}$$

where W_t and $\tilde{W}_t$ are standard but potentially correlated Brownian motions. Moreover, for any $\gamma \in [0,1)$, there is a sequence of stopping times $\{\tau_m\}_{m \ge 1}$ increasing to ∞, and deterministic function γ_m such that $\int_{\mathbb{R}^d} \gamma_m(z)^\gamma \lambda(dz) < \infty$ and that $|\delta(\omega,t,z)|_\infty \wedge 1 \le \gamma_m(z)$ and $|\tilde{\delta}(\omega,t,z)|_\infty \wedge 1 \le \gamma_m(z)$ for all (ω,t,z) with $t \le \tau_m(\omega)$. In addition, b_t and $\tilde{b}_t$ are locally bounded and progressively measurable, and $\tilde{\sigma}_t$ is càdlàg. Finally, the process c_t is an Itô semimartingale. c_{t-} and c_t are positive-definite.

To deal with the curse of dimensionality, and as an alternative to the previous approach, we consider a continuous-time semiparametric approximate factor model:

$$Y_t = Y_0 + \int_0^t \beta_s dX_s^c + \sum_{s \le t} \overline{\beta}_s \Delta X_s + Z_t, \tag{4}$$

where Y is a d-dimensional vector process, X is a r-dimensional observable factor process with X^c being its continuous component and ΔX_s the jump size of X at s, Z is the idiosyncratic component,

[2]Models with factional Brownian motion components are unfortunately excluded.

and β_t and $\overline{\beta}_t$ are time-varying factor loading matrices of size $d \times r$ for the continuous and jump components, respectively. This model, though cast in continuous time, is motivated from the discrete-time factor model widely used in macroeconomics and empirical finance, see e.g. Ross (1976), Chamberlain and Rothschild (1983), Fama and French (1993), and Stock and Watson (2002). The continuous-time model is a natural choice for modeling high-frequency data, particularly for intraday transaction prices. In contrast with a standard factor model, the usual "alpha term", i.e. the excess return, is absorbed into the residual Z_t for convenience, as it does not play a role for risk measurement.

Assumption 2. *Suppose the vector of log asset prices Y follows the continuous-time factor model given by (4), in which X is a general Itô semimartingale $X_t = X_0 + X_t^c + X_t^d$, where*

$$X_t^c = \int_0^t h_s \, ds + \int_0^t \eta_s dB_s, \quad \text{and} \quad X_t^d = (\varepsilon 1_{\{|\varepsilon|_\infty \le 1\}}) * (\zeta - \xi)_t + (\varepsilon 1_{\{|\varepsilon|_\infty > 1\}}) * \zeta_t.$$

Its spot covariance, denoted as $e_t = \eta_t \eta_t^\intercal$, follows

$$e_t = e_0 + \int_0^t \tilde{h}_s ds + \int_0^t \tilde{\eta}_s d\tilde{B}_s + (\tilde{\varepsilon} 1_{\{|\tilde{\varepsilon}|_\infty \le 1\}}) * (\zeta - \xi)_t + (\tilde{\varepsilon} 1_{\{|\tilde{\varepsilon}|_\infty > 1\}}) * \zeta_t. \tag{5}$$

In addition, Z_t is another Itô semimartingale satisfying $Z_t = Z_0 + Z_t^c + Z_t^d$, where

$$Z_t^c = \int_0^t \bar{h}_s ds + \int_0^t \gamma_s d\bar{B}_s, \quad \text{and} \quad Z_t^d = (\bar{\varepsilon} 1_{\{|\bar{\varepsilon}|_\infty \le 1\}}) * (\zeta - \xi)_t + (\bar{\varepsilon} 1_{\{|\bar{\varepsilon}|_\infty > 1\}}) * \zeta_t.$$

B_t, $\tilde{B}_t$, and $\bar{B}_t$ are potentially correlated Brownian motions. Moreover, for any $\gamma \in [0, 1)$, there is a sequence of stopping times (τ_n) increasing to ∞, and deterministic function γ_n such that $\int_{\mathbb{R}^d} \gamma_n(z)^\gamma \lambda(dz) < \infty$ and that $|\varepsilon(\omega, t, z)|_\infty \wedge 1 \le \gamma_n(z)$, $|\tilde{\varepsilon}(\omega, t, z)|_\infty \wedge 1 \le \gamma_n(z)$, and $|\bar{\varepsilon}(\omega, t, z)|_\infty \wedge 1 \le \gamma_n(z)$ for all (ω, t, z) with $t \le \tau_n(\omega)$. In addition, h_t, $\tilde{h}_t$, and $\bar{h}_t$ are locally bounded and progressively measurable, and $\tilde{\eta}_t$ is càdlàg. Finally, the process η_t, β_t, γ_t are Itô semimartingales, and e_{t-} and e_t are positive-definite, and $\gamma_t \gamma_t^\intercal$ is a positive-definite block-diagonal matrix.

In addition, similar to the standard OLS regression, we need to introduce the exogeneity condition for the purpose of identification:

Assumption 3. *$[Z_s^c, \int_0^s \beta_h dX_h^c] = 0$, and $[Z_s^d, \sum_{h \le s} \overline{\beta}_h \Delta X_h] = 0$, for any $0 \le s \le t$, where $[\cdot, \cdot]$ denotes the quadratic covariation.*

Remark 1. The above Assumptions 2 and 3 impose the following structure on the spot covariance matrix of Y:

$$c_s = \beta_s e_s \beta_s^\intercal + \gamma_s \gamma_s^\intercal, \quad 0 \le s \le t,$$

where $\gamma_s \gamma_s^\intercal$ is a block diagonal matrix. We use $\mathcal{S}$ to denote the set of entries of $\gamma_s \gamma_s^\intercal$ within the blocks on the diagonal. Therefore, for any $(i, j) \notin \mathcal{S}$, $(\gamma_s \gamma_s^\intercal)_{i,j} = 0$.

When $\gamma\gamma^\intercal$ is a block diagonal matrix, we can this model an *approximate* factor model, as opposed to a *strict* factor model, the case when $\gamma\gamma^\intercal$ is a diagonal matrix. The block-diagonal covariance matrix is motivated from a recent paper by Fan, Furger, and Xiu (2014), who document a block-diagonal structure of the residual covariance matrix using a variety of factor models. The blocks are formed by sorting the GICS codes of stocks.

Remark 2. Under either Assumption 1 or 2, we have

$$\mathbb{E}(c_s|\mathcal{F}_{s-}) = c_{s-},$$

for any non-random $s \in [0, t]$, because c_t is an Itô semimartingale. This simplifies our optimization problem to:

$$f(c_{s-}, A, b) := \min_w w^\intercal c_{s-} w, \quad \text{subject to } Aw \leq b, \tag{6}$$

so that ω^* is a function of A, b, and c_{s-}. We write it as ω^*_{s-}.

Remark 3. The realized volatility of the optimal portfolio is $\mathrm{RV}_t = \int_0^t (\omega^*_{s-})^\intercal c_s \omega^*_{s-}$, which is equal to $\int_0^t \omega_s^{*\intercal} c_s \omega^*_s ds$, due to absolute continuity of the Lebesgue Integral. Since ω^* is the solution to (6), the portfolio risk RV_t equals

$$\mathrm{RV}_t(A, b) = \int_0^t f(c_s, A, b) ds, \tag{7}$$

where we also highlight its dependence on the constraints defined by A and b.

Very frequent rebalancing may increase the transaction cost dramatically, which may not be feasible in practice. Therefore, it is perhaps more relevant to consider the risk of a portfolio in discrete-time. Below we point out that the discretization error associated with the realized volatility measure defined in continuos-time is negligible relative to the statistical estimation error to be discussed later, as long as the portfolio rebalance frequency is of certain range.

Theorem 1. *Suppose a portfolio is rebalanced in discrete-time at δ_n intervals. For a fixed time period $[0, t]$, the realized risk of this portfolio satisfies*

$$\sum_{l=1}^{[t/\delta_n]} {\omega^*}^\intercal_{(l-1)\delta_n} \Big(\int_{(l-1)\delta_n}^{l\delta_n} c_s ds \Big) {\omega^*}_{(l-1)\delta_n} - \int_0^t \omega_s^{*\intercal} c_s \omega^*_s ds = O_p(\delta_n),$$

as $\delta_n \to 0$, where the portfolio strategy ω^*_t and the spot covariance c_t are Itô semimartingales.

Remark 4. Suppose that the data is sampled at a frequency Δ_n, and that the rebalancing window is $m_n \Delta_n$. Theorem 1 implies that as long as $m_n = o(\Delta_n^{-1/2})$, the approximation error due to continuous rebalancing is of order $O_p(m_n\Delta_n) = o_p(\Delta_n^{1/2})$, which is dominated by the statistical estimation error $O_p(\Delta_n^{1/2})$ shown below, so that our target RV_t based on continuous rebalancing remains relevant and valid. Adopting it facilities our statistical inference.

3 Constrained Portfolio Optimization

Prior to the econometric analysis, we investigate the optimization problem (6), which is essential for the subsequent analysis. In particular, we need to investigate the sensitivity of the function $f(c, A, b)$ with respect to its matrix input c, i.e. the smoothness of $f(c, A, b)$ as a function of c. We will use results developed in Shapiro (1985) to establish closed form results for the derivatives of (6).

Before we begin, it should be noted that we are not in the same regime that Shapiro (1985) studied. Most obviously, they consider problems that involve equality constraints, whereas we consider general inequality constraints. The extension, therefore, will require we first partition our

Throughout, we will fix $c \in \mathcal{M}_{d\times d}^{++}$.

3.1 Smoothness of the Constrained QP

Prior to the econometric analysis, we investigate the optimization problem (6), which is essential for the subsequent analysis. In particular, we need to investigate the sensitivity of the function $f(c, A, b)$ with respect to its matrix input c, i.e. the smoothness of $f(c, A, b)$ as a function of c. In this section, we establish the existence of and provide closed forms for the first Hadamard derivatives of this function with respect to c.

We now need a bit of notation to state the main theorem of this section. Let ω^* and λ^* be the optimal portfolio and associated dual solution, where we hide the dependence on c in order to simplify the notation. Given a portfolio and its dual solution, one may partition the set of constraints into the following sets:

$$\begin{aligned}
\mathcal{A}(c) &= \{i : A_i\omega^* - b_i = 0\},\\
\mathcal{D}(c) &= \{i : \lambda_i^* = 0, A_i\omega^* - b_i = 0\},\\
\mathcal{I}(c) &= \{i : \lambda_i^* = 0\}.
\end{aligned}$$

These are commonly called the active, degenerate, and inactive constraint sets, respectively. This notation differs slightly from the usual notation, in that the sets are not mutually exclusive. Although, it is clear that $\mathcal{D}(c) = \mathcal{A}(c) \cap \mathcal{I}(c)$, we emphasize it with a separate notation for clarity. Finally, let $A_{\mathcal{A}}$ and $b_{\mathcal{A}}$ be the subset of rows that correspond to indices in the active set.

We will use the not In the following theorem, by the first-order derivative of $f(c)$ we mean:

$$\lim_{h\to 0} \frac{f(c + he_ie_j^\intercal) - f(c)}{h} = \partial_{ij} f(c).$$

And by second-order one-sided (directional) derivative we mean:

$$\begin{aligned}
\lim_{h\to 0^+} \frac{\partial_{ij} f(c + he_ke_\ell^\intercal) - \partial_{ij} f(c)}{h} &= \partial^{2,+}_{ij,k\ell} f(c),\\
\lim_{h\to 0^-} \frac{\partial_{ij} f(c + he_ke_\ell^\intercal) - \partial_{ij} f(c)}{h} &= \partial^{2,-}_{ij,k\ell} f(c),
\end{aligned}$$

which is defined in the sense of Gâteaux, see e.g. Shapiro (1990). If $\partial^{2,+}_{ij,k\ell} f(c) = \partial^{2,-}_{ij,k\ell} f(c)$, then the second-order derivative exists.

Theorem 2. *Assume c is positive definite, and that the domain of c is closed and connected, then we have*

i. The optimal objective value is given by:

$$f(c, A, b) = -(\lambda^*)^\intercal b, \tag{8}$$

where λ^* is the solution to the dual problem of (6).

ii. The first-order derivative of f with respect to entry c_{ij} is given by:

$$\partial_{ij} f(c, A, b) = E_{ij} = E_{ji}, \tag{9}$$

where

$$E = c^{-1} A^\intercal \lambda^* (\lambda^*)^\intercal A c^{-1} = \omega^* \omega^{*\intercal}. \tag{10}$$

iii. The second-order one-sided derivative with respect to c_{ij} and $c_{k\ell}$ direction is given by:

$$\partial^{2,+}_{ij,k\ell} f(c, A, b) = E_{jk}(F_{li} - c^{-1}_{li}) + E_{li}(F_{jk} - c^{-1}_{jk}), \tag{11}$$

where

$$F = c^{-1} A^\intercal_{\mathcal{A}} (A_{\mathcal{A}} c^{-1} A^\intercal_{\mathcal{A}})^{-1} A_{\mathcal{A}} c^{-1}. \tag{12}$$

If $\mathcal{D}(c) = \emptyset$, then the second-order derivative exists at c; the two one-sided limits exist and are equal.

Conversely, if $\mathcal{D}(c) \neq \emptyset$, then the second-order derivative may not exist at c; the two one-sided limits exist but may not be equal.

3.2 A Concrete Example

A central issue in the application of Theorem 2 is the calculation of the active set $\mathcal{A}(c)$. This quantity appears in the second-order derivatives and, as will we see, is critical for the econometric analysis. The actual calculation of the active set is one that must be implemented differently for each different optimizing allocation problem used. In our simulations and empirics we use a single model for calculating optimal portfolios, as suggested in Fan, Zhang, and Yu (2012):

$$\min_w w^\intercal c w, \quad \text{subject to } \omega^\intercal 1 = 1, \|\omega\|_1 \leq \gamma. \tag{13}$$

In this case, we can rewrite the $\mathbb{L}_1$ constraint into 2^d linear constraints in d-dimensions, namely $\tilde{A}\omega = \tilde{b}$, where $\tilde{A} \in \mathbb{R}_{2^d \times d}$ and $\tilde{b} \in \mathbb{R}_{2^d}$. One simple way to characterize A is in terms of a base expansion matrix which contains the numbers $1, 2, \cdots, d$ in their d-digit base 2 expansion. Let the rows of a matrix $B_{i,:} = i$ base 2. Then,

$$\tilde{A} = 2B - 1, \qquad \tilde{b}_i = \gamma.$$

Finally, combining the constraints from the equality constraint $\omega^\intercal 1 = 1$ gives us

$$A = \begin{pmatrix} & \tilde{A} & \\ 1 & \cdots & 1 \\ -1 & \cdots & -1 \end{pmatrix}, \qquad b = \begin{pmatrix} \tilde{b} \\ 1 \\ -1 \end{pmatrix}.$$

The theorem requires that the choice of $A_{\mathcal{A}}$ be of full-row rank. Admittedly, this can be a complicating aspect to the practical application of our methodology. However, in the above allocation problem, given just the optimized portfolio and the exposure constraint, we can easily calculate the full row-rank set of active constraints.

To begin, it is obvious that

$$\begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix}^\intercal \omega \leq -1$$

will always be an active constraint. This is simply because the allocation can always take less risk by allocating less. After this it becomes less obvious what constraints are active. When, $\gamma = \infty$ there are no other active constraints (there are no more constraints!). When, $\gamma < \infty$ there are possibly more active constraints. The additional constraints have the form of $\operatorname{sgn}(\omega)$, where $\operatorname{sgn}(\cdot)$ is the usual sign function with $\operatorname{sgn}(0) = 0$. For each zero in the portfolio there will be an active constraint. If we number the m zeros in the portfolio with $j = 1, \cdots, m$ then the j-th constraint vector is given by

$$a_{j,i} = \begin{cases} \operatorname{sgn} \omega_i & \text{if } \operatorname{sgn} \omega_i \neq 0 \\ -1 & \text{if } \operatorname{sgn} \omega_i = 0 \text{ and } i \text{ is the } j\text{-th zero.} \\ 1 & \text{otherwise} \end{cases} \tag{14}$$

In the case where the portfolio allocations are negative we need one more constraint,

$$\tilde{a}_i = \begin{cases} \operatorname{sgn} \omega_i & \text{if } \operatorname{sgn} \omega_i \neq 0 \\ 1 & \text{otherwise} \end{cases}. \tag{15}$$

The final active constraint matrix which has full row-rank and the corresponding b are given by:

$$A_{\mathcal{A}} = \begin{pmatrix} -1 \\ a_1 \\ \vdots \\ a_m \\ \tilde{a} \end{pmatrix} \quad \text{and} \quad b_{\mathcal{A}} = \begin{pmatrix} -1 \\ c \\ \vdots \\ c \\ c \end{pmatrix}.$$

It is worth noting that an active set optimization algorithm would also supply this matrix, *if it were applied exactly on (13)*. However, most implementations choose to solve the optimization in a higher dimensional space, in order to encode the $\mathbb{L}_1$-norm in a linear number of constraints. This transition to higher dimension causes the optimization to become a positive semi-definite quadratic programing, for which our theorem is not applicable. Specifically, it breaks down because the theorem relies on the invertibility of the matrix c. In the higher-dimensional transformed problem this matrix is no longer invertible.

4 Econometric Analysis

4.1 Portfolios with Optimal Weights

We now proceed to the inference. To fix ideas, denote the distance between adjacent observations by Δ_n. Let $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$, for $1 \le i \le [t/\Delta_n]$. To estimate the spot covariance of X, we form a non-overlapping window of k_n observations, so that the length of the interval covered by a window is $k_n\Delta_n$.

Within the i-th block, we estimate the spot covariance matrix with the truncated realized variance over the small window, that is,

$$\widehat{c}_{ik_n\Delta_n} = \frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\left(\Delta_{ik_n+j}^n Y\right)\left(\Delta_{ik_n+j}^n Y\right)^\intercal 1_{\left\{|\Delta_{ik_n+j}^n Y|_\infty \le u_n\right\}}. \tag{16}$$

The natural estimator of $\mathrm{RV}_t(A,b)$ is given below, which can be shown to be consistent:

$$\Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]} f(\widehat{c}_{ik_n\Delta_n}, A, b) \xrightarrow{\mathrm{P}} \mathrm{RV}_t(A,b).$$

However, there is some asymptotic bias associated with this naive estimator, as a result of the first-step estimation of $\widehat{c}_{ik_n\Delta_n}$. It can also be shown from the discussions in Jacod and Rosenbaum (2013) that the asymptotic bias hinders the statistical inference, in that as $k_n \to \infty$,

$$k_n\left(k_n\Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]} f(\widehat{c}_{ik_n\Delta_n}, A, b) - \mathrm{RV}_t(A,b)\right)$$
$$\xrightarrow{\mathrm{P}} \frac{1}{2}\int_0^t \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} f(c_s, A, b)\left(c_{jl,s}c_{km,s} + c_{jm,s}c_{kl,s}\right) ds.$$

Therefore, bias correction is necessary to allow a central limit result. The new estimator is given by:

$$\widehat{\mathrm{RV}}_t(A,b) = k_n\Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]} f(\widehat{c}_{ik_n\Delta_n}, A, b)\left(1 + \frac{1}{k_n}(d - d_{\mathcal{A}})\right),$$

where $d_{\mathcal{A}} = dim(A_{\mathcal{A}}c^{-1}A_{\mathcal{A}}^\intercal)$, and $A_{\mathcal{A}}$ is the subset of rows that correspond to indices in the active set $\mathcal{A}(c)$. The second term inside the summation is a correction to the aforementioned asymptotic bias, for which we give a new notation:

$$\text{Bias}(c; \text{RV}) = -\frac{1}{k_n}(d - d_{\mathcal{A}})(\omega^{*\intercal} c \omega^*).$$

The new estimator satisfies the following central limit theory:

Theorem 3. *Suppose Assumption 1 holds and that* $u_n \asymp \Delta_n^{\varpi}$*, with* $5/(12 - 2\gamma) \leq \varpi < 1/2$*. For any fixed* t*, and as* $\Delta_n \to 0$*,* $k_n \to \infty$ *with* $k_n^2 \Delta_n \to 0$ *and* $k_n^3 \Delta_n \to \infty$*, we have*

$$\frac{1}{\sqrt{\Delta_n}}\left(\widehat{\text{RV}}_t(A, b) - \text{RV}_t(A, b)\right) \xrightarrow{\text{L}-s} \mathcal{W}_t,$$

where $\mathcal{W}$ *is a standard normal random variable defined on the extension of the original probability space, and its covariance is given by*

$$\mathbb{E}(\mathcal{W}_t^2 | \mathcal{F}) = 2\int_0^t (\omega_s^{*\intercal} c_s \omega_s^*)^2 ds.$$

where ω_s^* *is the optimal portfolio allocation that solves (6) at time* s*.*

An alternative strategy is to take advantage of the factor structure in the data. Similar to the standard factor model with observable factors, we use OLS type of regressions to estimate the factor loadings. The assumed factor structure essentially only helps regulate the covariance matrix of the residual, by replacing the off-diagonal entries with zeros. We can write such an estimator in a compact form.

We stack Y and X together in $U = (Y^\intercal, X^\intercal)^\intercal$. Therefore, we have

$$C_s := \frac{[dU_s, dU_s]^c}{ds} = \begin{pmatrix} \beta_s e_s \beta_s^\intercal + \gamma_s \gamma_s^\intercal & \beta_s e_s \\ e_s \beta_s^\intercal & e_s \end{pmatrix} =: \begin{pmatrix} C_{11,s} & C_{12,s} \\ C_{21,s} & C_{22,s} \end{pmatrix}.$$

where $C_{ij,s}$ denotes the (i,j) block of C_s. It is clear that we can estimate $C_{ik_n\Delta_n}$ in the same manner as we estimate $c_{i\Delta_n}$:

$$\widehat{C}_{ik_n\Delta_n} = \frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\left(\Delta_{ik_n+j}^n U\right)\left(\Delta_{ik_n+j}^n U\right)^\intercal 1_{\{|\Delta_{ik_n+j}^n U|_\infty \leq u_n\}}. \tag{17}$$

Then, our factor-based estimator of $c_{ik_n\Delta_n}$ can be written as

$$\begin{aligned}\widehat{c}_{ik_n\Delta_n} =& g(\widehat{C}_{ik_n\Delta_n}) \\ :=& \mathcal{M}_{\mathcal{S}}\left(\widehat{C}_{11,ik_n\Delta_n} - \widehat{C}_{12,ik_n\Delta_n}\widehat{C}_{22,ik_n\Delta_n}^{-1}\widehat{C}_{21,ik_n\Delta_n}\right) + \widehat{C}_{12,ik_n\Delta_n}\widehat{C}_{22,ik_n\Delta_n}^{-1}\widehat{C}_{21,ik_n\Delta_n},\end{aligned} \tag{18}$$

where $\mathcal{M}_{\mathcal{S}}(M) = M_{ij} 1\left((i,j) \in \mathcal{S}\right)$.

Moreover, $g(C_{ik_n\Delta_n}) = c_{ik_n\Delta_n}$. Since the function g is C^∞, we can establish the following CLT for the bias corrected estimator:

$$\widehat{\text{RV}}_t^F(A, b) = k_n\Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]}\Big\{(f \circ g)(\widehat{C}_{ik_n\Delta_n})$$

$$-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d+r}\partial^2_{jk,lm}(f\circ g)(\widehat{C}_{ik_n\Delta_n})\left(\widehat{C}_{jl,ik_n\Delta_n}\widehat{C}_{km,ik_n\Delta_n}+\widehat{C}_{jm,ik_n\Delta_n}\widehat{C}_{kl,ik_n\Delta_n}\right)\Big\}. \tag{19}$$

Theorem 4. *Suppose Assumptions 2 and 3 hold and that $u_n \asymp \Delta_n^{\varpi}$, with $5/(12-2\gamma) \leq \varpi < 1/2$. For any fixed t, and as $\Delta_n \to 0$, $k_n \to \infty$ with $k_n^2\Delta_n \to 0$ and $k_n^3\Delta_n \to \infty$, we have*

$$\frac{1}{\sqrt{\Delta_n}}\left(\widehat{\mathrm{RV}}_t^F(A,b)-\mathrm{RV}_t(A,b)\right)\xrightarrow{\text{L}-s}\mathcal{W}_t^F,$$

where $\mathcal{W}^F$ is a standard normal random variable defined on the extension of the original probability space, and its covariance is given by

$$\mathbb{E}((\mathcal{W}_t^F)^2|\mathcal{F}) = 2\int_0^t 2\left(\mathrm{Tr}\left((\omega_s^{*\intercal}c_s\omega_s^*)^2\right)-\mathrm{Tr}\left(\mathcal{M}_{\mathcal{S}}^-(\omega_s^*\omega_s^{*\intercal})\gamma_s\gamma_s^\intercal\mathcal{M}_{\mathcal{S}}^-(\omega_s^*\omega_s^{*\intercal})\gamma_s\gamma_s^\intercal\right)\right)ds.$$

where ω_s^* is the optimal portfolio allocation that solves (6) at time s. The bias-correction term can be written explicitly as

$$\begin{aligned}
Bias(C;RV_t^F) = &-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d+r}\partial^2_{jk,lm}(f\circ g)(C)\left(C_{jl,}C_{km}+C_{jm}C_{kl}\right)\\
=&\frac{1}{k_n}(d-d_{\mathcal{A}})\omega^{*\intercal}c\omega^*+\frac{1}{k_n}(\mathrm{Tr}[M_{\mathcal{S}}^-(E)\gamma\gamma^\intercal M_{\mathcal{S}}^-(F-c^{-1})\gamma\gamma^\intercal]\\
&+\mathrm{Tr}[M_{\mathcal{S}}(E)\gamma\gamma^\intercal]\mathrm{Tr}[M_{\mathcal{S}}(F-c^{-1})\gamma\gamma^\intercal]-\mathrm{Tr}[M_{\mathcal{S}}(E)\gamma\gamma^\intercal M_{\mathcal{S}}(F-c^{-1})\gamma\gamma^\intercal]),
\end{aligned}$$

where E and F are defined in (10) and (12), $\mathcal{M}_{\mathcal{S}}(M) = M_{ij}1\left((i,j)\in\mathcal{S}\right)$, and $\mathcal{M}_{\mathcal{S}}^-(M) = M_{ij}1\left((i,j)\notin\mathcal{S}\right)$.

Example 1. Consider the global minimum variance (GMV) portfolio, in which case we have $A = (\mathbf{1}^\intercal, -\mathbf{1}^\intercal)^\intercal$ and $b = (1,-1)^\intercal$, we have

$$\begin{aligned}
\omega^* =&\frac{c^{-1}\mathbf{1}}{\mathbf{1}^\intercal c^{-1}\mathbf{1}},\quad f(c,A,b)=\frac{1}{\mathbf{1}^\intercal c^{-1}\mathbf{1}},\quad \partial_{jk}f=\frac{\mathbf{1}^\intercal c^{-1}e_je_k^\intercal c^{-1}\mathbf{1}}{(\mathbf{1}^\intercal c^{-1}\mathbf{1})^2},\quad\text{and}\\
\partial^2_{jk,lm}f =&\frac{2}{(\mathbf{1}^\intercal c^{-1}\mathbf{1})^3}(\mathbf{1}^\intercal c^{-1}e_je_k^\intercal c^{-1}\mathbf{1})(\mathbf{1}^\intercal c^{-1}e_le_m^\intercal c^{-1}\mathbf{1})\\
&-\frac{1}{(\mathbf{1}^\intercal c^{-1}\mathbf{1})^2}(\mathbf{1}^\intercal c^{-1}e_le_m^\intercal c^{-1}e_je_k^\intercal c^{-1}\mathbf{1}+\mathbf{1}^\intercal c^{-1}e_je_k^\intercal c^{-1}e_le_m^\intercal c^{-1}\mathbf{1}).
\end{aligned}$$

where $\mathbf{1}$ is a d-dimensional vector of 1s, and e_j is the unit vector with the j-th entry equal to 1. The bias correction term of $\widehat{\mathrm{RV}}_t(A,b)$ is

$$-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}f(\widehat{c}_{i\Delta_n},A,b)\left(\widehat{c}_{jl,i\Delta_n}\widehat{c}_{km,i\Delta_n}+\widehat{c}_{jm,i\Delta_n}\widehat{c}_{kl,i\Delta_n}\right)=\frac{d-1}{k_n}\frac{1}{\mathbf{1}^\intercal\widehat{c}_{i\Delta_n}^{-1}\mathbf{1}}.$$

The asymptotic variance of $\widehat{\mathrm{RV}}_t(A,b)$ is given by

$$Avar^{\mathrm{GMV}}(\widehat{RV}_t) = 2\int_0^t(\mathbf{1}^\intercal c_s^{-1}\mathbf{1})^{-2}ds.$$

Example 2. Consider once again the GMV portfolio, for which we have closed-form formula of the asymptotic variance. By direct calculations, we note that

$$\partial(f\circ g) = (\mathbf{1}^\intercal c^{-1}\mathbf{1})^{-2}\times\begin{pmatrix} \left(\mathrm{Diag}(c^{-1}\mathbf{1})\right)^2 & (c^{-1}\mathbf{1})(\beta^\intercal c^{-1}\mathbf{1})^\intercal - \left(\mathrm{Diag}(c^{-1}\mathbf{1})\right)^2\beta \\ \cdot & \beta^\intercal\left(\mathrm{Diag}(c^{-1}\mathbf{1})\right)^2\beta - (\beta^\intercal c^{-1}\mathbf{1})(\beta^\intercal c^{-1}\mathbf{1})^\intercal \end{pmatrix},$$

where the $(2,1)$ block is given by the transpose of the $(1,2)$ block. Therefore, the asymptotic variance of $\widehat{\mathrm{RV}}_t^F(A,b)$ is given by

$$\begin{aligned} &Avar^{\mathrm{GMV}}(\widehat{RV}_t^F)\\ =&2\int_0^t(\mathbf{1}^\intercal c_s^{-1}\mathbf{1})^{-4}\times\Big\{\mathrm{Tr}\left(c_s^{-1}\mathbf{1}\mathbf{1}^\intercal c_s^{-1}\mathbf{1}\mathbf{1}^\intercal c_s^{-1}\beta_s e_s\beta_s^\intercal\right)+\mathrm{Tr}\left(c_s^{-1}\mathbf{1}\mathbf{1}^\intercal c_s^{-1}\beta_s e_s\beta_s^\intercal c_s^{-1}\mathbf{1}\mathbf{1}^\intercal c_s^{-1}\gamma_s\gamma_s^\intercal\right)\\ &+\mathrm{Tr}\left(\left(\mathrm{Diag}(c_s^{-1}\mathbf{1})\right)^2\gamma_s\gamma_s^\intercal\left(\mathrm{Diag}(c_s^{-1}\mathbf{1})\right)^2\gamma_s\gamma_s^\intercal\right)\Big\}ds. \end{aligned}$$

4.2 Portfolios with Known Weights

In contrast, if portfolio weights $\{\omega_s, 0\le s\le t\}$ were provided without any statistical uncertainty, e.g. the equally weighted portfolio, then the portfolio risk is given by $\int_0^t\omega_s^\intercal c_s\omega_s ds$. Its estimator is therefore given by:

$$\widehat{\mathrm{RV}}_t(\omega) = k_n\Delta_n\sum_{i=0}^{[t/(k_n\Delta_n)]}\omega_{ik_n\Delta_n}^\intercal\widehat{c}_{ik_n\Delta_n}\omega_{ik_n\Delta_n}.$$

The next theorem provides its asymptotic variance. Its proof follows directly from part of the proof in Theorem 3, hence is omitted.

Theorem 5. *Suppose Assumption 1 holds and that $u_n\asymp\Delta_n^{\varpi}$, with $5/(12-2\gamma)\le\varpi<1/2$. For any fixed t, and as $\Delta_n\to 0$, $k_n\to\infty$ with $k_n^2\Delta_n\to 0$ and $k_n^3\Delta_n\to\infty$, we have*

$$\frac{1}{\sqrt{\Delta_n}}\left(\widehat{\mathrm{RV}}_t(\omega)-\int_0^t\omega_s^\intercal c_s\omega_s ds\right)\overset{\mathrm{L}-s}{\longrightarrow}\widetilde{\mathcal{W}}_t,$$

where $\widetilde{\mathcal{W}}$ is a standard normal random variable defined on the extension of the original probability space, and its covariance is given by

$$\mathbb{E}(\widetilde{\mathcal{W}}_t^2|\mathcal{F}) = 2\int_0^t(\omega_s^\intercal c_s\omega_s)^2 ds.$$

Remark 5. Assuming perfect knowledge of the optimal portfolio weights, the asymptotic variance of $\widehat{\mathrm{RV}}_t(\omega)$ is the same as $\widehat{\mathrm{RV}}_t(A,b)$ in Example 1. That is, the first stage estimation of the covariance matrix only contributes to an asymptotic bias. However, without correcting it, the estimator does not admit a central limit theorem. The same result applies to the factor model based estimators.

The factor based estimator of the portfolio risk with known weights is given by:

$$\widehat{\mathrm{RV}}_t^F(\omega) = k_n\Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]} \left\{ \omega_{ik_n\Delta_n}^\intercal g(\widehat{C}_{ik_n\Delta_n})\omega_{ik_n\Delta_n} \right\}.$$

And, we have the next central limit result.

Theorem 6. *Suppose Assumptions 2 and 3 hold and that $u_n \asymp \Delta_n^{\varpi}$, with $5/(12-2\gamma) \leq \varpi < 1/2$. For any fixed t, and as $\Delta_n \to 0$, $k_n \to \infty$ with $k_n^2\Delta_n \to 0$ and $k_n^3\Delta_n \to \infty$, we have*

$$\frac{1}{\sqrt{\Delta_n}}\left(\widehat{\mathrm{RV}}_t^F(\omega) - \int_0^t \omega_s^\intercal c_s \omega_s ds\right) \stackrel{\mathrm{L}-s}{\longrightarrow} \widetilde{\mathcal{W}}_t^F,$$

where $\widetilde{\mathcal{W}}^F$ is a standard normal random variable defined on the extension of the original probability space, and its covariance is given by

$$\mathbb{E}((\widetilde{\mathcal{W}}_t^F)^2|\mathcal{F}) = 2\int_0^t 2\left(\mathrm{Tr}\left((\omega_s^\intercal c_s\omega_s)^2\right) - \mathrm{Tr}\left(\mathcal{M}_{\mathcal{S}}^-(\omega_s\omega_s^\intercal)\gamma_s\gamma_s^\intercal\mathcal{M}_{\mathcal{S}}^-(\omega_s\omega_s^\intercal)\gamma_s\gamma_s^\intercal\right)\right) ds.$$

Remark 6. In practice, we choose $k_n = m_n$, i.e. our portfolio is rebalanced when a new spot covariance estimate becomes available. If this choice of k_n satisfies the conditions that $k_n^2\Delta_n \to 0$ and $k_n^3\Delta_n \to \infty$, then the statistical estimation error documented in Theorems 3 - 6 dominates the discretization error given by Theorem 1. As a result, our inference is relevant for discrete-time rebalanced portfolios.

5 Monte Carlo Simulations

5.1 Asymptotic Efficiency and Asymptotic Bias

From Theorems 3 and 4, we could compare in closed-form the asymptotic efficiency of the alternative approaches. The difference in the asymptotic variances of sample covariance estimator and that of the strict factor-based covariance is given by

$$Avar(\widehat{RV}_t) - Avar(\widehat{RV}_t^F) = 2\int_0^t \left\{ \mathrm{Tr}\left[(\omega_s^{*\intercal}\gamma_s\gamma_s^\intercal\omega_s^*)^2\right] - \mathrm{Tr}\left[\left((\mathrm{Diag}(\omega_s^*))^2\gamma_s\gamma_s^\intercal\right)^2\right] \right\} ds \geq 0, \quad (20)$$

hence the factor-based approach is more efficient, since it uses the knowledge of the factor structure. The efficiency gain comes from the estimation of the spot covariance matrix when evaluating the portfolio risk, because from Theorems 5 and 6, the estimation error in the optimization step only contributes to an asymptotic bias.

Figure 1 provides numerical calculations of the asymptotic relative efficiency between the two estimators, defined as $Avar(\widehat{RV}_t)/Avar(\widehat{RV}_t^F)$, for different risk exposure constraints and across an increasing dimensionality. One intriguing fact is that for the optimal portfolio, the efficiency gain typically becomes more evident when dimension d grows. As the exposure constraint becomes increasingly binding, the percentage of assets selected shrinks, which diminishes the gain in efficiency, as the efficiency only depends on the second stage estimation. As to the equal weighted portfolio, the relative efficiency gain is little, and it decreases as dimension grows larger.

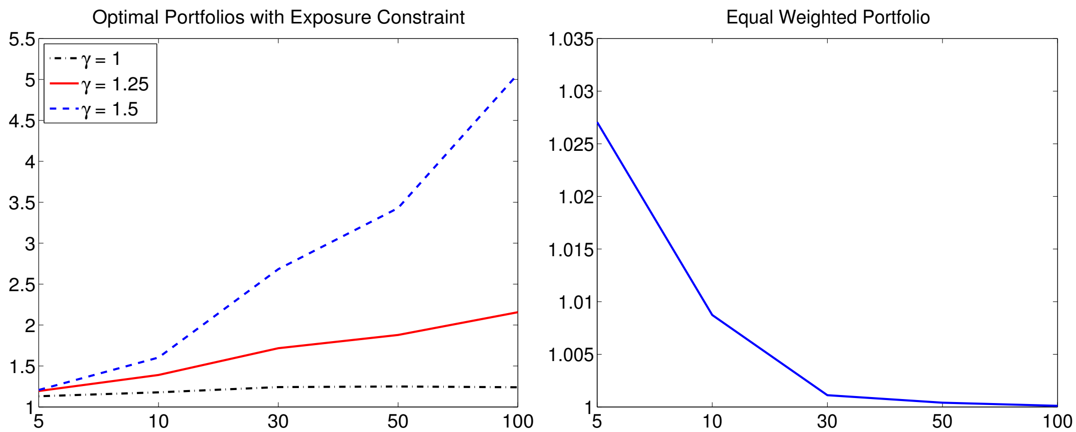

Figure 1: Relative Efficiency of Nonparametric and Semiparametric Estimators

Note: On the left panel, we plot the asymptotic relative efficiency between the nonparametric and semiparametric estimators for the optimal portfolios with different choices of exposure constraints $\|\omega\|_1 \leq \gamma$. On the right panel, we plot the efficiency for the equal-weighted portfolio. The plots are based on the simulation setting detailed in Section 5.

5.2 Finite Sample Performance of the Estimators

In this section we examine the finite sample properties of our estimator. We sample 1000 paths from a continuous-time r-factor model of p assets specified as:

$$dY_{i,t} = \sum_{j=1}^{r} \beta_{i,j,t} dX_{j,t} + dZ_{i,t}, \quad dX_{j,t} = b_j dt + \sigma_{j,t} dW_{j,t} + dJ_{j,t}, \quad dZ_{i,t} = \gamma_{i,t} dB_{i,t} + dM_{i,t},$$

where $i = 1, 2, \ldots, p$, and $j = 1, 2, \ldots, r$. X_j is the jth observable factor. One of the Xs is deemed as the market factor, so that its associated βs are positive.

We allow for time-varying $\sigma_{j,t}$ and $\beta_{i,j,t}$, which evolve according to the following system of equations:

$$d\sigma_{j,t}^2 = \kappa_j(\theta_j - \sigma_{j,t}^2)dt + \eta_j \sigma_{j,t} d\widetilde{W}_{j,t},$$

$$d\beta_{i,j,t} = \begin{cases} \kappa_i(\theta_{i,j} - \beta_{i,j,t}) + \xi_i \sqrt{\beta_{i,j,t}} d\widetilde{B}_{i,j,t} & \text{if the } i\text{th factor is the "market"}, \\ \kappa_i(\theta_{i,j} - \beta_{i,j,t}) + \xi_i d\widetilde{B}_{i,j,t} & \text{otherwise.} \end{cases}$$

where $\widetilde{B}_t$ is a $(p \times r)$-dimensional standard Brownian motion, $\widetilde{W}_t$ is a r-dimensional standard Brownian motion with $\mathbb{E}[dW_{j,t} d\widetilde{W}_{j,t}] = \rho_j dt$.

Each jump process $J_{j,t}$ or $M_{i,t}$ is an independent Poisson with jump sizes following the Gaussian distribution, with a zero mean and a calibrated variance such that the expected quadratic variation of jumps is equal to 1/3 of the expected quadratic variation of the continuous part of X_j or Z_i.

We simulate three different settings, including sampling every 15 seconds and every minute within a week (T = 5/252), and sampling at 5-minute intervals within one month (T = 1/12). The number

	Factor Vol				β			Residual Vol
	κ	θ	η	ρ	κ	θ	ξ	
X_1	3	.09	.3	-.6	1	$U[.25, 3.25]$	.5	$diag\{U[.2, .5]\}$
X_2	4	.04	.4	-.4	2	$\mathcal{N}(0, 1)$	.6	
X_3	5	.06	.3	-.25	3	$\mathcal{N}(0, 1)$	.7	

Table 1: Parameter Values Used in Simulations

Note: Jump intensity for factors and residuals are selected so that every individual path has an expected number of jumps equal to 2. This expectation is held constant to ensure so that the choice of T does not have an impact on the number of jumps the estimator encounters.

of factors is fixed at 3, and the number of assets is chosen from $\{5, 10, 30, 50, 100\}$. We choose $k_n = [\log(d)\Delta_n^{-2/5}]$. All other parameter values are given in Table 1.

We first verify Theorems 3 and 4. The estimation results are provided in Table 2. In all scenarios, the factor-based estimator dominates the fully nonparametric approach. The advantage is due to the portfolio optimization step, as the optimal weights depend on the inverse of the local covariance estimates. Since the factor-based semiparametric estimator has a closed-form inverse, it has a better finite sample performance. Also, the factor-based approach is more efficient than the nonparametric approach, in particular when the dimension becomes larger, which echoes our discussion on the relative efficiency.

We next simulate the case for equal weighted portfolios and report the estimates in Table 3. The results agree with what Theorems 5 and 6 predict. The RMSEs are much closer to 1, compared to the previous case, even for the case with 30 assets. The two estimators yield almost identical performance, which is not surprising given our prior discussion on the relative efficiency and the fact that no optimization is required for this case.

Figure 2 plots the histograms for the standardized estimates of the risk for both the equal weighted portfolio and the optimal portfolio using the two estimators. The dimension of the universe of assets is 100. We find that for the optimal portfolio, the finite sample approximation for the sample-covariance based estimator is not satisfactory, whereas the factor-based estimator performs quite well. As to the equal weighted portfolios, both estimators are equally good. This is not surprising because as dimension increases, the inverse of the sample covariance estimator behaves rather poorly, in contrast with the factor-based approach. It is the inverse of the spot covariance matrix that determines the portfolio allocation in the constrained optimization step.

6 Empirical Work

We apply our methodology to the high-frequency price data for the S&P 100 constituents from January 2006 to January 2012. The data are collected from the Trade and Quote (TAQ) database from the New York Stock Exchange (NYSE). The stocks are sampled at 5-minute intervals to address

	Nonparametric Estimates				Semiparametric Estimates			
	Δ = 15 seconds, T = 1 week				Δ = 15 seconds, T = 1 week			
	True	Bias	Stdev	RMSE	True	Bias	Stdev	RMSE
d=5	6.006	0.018	0.111	1.073	6.006	-0.002	0.105	1.046
d=10	3.238	0.017	0.057	1.094	3.238	0.003	0.048	0.996
d=30	1.154	0.011	0.022	1.299	1.154	0.002	0.014	1.086
d=50	0.677	0.009	0.015	1.522	0.677	0.001	0.007	1.031
d=100	0.351	0.007	0.009	2.015	0.351	0.001	0.003	1.077
	Δ = 1 minute, T = 1 week				Δ = 1 minute, T = 1 week			
	True	Bias	Stdev	RMSE	True	Bias	Stdev	RMSE
d=5	5.651	0.090	0.222	1.175	5.651	0.023	0.204	1.072
d=10	3.215	0.064	0.122	1.298	3.215	0.013	0.093	1.021
d=30	1.122	0.042	0.055	1.790	1.122	0.005	0.028	1.091
d=50	0.681	0.033	0.039	2.200	0.681	0.004	0.014	1.104
d=100	0.373	0.032	0.030	3.480	0.373	0.003	0.006	1.164
	Δ = 5 minutes, T = 1 month				Δ = 5 minutes, T = 1 month			
	True	Bias	Stdev	RMSE	True	Bias	Stdev	RMSE
d=5	23.935	0.368	1.047	1.176	23.935	0.033	0.972	1.084
d=10	12.972	0.286	0.575	1.328	12.972	0.081	0.443	1.073
d=30	4.911	0.120	0.242	1.623	4.911	0.027	0.128	1.165
d=50	3.049	0.099	0.172	1.966	3.049	0.021	0.072	1.190
d=100	1.370	0.042	0.115	2.634	1.370	0.011	0.027	1.231

Table 2: Simulation Results for Optimized Portfolios

Note: In this simulation, we estimate the realized volatility of a portfolio with d assets, rebalanced continuously on an interval $[0, T]$. The portfolio weights are chosen by solving the optimization problem: $\min_w w^\intercal c w$, subject to $\|w\|_1 \leq 1.5$. The True value, and the Bias and the Stdev of the estimates are scaled by 10^3. The RMSE is calculated for the standardized estimates using the estimated asymptotic standard error. The closer the RMSE is relative to 1, the better the finite sample performance of the estimator is.

	Nonparametric Estimates				Semiparametric Estimates			
	Δ = 15 seconds, T = 1 week				Δ = 15 seconds, T = 1 week			
	True	Bias	Stdev	RMSE	True	Bias	Stdev	RMSE
d=5	9.391	0.008	0.169	1.044	9.391	-0.015	0.167	1.045
d=10	8.167	0.012	0.139	1.018	8.167	-0.006	0.138	1.023
d=30	7.359	0.007	0.129	1.076	7.359	-0.009	0.129	1.081
d=50	6.881	0.000	0.123	1.051	6.881	-0.014	0.122	1.066
d=100	6.799	0.001	0.125	1.098	6.799	-0.012	0.125	1.109
	Δ = 1 minute, T = 1 week				Δ = 1 minute, T = 1 week			
	True	Bias	Stdev	RMSE	True	Bias	Stdev	RMSE
d=5	8.872	0.087	0.335	1.078	8.872	0.005	0.324	1.046
d=10	8.119	0.051	0.289	1.056	8.119	-0.018	0.283	1.043
d=30	7.056	0.017	0.267	1.107	7.056	-0.040	0.262	1.130
d=50	6.932	0.028	0.255	1.078	6.932	-0.029	0.253	1.110
d=100	7.207	0.028	0.275	1.166	7.207	-0.031	0.272	1.192
	Δ = 5 minutes, T = 1 month				Δ = 5 minutes, T = 1 month			
	True	Bias	Stdev	RMSE	True	Bias	Stdev	RMSE
d=5	37.982	0.461	1.525	1.095	37.982	-0.028	1.466	1.034
d=10	32.671	0.323	1.306	1.039	32.671	-0.025	1.288	1.029
d=30	31.035	0.129	1.277	1.129	31.035	-0.181	1.257	1.168
d=50	30.935	0.076	1.302	1.166	30.935	-0.251	1.289	1.232
d=100	26.645	0.086	1.206	1.191	26.645	-0.183	1.195	1.239

Table 3: Simulation Results for Equal Weighted Portfolios

Note: In this simulation, we estimate the realized volatility of a equal weighted portfolio with d assets, rebalanced continuously on an interval $[0, T]$. The portfolio weights are fixed at $1/d$ for each asset. The True value, and the Bias and the Stdev of the estimates are scaled by 10^3. The RMSE is calculated for the standardized estimates using the estimated asymptotic standard error. The closer the RMSE is relative to 1, the better the finite sample performance of the estimator is.

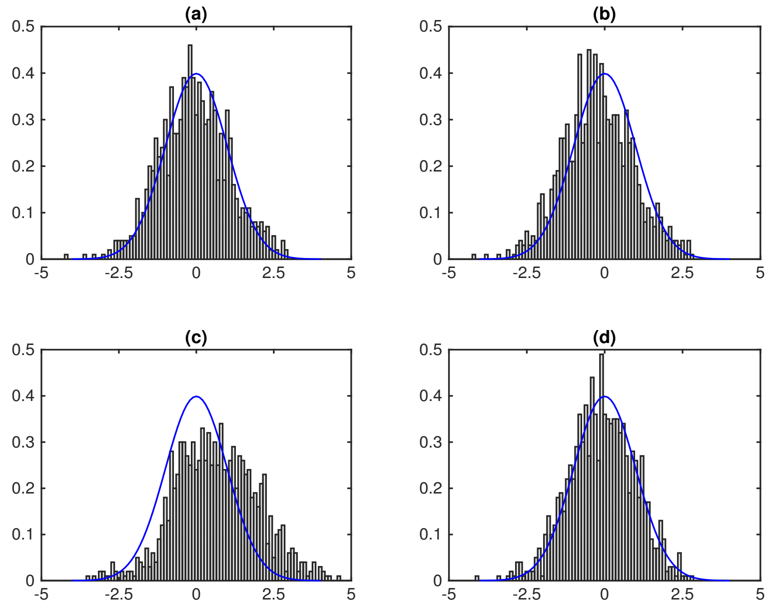

Figure 2: Histograms of the Standardized Estimates

Note: In this figure, we plot the histograms of the standardized estimates. The upper panel shows the case with equal weighted portfolios, and the lower panel illustrates the case of the optimal portfolio with weights selected by solving the optimization problem: $\min_w w^\intercal c w$, subject to $\|w\|_1 \leq 1.5$. The left panel uses the nonparametric sample-covariance-based estimators, whereas the right panel uses the semiparametric factor-based approach. The number of assets is $d = 100$, which are simulated using a factor model with $r = 3$ factors. The sampling frequency is $\Delta = 15$ seconds, and the data span T = 1 week.

the issues of microstructure noise and asynchronous trading.

The goal of our empirical analysis is to investigate the in-sample optimism bias for the realized volatility of an optimized portfolio. In this sense, our results can also be interpreted as a risk *forecast* framework. For each week t, we estimate the covariance matrix using 1-minute returns of week $t-1$, to obtain the $\widehat{c}_{t-1}$. Next we obtain the optimal weight $\widehat{\omega}^*_{t-1}$ by solving the portfolio optimization problem subject to the exposure constraint. Then we calculate the bias-adjusted realized volatility of the optimal portfolio. The new portfolio risk adjusts the in-sample optimism bias. We then calculate the realized volatility using week t's data and the optimal weight $\widehat{\omega}^*_{t-1}$, as our out-of-sample volatility of the optimal portfolio built on week $t-1$. Figure 3 compares the bias-unadjusted volatility, the adjusted volatility, and the out-of-sample volatility. We find that the bias-unadjusted volatility dramatically underestimate the next-period volatility, namely, the in-sample optimism. Moreover, we find the adjusted volatility matches the substantially helps the prediction. The 3-factor model does a poor job of correcting the in sample optimism, while the 12-factor model does a much better job. This is due to the fact that the 3-factor strict model is strongly mispecified empirically, whereas

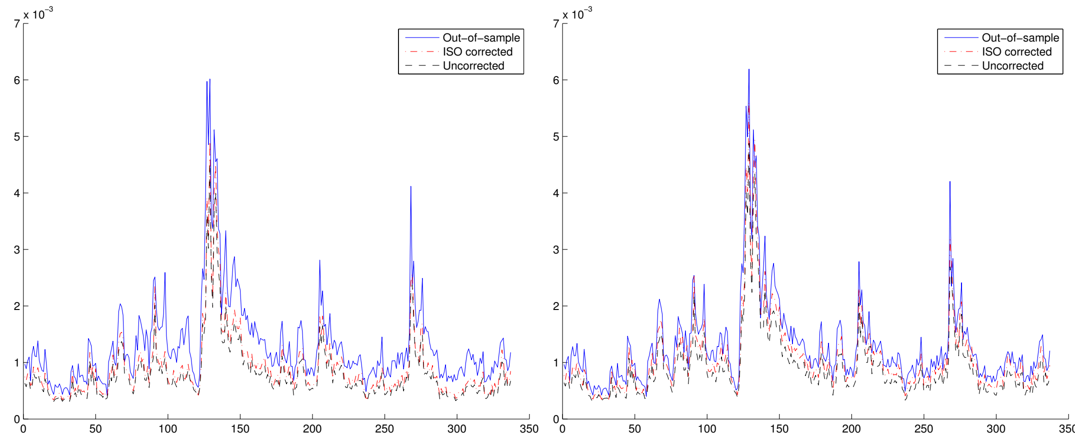

Figure 3: Out-of-sample volatility comparison

Note: Shown above are the volatilities of weekly rebalanced portfolios based on two different strict factor models. Left panel: 3-factor model based portfolios; right panel: 12-factor model based portfolios. In each panel, there are three lines. The Out-of-sample line is the calculated out-of-sample volatility of the optimized portfolio over the subsequent one week. The ISO-corrected line is the in-sample optimism corrected prediction for the next week portfolio volatility, $\hat{\omega}^\intercal g(\hat{C})\hat{\omega} + 2 \cdot \text{Bias}(C; RV_t^F)$. The Uncorrected line is the naive estimate of next week volatility, $\hat{\omega}^\intercal g(\hat{C})\hat{\omega}$.

the 12-factor strict model fits the data much better, e.g. Fan, Furger, and Xiu (2014).

References

Aït-Sahalia, Y., I. Kalnina, and D. Xiu (2014): "The Idiosyncratic Volatility Puzzle: A Reassessment at High Frequency," Discussion paper, The University of Chicago.

Aït-Sahalia, Y., and D. Xiu (2014): "Principal Component Analysis of High Frequency Data," Discussion paper, Princeton University and the University of Chicago.

Basak, G. K., R. Jagannathan, and T. Ma (2009): "Jackknife Estimator for Tracking Error Variance of Optimal Portfolios," *Management Science*, 55(6), 990–1002.

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

Comte, F., and E. Renault (1998): "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

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

Fan, J., Y. Fan, and J. Lv (2008): "High dimensional covariance matrix estimation using a factor model," *Journal Econometrics*, 147, 186–197.

Fan, J., A. Furger, and D. Xiu (2014): "Incorporating Global Industrial Classification Standard into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator with High Frequency Data," Discussion paper, University of Chicago.

Fan, J., Y. Li, and K. Yu (2012): "Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection," *Journal of the American Statistical Association*, 107, 412–428.

Fan, J., Y. Liao, and M. Mincheva (2011): "High Dimensional Covariance Matrix Estimation in Approximate Factor Models," *The Annals of Statistics*, 39, 3320–3356.

——— (2013): "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society Series B*, 75, 603–680.

Fan, J., Y. Liao, and X. Shi (2013): "Risks of large portfolios," *manuscript*.

Fan, J., J. Zhang, and K. Yu (2012): "Vast portfolio selection with gross-exposure constraints," *Journal of the American Statistical Association*, 107, 592–606.

Foster, D., and D. B. Nelson (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64, 139–174.

Frost, P. A., and J. E. Savarino (1988): "For better performance: Constrain portfolio weights," *The Journal of Portfolio Management*, 15(1), 29–34.

Hautsch, N., L. M. Kyj, and P. Malec (2013): "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?," *Journal of Applied Econometrics*.

Jacod, J., and M. Rosenbaum (2013): "Quarticity and Other Functionals of Volatility: Efficient Estimation," *Annals of Statistics*, 41, 1462–1484.

Jagannathan, R., and T. Ma (2003): "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 1651–1684.

Jobson, J., and R. M. Korkie (1981): "Putting Markowitz theory to work," *The Journal of Portfolio Management*, 7(4), 70–74.

Kalnina, I., and D. Xiu (2014): "Nonparametric Estimation of the Leverage Effect using Information from Derivatives Markets," Discussion paper, Université de Montréal and University of Chicago.

Kristensen, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach," *Econometric Theory*, 26(1), pp. 60–93.

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

——— (2004b): "Honey, I shrunk the sample covariance matrix," *Journal of Multivariate Analysis*, 88, 365–411.

——— (2012): "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, 40, 1024–1060.

Li, J., V. Todorov, and G. Tauchen (2013): "Inference Theory on Volatility Functional Dependencies," Discussion paper, Duke University.

Li, J., and D. Xiu (2013): "Spot Variance Regressions," Discussion paper, Duke University and University of Chicago.

Lunde, A., N. Shephard, and K. Sheppard (2014): "Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels," Discussion paper, Aarhus University.

Markowitz, H. (1952): "Portfolio Selection," *Journal of Finance*, 7(1), 77–91.

Michaud, R. O. (1989): "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?," *Financial Analysts Journal*, 45(1), 31–42.

Mykland, P. A., and L. Zhang (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1445.

Renò, R. (2008): "Nonparametric Estimation of the Diffusion Coefficient of Stochastic Volatility Models," *Econometric Theory*, 24, 1174–1206.

Ross, S. A. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Shapiro, A. (1985): "Second-Order Derivatives of Extremal-Value Functions and Optimality Conditions for Semi-Infinite Programs," *Mathematics of Operations Research*, 10(2), 207–219.

Shapiro, A. (1990): "On Concepts of Directional Differentiability," *Journal of Optimization Theory and Applications*, 66(3).

Stock, J. H., and M. W. Watson (2002): "Forecasting using Principal Components from a Large Number of Predictors," *Journal of American Statistical Association*, 97, 1167–1179.

Tao, M., Y. Wang, and H. H. Zhou (2013): "Optimal Sparse Volatility Matrix Estimation for High Dimensional Itō Processes With Measurement Errors," *Annals of Statistics*, 41, 1816–1864.

Wang, Y., and J. Zou (2010): "Vast volatility matrix estimation for high-frequency financial data," *Annals of Statistics*, 38, 943–978.

Appendix

A Proof of Theorem 1

Proof. Note that the right-hand side scaled by δ_n^{-1} can be written as

$$\frac{1}{\delta_n}\sum_{l=1}^{[t/\delta_n]}\int_{(l-1)\delta_n}^{l\delta_n}\Big({\omega^*}^{\intercal}_{(l-1)\delta_n}c_s{\omega^*}_{(l-1)\delta_n}-{\omega^*}^{\intercal}_{s}c_s{\omega^*}_{s}\Big)ds:=\sum_{l=1}^{[t/\delta_n]}\xi_l,$$

where ξ_l is $\mathcal{F}_{(l-1)\delta_n}$-measurable. Note that by the standard estimates for semimartingales, see e.g. (4.3) in Jacod and Rosenbaum (2013):

$$\begin{aligned}
&\Big|\mathbb{E}\Big({\omega^*}^{\intercal}_{(l-1)\delta_n}c_s{\omega^*}_{(l-1)\delta_n}-{\omega^*}^{\intercal}_{s}c_s{\omega^*}_{s}|\mathcal{F}_{(l-1)\delta_n}\Big)\Big|\\
=&|\mathbb{E}\Big({\omega^*}^{\intercal}_{(l-1)\delta_n}c_s{\omega^*}_{(l-1)\delta_n}-{\omega^*}^{\intercal}_{(l-1)\delta_n}c_{(l-1)\delta_n}{\omega^*}_{(l-1)\delta_n}|\mathcal{F}_{(l-1)\delta_n}\Big)\Big|\\
&+\Big|\mathbb{E}\Big({\omega^*}^{\intercal}_{(l-1)\delta_n}c_{(l-1)\delta_n}{\omega^*}_{(l-1)\delta_n}-{\omega^*}^{\intercal}_{s}c_s{\omega^*}_{s}|\mathcal{F}_{(l-1)\delta_n}\Big)\Big|\le K(s-(l-1)\delta_n),
\end{aligned}$$

which implies that $|\mathbb{E}(\xi_l|\mathcal{F}_{(l-1)\delta_n})|\le K\delta_n$. And similarly, we have

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{(l-1)\delta_n\le s\le l\delta_n}\Big|{\omega^*}^{\intercal}_{(l-1)\delta_n}c_s{\omega^*}_{(l-1)\delta_n}-{\omega^*}^{\intercal}_{s}c_s{\omega^*}_{s}\Big||\mathcal{F}_{(l-1)\delta_n}\Big)\\
=&\mathbb{E}\Big(\sup_{(l-1)\delta_n\le s\le l\delta_n}\Big|{\omega^*}^{\intercal}_{(l-1)\delta_n}c_s{\omega^*}_{(l-1)\delta_n}-{\omega^*}^{\intercal}_{(l-1)\delta_n}c_{(l-1)\delta_n}{\omega^*}_{(l-1)\delta_n}\Big||\mathcal{F}_{(l-1)\delta_n}\Big)\\
&+\mathbb{E}\Big(\sup_{(l-1)\delta_n\le s\le l\delta_n}\Big|{\omega^*}^{\intercal}_{(l-1)\delta_n}c_{(l-1)\delta_n}{\omega^*}_{(l-1)\delta_n}-{\omega^*}^{\intercal}_{s}c_s{\omega^*}_{s}\Big||\mathcal{F}_{(l-1)\delta_n}\Big)\le K\delta_n^{1/2},
\end{aligned}$$

so that $\mathbb{E}(|\xi_l|^2|\mathcal{F}_{(l-1)\delta_n})\le K\delta_n$.

Let $\xi'_l=\mathbb{E}(\xi_l|\mathcal{F}_{(l-1)\delta_n})$, and $\xi''_l=\xi_l-\xi'_l$. Then we have

$$\sup_{k\le[t/\delta_n]}\Big|\sum_{l=1}^{k}\xi_l\Big|\le\sum_{l=1}^{[t/\delta_n]}\Big|\xi'_l\Big|+\sup_{k\le[t/\delta_n]}\Big|\sum_{l=1}^{k}\xi''_l\Big|.$$

Since $\sum_{l=1}^{[t/\delta_n]}|\xi'_l|\le K$, and by Doob's inequality, $\mathbb{E}(\sup_{k\le[t/\delta_n]}|\sum_{l=1}^{k}\xi''_l|)\le\sum_{l=0}^{[t/\delta_n]}\mathbb{E}((\xi_l)^2)\le K$, we have $\sum_{l=1}^{[t/\delta_n]}\xi_l=O_p(1)$. □

Theorem 7. *The set $D=\{c:\mathcal{D}(c)\neq\emptyset\}$ is contained in the finite union of hyperplanes, each of which are measure zero.*

Proof. Throughout fix $c\in D$. This implies that there exists a j, $A_{\mathcal{S}}$, and $b_{\mathcal{S}}$, such that

$$e_j^{\intercal}\big(A_{\mathcal{S}}c^{-1}A_{\mathcal{S}}^{\intercal}\big)^{-1}b_{\mathcal{S}}=0$$

Here, $A_{\mathcal{S}}$ and $b_{\mathcal{S}}$ are subsets of the constraints such that $A_{\mathcal{S}}$ have full row rank. From here on, we will abuse the notation and drop the subscript.

We begin with the observation that for any $B \in \mathcal{M}_+^{d\times d}$, we have the decomposition from the Woodbury formula:

$$\begin{aligned}&\Big(A\big(c+h(B-C)\big)^{-1}A^\intercal\Big)^{-1}\\&=\Big(Ac^{-1}A^\intercal\Big)^{-1}+h\Big(Ac^{-1}A^\intercal\Big)^{-1}Ac^{-1}(B-c)\Big[I-h(F-c^{-1})(B-c)\Big]^{-1}c^{-1}A^\intercal\Big(Ac^{-1}A^\intercal\Big)^{-1}\end{aligned}$$

We may formally expand $\big[I-h(F-c^{-1})(B-c)\big]^{-1}$ as a power series, if $h\|(F-c^{-1})(B-c)\|<1$. So will limit $h\in(-a,a)$, where $a=\frac{1}{\|(F-c^{-1})(B-c)\|}$. Recall, the formal power series is:

$$\big[I-h(F-c^{-1})(B-c)\big]^{-1}=\sum_{i=1}^{\infty}h^i\big((F-c^{-1})(B-c)\big)^i$$

Thus, we see

$$\Big(A\big(c+h(B-C)\big)^{-1}A^\intercal\Big)^{-1}=\phi(h)$$

where

$$\phi(h)=\Big(Ac^{-1}A^\intercal\Big)^{-1}+\Big(Ac^{-1}A^\intercal\Big)^{-1}Ac^{-1}\left[\sum_{i=0}^{\infty}h^{i+1}Q_i\right]c^{-1}A^\intercal\Big(Ac^{-1}A^\intercal\Big)^{-1}$$
$$Q_i=(B-c)\big((F-c^{-1})(B-c)\big)^i$$

With our choice of c,j,A,b, it is clear that $e_j\phi(h)b=0$. At this point, let us fix $B\in\mathcal{M}_+^{d\times d}$ such that

$$e_j^\intercal\big(AB^{-1}A^\intercal\big)^{-1}b=0$$

One can see that $\phi(h)$ is analytic. Thus, $e_j\phi(h)b$ is constant iff $e_j\phi^{(n)}(0)b=0$ for every $n\geq 1$. We will only need the first derivative, which is given by:

$$\phi'(h)=\Big(Ac^{-1}A^\intercal\Big)^{-1}Ac^{-1}\left[\sum_{i=1}^{\infty}(i+1)h^iQ_i\right]c^{-1}A^\intercal\Big(Ac^{-1}A^\intercal\Big)^{-1}$$

$$\begin{aligned}\phi'(0)&=\Big(Ac^{-1}A^\intercal\Big)^{-1}Ac^{-1}Q_0c^{-1}A^\intercal\Big(Ac^{-1}A^\intercal\Big)^{-1}\\&=\Big(Ac^{-1}A^\intercal\Big)^{-1}Ac^{-1}(B-c)c^{-1}A^\intercal\Big(Ac^{-1}A^\intercal\Big)^{-1}\\&=EBE^\intercal-\Big(Ac^{-1}A^\intercal\Big)^{-1}\end{aligned}$$

,where $E=\big(Ac^{-1}A^\intercal\big)^{-1}Ac^{-1}$.

Thus,

$$e_j^\intercal\phi'(0)b=e_j^\intercal EBE^\intercal b-e_j\big(Ac^{-1}A^\intercal\big)^{-1}b=e_j^\intercal EBE^\intercal b$$

Then, it is clear that a necessary condition for $e_j\phi(h)b$ to be constant is that $e_j^\intercal EBE^\intercal b=0$. Now, this clearly defines a hyperplane in $\mathcal{M}_+^{d\times d}$.

□

B Proof of Theorem 2

Proof. To begin the proof we write the Lagrangian and KKT conditions associated with (6):

$$L(w, \lambda) = \frac{1}{2}\omega^\intercal c\omega + \lambda^\intercal(Aw - b). \tag{21}$$

For this problem the KKT conditions are given by:

$$c\omega + A^\intercal\lambda = 0, \tag{22}$$

$$A\omega - b \leq 0, \tag{23}$$

$$\lambda \geq 0, \tag{24}$$

$$\lambda_i(A_i\omega - b_i) = 0, \quad \text{for any } i. \tag{25}$$

Since c is positive definite, solving (22) for ω we get that $\omega^* = -c^{-1}A^\intercal\lambda^*$. Plugging back into the Lagrangian function, we arrive at the dual problem:

$$\max_\lambda -\lambda^\intercal b - \frac{1}{2}\lambda^\intercal Ac^{-1}A^\intercal\lambda, \quad \text{subject to } \lambda \geq 0. \tag{26}$$

We will find it simpler to work with the dual problem and so we will calculate all derivatives of f through its dual. The fact that Strong Duality holds in this canonical problem allows us to only consider the derivative calculation from the perspective of the dual.

For a given set of active constraints, $\mathcal{A}(c)$, the dual problem has a particularly simple closed-form solution. Let $\lambda_\mathcal{A}$ be the subset of the rows of λ that are possibly non-zero, setting all other entries to zero. Under the additional assumption that the rows of $A_\mathcal{A}$ are linearly independent we have,

$$\lambda^*_\mathcal{A} = -(A_\mathcal{A}c^{-1}A^\intercal_\mathcal{A})^{-1}b_\mathcal{A}, \qquad \omega^* = c^{-1}A^\intercal(A_\mathcal{A}c^{-1}A^\intercal_\mathcal{A})^{-1}b_\mathcal{A}. \tag{27}$$

Plugging this back into the dual problem, and asserting strong duality, we find that the optimum objective value is,

$$\frac{1}{2}(\omega^*)^\intercal c\omega^* = -(\lambda^*)^\intercal b - \frac{1}{2}(\lambda^*)^\intercal Ac^{-1}A^\intercal\lambda^* = \frac{1}{2}b^\intercal_\mathcal{A}(A_\mathcal{A}c^{-1}A^\intercal_\mathcal{A})^{-1}b_\mathcal{A} = -\frac{1}{2}(\lambda^*)^\intercal b. \tag{28}$$

We will need one more piece of notation. For a given subset of constraints $\mathcal{S}$ such that the rows of $A_\mathcal{S}$ are linearly independent, let $K(\mathcal{S})$ be the collection of matrices c that satisfy both of the following inequalities:

$$-(A_\mathcal{S}c^{-1}A^\intercal_\mathcal{S})^{-1}b_\mathcal{S} \geq 0, \tag{29}$$

$$Ac^{-1}A^\intercal_\mathcal{S}(A_\mathcal{S}c^{-1}A^\intercal_\mathcal{S})^{-1}b_\mathcal{S} \leq b. \tag{30}$$

By construction, any pair of

$$\lambda_\mathcal{S} = -(A_\mathcal{S}c^{-1}A^\intercal_\mathcal{S})^{-1}b_\mathcal{S}, \quad w = c^{-1}A^\intercal_\mathcal{S}(A_\mathcal{S}c^{-1}A^\intercal_\mathcal{S})^{-1}b_\mathcal{S} \tag{31}$$

will satisfy the first KKT condition. If, additionally, it satisfies the above inequalities (29), (30), this pair is a KKT point. Moreover, since the primal is positive definite, this pair is then optimal. Thus, for $c \in K(\mathcal{S})$, the optimal pair is given by (31).

To calculate the derivatives we will use the following lemma:

Lemma 1. *Under the assumptions of Theorem 2, we have*

i. the sets $K(\mathcal{S})$ are closed and cover the domain of c;

ii. for every c there is some $\mathcal{S}$ such that c is a limit point in $K(\mathcal{S})$;

iii. for every c and direction UV, there is an $\bar{h}$ and $\mathcal{S}$, such that $(c + hUV) \in K(\mathcal{S})$ for all $0 \leq h \leq \bar{h}$.

Proof. (i.) The inequalities (29) and (30) involve the level set of a continuous (actually C^∞) function of c. Thus they are closed in the domain of c. It is clear that the sets $K(\mathcal{S})$ will cover the domain of c, since $\mathcal{S}$ may be the empty set.

(ii.) Partition the entire domain of c into the closed sets of the form $K(\mathcal{S})$ and call the sets $K(\mathcal{S}_1), ..., K(\mathcal{S}_M)$. Here, M will be large but finite. Suppose, to reach a contradiction, that c is an isolated point in each of sets to which it belongs. Since there are only finitely many sets in our partition, it must be the case that c is an isolated point in the domain. This is a contradiction. Thus, there must be a set $K(\mathcal{S})$ such that c is a limit point in $K(\mathcal{S})$.

(iii.) Fix c and a direction UV. Part (ii.) of the lemma states that there is a set $K(\mathcal{S})$ so that for every $\bar{h} > 0$ there is an h, $0 < h \leq \bar{h}$, so that $c \in K(\mathcal{S})$ and $(c + hUV) \in K(\mathcal{S})$.

On the other hand, since the inequalities (29) and (30) involve the level sets of C^∞ functions of c, we know that there exists a one-sided neighborhood, $(0, \bar{h})$ for which either $(c+hUV) \in K(\mathcal{S})$ for all $h \in (0, \bar{h})$ or $(c + hUV) \notin K(\mathcal{S})$ for all $h \in (0, \bar{h})$. This is because a function which is differentiable on a compact set can only enter or leave a given compact set in the function's range finitely many times. But now, since we chose $K(\mathcal{S})$ to be the set for which c is a limit point, we know that it must be the case that $(c + hUV) \in K(\mathcal{S})$ for all $h \in (0, \bar{h})$. □

As a first application of Lemma (2), we will show that $f(c; A, B)$ is locally Lipschitz. Recall the definition of locally Lipschitz, for every c there is a constant K and an open neighborhood O, about c for which if $c_1, c_2 \in O$, then $|f(c_1) - f(c_2)| <= M\|c_1 - c_2\|$.

Lemma 2. *$f(c; A, B)$ is locally Lipschitz.*

We make use of the following lemma to prove it.

Lemma 3. *Fix a c and direction UV. There are constants $\bar{h}$ and M, such that for every $\tilde{c} = c+hUV$, with $0 \leq h \leq \bar{h}$, $|f(\tilde{c}) - f(c)| \leq M\|\tilde{c} - c\|$.*

Proof. Fix c and a direction UV. Let $\bar{h}$ and $\mathcal{S}$ be the value and set provided by Lemma (2), respectively. That is, $\tilde{c}(h) = c + hUV \in K(S)$ for all $0 \leq h \leq \bar{h}$. Let $g(h) = f(c(h)) = b_{\mathcal{S}}^\intercal \left(A_{\mathcal{S}}\tilde{c}(h)^{-1}A_{\mathcal{S}^\intercal}\right)^{-1} b_{\mathcal{S}}$ be defined on the interval $h \in [0, \bar{h}]$. Since, the set of active constraints $\mathcal{S}$ does not change over the range g, it is clear that g is C^∞ as a function of h, and so it is Lipschitz. □

Proof of Lemma 2 . Fix a c.

(2) Now, there are only finitely many possible sets of active constraints. Let $K(\mathcal{S}_1), \ldots, K(\mathcal{S}_m)$ be the sets which intersect every neighborhood of c. For each set $K(\mathcal{S}_i)$, the □

With this lemma in hand the calculation of the one-sided directional derivatives becomes one of simple matrix calculations. To fix terms, the directional derivatives in the direction of a matrix UV is,

$$\lim_{h \to 0^+} \frac{f(c + hUV) - f(c)}{h}. \tag{32}$$

Lemma 1 states that there is a one-sided neighborhood over which the KKT set and therefore a certain active set of constraints remains constant. We may then fix a certain active set, $\mathcal{A}$, over the limit.

$$\lim_{h \to 0^+} \frac{f(c + hUV) - f(c)}{h} = \lim_{h \to 0^+} -\frac{1}{2} \frac{b_{\mathcal{A}}^\intercal (A_{\mathcal{A}}(c + hUV)^{-1} A_{\mathcal{A}}^\intercal)^{-1} b_{\mathcal{A}} - b_{\mathcal{A}}^\intercal (A_{\mathcal{A}} c^{-1} A_{\mathcal{A}}^\intercal)^{-1} b_{\mathcal{A}}}{h}$$

As justified by the lemma we will drop the dependence on $\mathcal{A}$ to ease the notation.

To evaluate the derivatives we will need the following formulas, which are simply applications of the Morrison formula. For a fixed U and V,

$$\begin{aligned}
(c + hUV)^{-1} &= c^{-1} - hc^{-1}U(I + hVc^{-1}U)^{-1}Vc^{-1} \\
&= c^{-1} - hc^{-1}UVc^{-1} + o(h), \\
(A(c + hUV)^{-1}A^\intercal)^{-1} &= (Ac^{-1}A^\intercal - hAc^{-1}U(I + hVc^{-1}U)^{-1}Vc^{-1}A^\intercal)^{-1} \\
&= (Ac^{-1}A^\intercal)^{-1} - h(Ac^{-1}A^\intercal)^{-1}\tilde{U}(I + h\tilde{V}(Ac^{-1}A^\intercal)^{-1}\tilde{U})^{-1}\tilde{V}(Ac^{-1}A^\intercal)^{-1} \\
&= (Ac^{-1}A^\intercal)^{-1} + h(Ac^{-1}A^\intercal)^{-1}\tilde{U}\tilde{V}(Ac^{-1}A^\intercal)^{-1} + o(h), \\
\tilde{U} &= Ac^{-1}U(I + hVc^{-1}U)^{-1/2}, \\
\tilde{V} &= (I + hVc^{-1}U)^{-1/2}Vc^{-1}A^\intercal.
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{f(c + hUV) - f(c)}{h} &= \frac{1}{2} \frac{[b^\intercal (A(c + hUV)^{-1}A^\intercal)^{-1}b - b^\intercal (Ac^{-1}A^\intercal)^{-1}b]}{h} \\
&= \frac{1}{2} \frac{[b^\intercal \left((Ac^{-1}A^\intercal)^{-1} + h(Ac^{-1}A^\intercal)^{-1}\tilde{U}\tilde{V}(Ac^{-1}A^\intercal)^{-1} + o(h)\right) b - b^\intercal (Ac^{-1}A^\intercal)^{-1}b]}{h} \\
&= \frac{1}{2} b^\intercal \left((Ac^{-1}A^\intercal)^{-1}\tilde{U}\tilde{V}(Ac^{-1}A^\intercal)^{-1}\right) b.
\end{aligned}$$

Taking limits gives

$$\begin{aligned}
\lim_{h \to 0^+} \frac{f(c + hUV) - f(c)}{h} &= \frac{1}{2} b^\intercal (Ac^{-1}A^\intercal)^{-1} Ac^{-1}UVc^{-1}A^\intercal (Ac^{-1}A^\intercal)^{-1} b & (33) \\
&= \frac{1}{2} (\lambda^*)^\intercal Ac^{-1}UVc^{-1}A^\intercal \lambda^*. & (34)
\end{aligned}$$

$$= \frac{1}{2}(\lambda^*)^\intercal A_{\mathcal{A}} c^{-1} U V c^{-1} A_{\mathcal{A}}^\intercal \lambda^* \tag{35}$$

The limit as $h \to 0^-$ is analogous, and one will arrive at the same formula, with a (potentially) different active set in place of $A_{\mathcal{A}}$. To be precise, let the active set as $h \to 0^-$ be $\mathcal{B}$. Then,

$$\lim_{h\to 0^-} \frac{f(c + hUV) - f(c)}{h} = \frac{1}{2}(\lambda^*)^\intercal A_{\mathcal{B}} c^{-1} U V c^{-1} A_{\mathcal{B}}^\intercal \lambda^* \tag{36}$$

Now, $\mathcal{A}$ and $\mathcal{B}$ must each be active sets for the optimization at c. An obvious property of an active set, $\mathcal{A}$ is that $\lambda_i^* = 0$ for $i \notin \mathcal{A}$ and $\lambda_i^* > 0$ implies $i \in \mathcal{A}$. Thus, $\lambda_i^* = 0$ for $i \in \mathcal{A}^\complement \cap \mathcal{B}^\complement$, and λ^* is supported on $\mathcal{A} \cap \mathcal{B}$. Thus, $(\lambda^*)^\intercal A_{\mathcal{A}} = (\lambda^*)^\intercal A = (\lambda^*)^\intercal A_{\mathcal{B}}$. This shows that the two sides of the derivative exist and are equal. In the sequel will call this, $\partial_{UV} f$.

We now turn to calculating the one-sided second order directional derivatives

$$\lim_{h\to 0^+} \frac{\partial_{UV} f(c + hWZ) - \partial_{UV} f(c)}{h}.$$

Expanding each occurrence of c in (34) using the matrix expansions we get five important terms.

$$\partial f_{UV}(c + hWZ) = \partial f_{UV}(c) + hM_1 + hM_2 + hM_3 + hM_4 + o(h), \tag{37}$$

where

$$\begin{aligned}
M_1 =& \frac{1}{2} b^\intercal \left((Ac^{-1}A^\intercal)^{-1} \tilde{W}\tilde{Z} (Ac^{-1}A^\intercal)^{-1} \right) Ac^{-1} UVc^{-1} A^\intercal (Ac^{-1}A^\intercal)^{-1} b, \\
M_2 =& \frac{1}{2} b^\intercal (Ac^{-1}A^\intercal)^{-1} Ac^{-1} UVc^{-1} A^\intercal \left((Ac^{-1}A^\intercal)^{-1} \tilde{W}\tilde{Z} (Ac^{-1}A^\intercal)^{-1} \right) b, \\
M_3 =& -\frac{1}{2} b^\intercal (Ac^{-1}A^\intercal)^{-1} A \left(c^{-1} WZ c^{-1} \right) UVc^{-1} A^\intercal (Ac^{-1}A^\intercal)^{-1} b, \\
M_4 =& -\frac{1}{2} b^\intercal (Ac^{-1}A^\intercal)^{-1} Ac^{-1} UV \left(c^{-1} WZ c^{-1} \right) A^\intercal (Ac^{-1}A^\intercal)^{-1} b.
\end{aligned}$$

This gives,

$$\begin{aligned}
&\lim_{h\to 0^+} \frac{\partial f_{UV}(c + hUV) - \partial f_{UV}(c)}{h} = \lim_{h\to 0^+} M_1 + M_2 + M_3 + M_4 \\
=& \frac{1}{2} b^\intercal \left((Ac^{-1}A^\intercal)^{-1} Ac^{-1} WZ c^{-1} A^\intercal (Ac^{-1}A^\intercal)^{-1} \right) \cdot Ac^{-1} UVc^{-1} A^\intercal (Ac^{-1}A^\intercal)^{-1} b \\
&+ \frac{1}{2} b^\intercal (Ac^{-1}A^\intercal)^{-1} Ac^{-1} UVc^{-1} A^\intercal \cdot \left((Ac^{-1}A^\intercal)^{-1} Ac^{-1} WZ c^{-1} A^\intercal (Ac^{-1}A^\intercal)^{-1} \right) b \\
&- \frac{1}{2} b^\intercal (Ac^{-1}A^\intercal)^{-1} A \left(c^{-1} WZ c^{-1} \right) UVc^{-1} A^\intercal (Ac^{-1}A^\intercal)^{-1} b \\
&- \frac{1}{2} b^\intercal (Ac^{-1}A^\intercal)^{-1} Ac^{-1} UV \left(c^{-1} WZ c^{-1} \right) A^\intercal (Ac^{-1}A^\intercal)^{-1} b.
\end{aligned}$$

Reorganizing and reintroducing the dependence on the active set where it is required helps us see why the second-order derivatives may only be defined up to one-sided limits. Noting that the above value is a scalar we may, through the trace, reorder the multiplications.

$$\lim_{h\to 0^+} \frac{\partial f_{UV}(c + hUV) - \partial f_{UV}(c)}{h}$$

$$=\frac{1}{2}\mathrm{Tr}\Bigg[\Big(c^{-1}A^\intercal\lambda\lambda^\intercal Ac^{-1}\Big)WZc^{-1}A_{\mathcal{A}}^\intercal(A_{\mathcal{A}}c^{-1}A_{\mathcal{A}}^\intercal)^{-1}A_{\mathcal{A}}c^{-1}UV-\Big(c^{-1}A^\intercal\lambda\lambda^\intercal Ac^{-1}\Big)WZc^{-1}UV$$
$$+\Big(c^{-1}A^\intercal\lambda\lambda^\intercal Ac^{-1}\Big)UVc^{-1}A_{\mathcal{A}}^\intercal(A_{\mathcal{A}}c^{-1}A_{\mathcal{A}}^\intercal)^{-1}A_{\mathcal{A}}c^{-1}WZ-\Big(c^{-1}A^\intercal\lambda\lambda^\intercal Ac^{-1}\Big)UVc^{-1}WZ)\Bigg].$$

In the equation above, each of the terms inside the large parentheses are continuous in c. However, since the Lemma 1 only insures the constancy of the active set on one-sided limits the term $A_{\mathcal{S}}^\intercal(A_{\mathcal{S}}c^{-1}A_{\mathcal{S}}^\intercal)^{-1}A_{\mathcal{S}}$ is not the same from both the left and the right. With this in mind it is becomes clear that this second-order derivative will not, in general, be continuous.

For the final part of the theorem, consider the case $\mathcal{D}(c)=\emptyset$. This implies $\lambda_i^*=0 \implies \omega_i^*>0$ and $\omega_i^*=0 \implies \lambda_i^*>0$. From the continuity of λ^* and ω^* as functions of c, we get that there exists a uniform h over all (i,j) such that for $0\le\bar{h}<h$

$$\begin{aligned}\mathcal{A}(c)&=\mathcal{A}(c+\bar{h}e_ie_j^\intercal),\\ \mathcal{I}(c)&=\mathcal{I}(c+\bar{h}e_ie_j^\intercal),\\ \mathcal{D}(c)&=\mathcal{D}(c+\bar{h}e_ie_j^\intercal)=\emptyset.\end{aligned}$$

□

Thus, $A_{\mathcal{A}}^\intercal(A_{\mathcal{A}}c^{-1}A_{\mathcal{A}}^\intercal)^{-1}A_{\mathcal{A}}$ is constant in this neighborhood of c and so the two one-sided derivatives are equal.

Lemma 4. *Under the conditions of Theorem 2, $\|E\|$ and $\|F\|$ are bounded by a polynomial in the entries of c.*

We will need the following lemma to establish it.

Lemma 5. $\det(B)=0 \iff \det(ABA^\intercal)=0$ *for some A, a full row rank k by p matrix with $k\le p$.*

Proof. "$\implies$" $\det(B)=0$ is equivalent to B being singular. Thus, there is a p-vector, $x\neq 0$, for which $x^\intercal Bx=0$ and the RHS statement holds.

"$\impliedby$" Suppose there is some A which is k by p and has full row rank for which $\det(ABA^\intercal)=0$. Again, this is equivalent to $ABA^\intercal$ being singular, and so there is a k-vector, $x\neq 0$, for which $x^\intercal ABA^\intercal x=0$. Since A has full row rank, $y^\intercal=x^\intercal A\neq 0$ and is a p-vector. But, then $y^\intercal By=0$ and so $\det(B)=0$. □

Proof of Lemma 4. To begin we will use Cramer's Rule throughout, so we remind ourselves of this identity,

$$c^{-1}=\frac{1}{\det(c)}\,\mathrm{adj}(c)$$

First, we will show that when $\det(c)\ge q>0$, we also have that there is a constant $q'_{\mathcal{B}}>0$ for which $\det(A_{\mathcal{B}}\,\mathrm{adj}(c)A_{\mathcal{B}}^\intercal)\ge q'_{\mathcal{B}}$, for any full row-rank submatrix $A_{\mathcal{B}}$ of A. From Lemma 5, $\det(c)$ and $\det(A_{\mathcal{B}}\,\mathrm{adj}(c)A_{\mathcal{B}}^\intercal)$ have the same zeros. Moreover, $\det(\mathrm{adj}(c))=\det(c)^{p-1}\ge q^{p-1}>0$, so that $\mathrm{adj}(c)$ is positive definite. Further, the entries of $\mathrm{adj}(c)$ are polynomial functions of the entries of c.

Letting, $q' = \min q_{\mathcal{B}}$, noting that there are finitely many sub matrices of A.

With this lower bound in hand, let us first consider the quantity $\|E\|$. Recall, $E = \omega^* \omega^{*\intercal}$ and Equation (27).

$$\begin{aligned}
\omega^* &= c^{-1} A^\intercal (A_{\mathcal{A}} c^{-1} A_{\mathcal{A}}^\intercal)^{-1} b_{\mathcal{A}} \\
&= \frac{1}{\det(c)} \operatorname{adj}(c) A_{\mathcal{A}}^\intercal \left(\frac{1}{\det(c)} A_{\mathcal{A}} \operatorname{adj}(c) A_{\mathcal{A}}^\intercal \right)^{-1} b_{\mathcal{A}} \\
&= \frac{1}{\det\left(A_{\mathcal{A}} \operatorname{adj}(c) A_{\mathcal{A}}^\intercal\right)} \operatorname{adj}(c) A_{\mathcal{A}}^\intercal \operatorname{adj}\left(A_{\mathcal{A}} \operatorname{adj}(c) A_{\mathcal{A}}^\intercal\right) b_{\mathcal{A}} \\
&\leq \frac{1}{q'} \operatorname{adj}(c) A_{\mathcal{A}}^\intercal \operatorname{adj}\left(A_{\mathcal{A}} \operatorname{adj}(c) A_{\mathcal{A}}^\intercal\right) b_{\mathcal{A}}
\end{aligned}$$

where the last inequality means entrywise comparison. This explicitly states that the entries of ω^* are bounded by polynomials in the entries of c. Then, of course, so is $\|E\|$. $\|F\|$ can be bounded similarly.

□

C Proof of Theorem 3

Proof. To prove this, we need a lemma, which will be used for the following-up proofs of Theorems 4 - 6 as well.

Lemma 6. *Suppose D is a closed set with Lebesgue measure 0 in the space of positive-definite matrices $\mathcal{M}_+^{d\times d}$. Outside D we have $f \in \mathbb{C}^3$ and $\left\|\partial^j f(c)\right\| \leq K(1 + \|c\|^{\zeta - j})$, for some $\zeta \geq 3$ and $j = 0, 1, 2, 3$. Then we have as $k_n^2 \Delta_n \to 0$, and $k_n^3 \Delta_n \to \infty$, as long as $(2\zeta - 1)/(4\zeta - 2\gamma) \leq \varpi < 1/2$, we have*

$$\frac{1}{\sqrt{\Delta_n}} \left(\Delta_n \sum_{i=0}^{[t/\Delta_n]} \left(f(\widehat{c}_{i\Delta_n}) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} f(\widehat{c}_{i\Delta_n}) (\widehat{c}_{jl,i\Delta_n} \widehat{c}_{km,i\Delta_n} + \widehat{c}_{jm,i\Delta_n} \widehat{c}_{kl,i\Delta_n}) 1_{\{\widehat{c}_{i\Delta_n} \in D^c\}} \right) - \int_0^t f(c_s) ds \right) \xrightarrow{\text{L}-s} \mathcal{W}_t,$$

where $\mathcal{W}$ is a Gaussian process defined on an extension of the original probability space, with conditional covariance matrix given by

$$\mathbb{E}(\mathcal{W}_t^\intercal \mathcal{W}_t | \mathcal{F}) = \int_0^t \sum_{j,k,l,m=1} \partial_{jk} f(c_s) \partial_{lm} f(c_s) (c_{jl,s} c_{km,s} + c_{jm,s} c_{kl,s}) ds.$$

Proof of Lemma 6. This lemma is an extension of Theorem 3.2 of Jacod and Rosenbaum (2013). By localization, we can find a sequence of stopping times $\{\tau_m^0\}_{m\geq 1}$ going to ∞, such that $\|c_s\| \leq K_m$, $\|\tilde{\sigma}_s \tilde{\sigma}_s^\intercal\| \leq K_m$, $|\tilde{\delta}^2(\omega, s, z)|_\infty \wedge 1 \leq K_m$, for any $0 \leq s \leq \tau_m^0$.

We then define an open set $O_{m,\varepsilon} = \{c \in \mathcal{M}_+^{d\times d} : dist(c, D \cap \bar{B}(0, K_m)) < \varepsilon\}$, where $dist(c, A)$ denotes the distance between a point $c \in \mathcal{M}_+^{d\times d}$ and some compact set A, and $\bar{B}(0, K_m) = \{c \in$

$\mathcal{M}_+^{d\times d} : \|c\| \leq K_m\}$. Since D is a closed set, $D \cap \bar{B}(0, K_m)$ is compact. For any $\varepsilon > 0$, we can define another two sequences of stopping times $\{\tau_m^1\}_{m\geq 1}$ and $\{\tau_m^2\}_{m\geq 1}$, such that $\tau_1^1 \leq \ldots \leq \tau_m^1 \leq \tau_m^2 \leq \tau_{m+1}^1 \leq \tau_{m+1}^2 \leq \ldots \infty$, $\tau_m^1 = \inf\{s \geq \tau_{m-1}^2 : c_s \in \bar{O}_{m,\varepsilon}\}$, and $\tau_m^2 = \inf\{s \geq \tau_m^1 : c_s \in O_{m,2\varepsilon}^c\}$, where by convention $\tau_0^2 = 0$. Note that $\left\|c_{\tau_m^2\wedge t} - c_{\tau_m^1\wedge t}\right\| \geq \varepsilon$, and that $\sum_{1\leq m\leq\infty} \left\|c_{\tau_m^2\wedge t} - c_{\tau_m^1\wedge t}\right\|^2$ converges to the quadratic variation of c which is bounded by some $K_m\cdot$ const, therefore there exists only a finite number of τ_m^1s and τ_m^2s which are smaller than t.

By Urysohn's lemma, there exists a $\mathbb{C}^\infty$ function $\chi_m : \mathcal{M}_+^{d\times d} \to [0,1]$ such that $\chi_m(c) = 0$ for $c \in D$, and $\chi_m(c) = 1$ for $c \in \bar{O}_{m,\varepsilon/2}^c$. Applying Theorem 3.2 in Jacod and Rosenbaum (2013), we have the central limit result for $f \cdot \chi_m$. Note that for any $s \in [0,t] \cap \bigcup_{m\geq 0}(\tau_m^2, \tau_{m+1}^1)$, $c_s \in \bar{O}_{m,\varepsilon}^c$, hence $f \cdot \chi_m(c_s) = f(c_s)$, and $\partial(f \cdot \chi_m)(c_s) = \partial f(c_s)$. Moreover, by the uniform convergence of $\widehat{c}_{i\Delta_n}$ to $(k_n\Delta_n)^{-1} \cdot \int_{i\Delta_n}^{(i+k_n-1)\Delta_n} c_s ds$, for any $i = 0, 1, 2, \ldots, [t/\Delta_n]$ and $i\Delta_n \in [0,t] \cap \bigcup_{m\geq 0}(\tau_m^2, \tau_{m+1}^1)$, there exists a large enough n, such that $\widehat{c}_{i\Delta_n} \in \bar{O}_{m,\varepsilon/2}^c$, hence $f \cdot \chi(\widehat{c}_{i\Delta_n}) = f(\widehat{c}_{i\Delta_n})$ and $\partial^2(f \cdot \chi)(\widehat{c}_{i\Delta_n}) = \partial^2 f(\widehat{c}_{i\Delta_n})$.

Finally, since c_t has a probability density, by Fubini's theorem, $\text{Leb}(\{s \in [0,t] : c_s \in \bar{O}_{m,2\varepsilon}\}) \to 0$ almost surely, as $\varepsilon \to 0$, since D has Lebesgue measure 0. Thereby, almost surely, we have

$$
\begin{aligned}
&\int_0^t f \cdot \chi(c_s) \to \int_0^t f(c_s)ds, \quad \text{and} \\
&\int_0^t \partial_{jk}(f \cdot \chi)(c_s)\partial_{lm}(f \cdot \chi)(c_s)(c_{jl,s}c_{km,s} + c_{jm,s}c_{kl,s})ds \to \int_0^t \partial_{jk} f(c_s)\partial_{lm} f(c_s)(c_{jl,s}c_{km,s} + c_{jm,s}c_{kl,s})ds,
\end{aligned}
$$

which concludes the proof. □

By Theorem 2, Lemma 6, and direct calculations, we have

$$
\sum_{j,k,l,m=1}^{d} \partial_{jk} f(c, A, b)\partial_{lm} f(c, A, b) \left(c_{jl}c_{km} + c_{jm}c_{kl}\right) = 2\text{Tr}(EcEc) = 2(\omega^{*\intercal}c\omega^*)^2,
$$

which implies the desired asymptotic variance. To obtain the bias-correction term, we note that:

$$
\begin{aligned}
&- \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 f(c, A, b) \left(c_{jl}c_{km} + c_{jm}c_{kl}\right) \\
=& - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \left(E_{kl}(F_{mj} - c_{mj}^{-1}) + E_{mj}(F_{kl} - c_{kl}^{-1})\right) \left(c_{jl}c_{km} + c_{jm}c_{kl}\right) \\
=& - \frac{1}{k_n} \left(\text{Tr}\left[cFcE\right] - \text{Tr}\left[cc^{-1}cE\right] + \text{Tr}\left[cF\right]\text{Tr}\left[cE\right] - d\text{Tr}\left[cE\right]\right) \\
=& \frac{1}{k_n}\left(\left(d + 1 - d_{\mathcal{A}}\right)\left(\omega^{*\intercal}c\omega^*\right) - \left(\omega^{*\intercal}A_{\mathcal{A}}^\intercal(A_{\mathcal{A}}c^{-1}A_{\mathcal{A}}^\intercal)^{-1}A_{\mathcal{A}}\omega^*\right)\right) \\
=& \frac{1}{k_n}\left(d - d_{\mathcal{A}}\right)\left(\omega^{*\intercal}c\omega^*\right),
\end{aligned}
$$

where we use (27) and (28) to obtain that $\omega^{*\intercal}A_{\mathcal{A}}^\intercal(A_{\mathcal{A}}c^{-1}A_{\mathcal{A}}^\intercal)^{-1}A_{\mathcal{A}}\omega^* = b_{\mathcal{A}}^\intercal(A_{\mathcal{A}}c^{-1}A_{\mathcal{A}}^\intercal)^{-1}b_{\mathcal{A}} = \omega^{*\intercal}c\omega^*$. □

D Proof of Theorem 4

Proof. Let $\mathcal{M}_{\mathcal{S}}(M) = M_{ij}1\left((i,j) \in \mathcal{S}\right)$ for some index set $\mathcal{S}$. Also, let Let $\mathcal{M}_{\mathcal{S}}^{-}(M) = M_{ij}1\left((i,j) \notin \mathcal{S}\right)$ We first calculate the derivatives of $f \circ g$, which will be useful for the following calculations. Applying the chain-rule, we have

$$\partial_{jk}(f \circ g) = \sum_{p,q=1}^{d} \partial_{pq} f(g) \partial_{jk} g_{pq}, \quad \text{and}$$

$$\partial^2_{jk,lm}(f \circ g) = \sum_{p,q,r,s=1}^{d} \partial^2_{pq,rs} f(g) \partial_{jk} g_{pq} \partial_{lm} g_{rs} + \sum_{p,q=1}^{d} \partial_{pq} f(g) \partial^2_{jk,lm} g_{pq}.$$

We hereby need the first and second order derivatives of g with respect to C. Note that

$$g(C) = \mathcal{M}_{\mathcal{S}}\left(C_{11} - C_{12}C_{22}^{-1}C_{21}\right) + C_{12}C_{22}^{-1}C_{21},$$

so we obtain

$$\begin{aligned}\partial_{jk} g_{pq} =& (C_{22}^{-1}C_{21})_{k-d,q} \cdot 1_{\{(p,q)\notin\mathcal{S},(j,k)\in C_{12},p=j\}} + (C_{12}C_{22}^{-1})_{p,j-d} \cdot 1_{\{(p,q)\notin\mathcal{S},(j,k)\in C_{21},q=k\}} \\ &+ 1_{\{(p,q)\in\mathcal{S},(j,k)\in C_{11},p=j,q=k\}} - (C_{12}C_{22}^{-1})_{p,j-d}(C_{22}^{-1}C_{21})_{k-d,q} \cdot 1_{\{(p,q)\notin\mathcal{S},(j,k)\in C_{22}\}},\end{aligned}$$

Note that

$$\partial_{jk}(f \circ g) = \sum_{p,q=1}^{d} \partial_{pq} f(g) \partial_{jk} g_{pq} = \sum_{p,q=1}^{d} E_{pq} \partial_{jk} g_{pq}$$

and

$$\begin{aligned}\partial^2_{jk,lm}(f \circ g) &= \sum_{p,q,u,v=1}^{d} \partial^2_{pq,uv} f(g) \partial_{jk} g_{pq} \partial_{lm} g_{uv} + \sum_{p,q=1}^{d} \partial_{pq} f(g) \partial^2_{jk,lm} g_{pq} \\ &= \sum_{p,q,u,v=1}^{d} \left(E_{qu}(F_{vp} - c_{vp}^{-1}) + E_{vp}(F_{qu} - c_{qu}^{-1})\right) \partial_{jk} g_{pq} \partial_{lm} g_{uv} + \omega^{*\intercal}(\partial^2_{jk,lm} g)\omega^*.\end{aligned}$$

To obtain the asymptotic variance, we note that

$$\begin{aligned}&\sum_{p,q,v,u=1}^{d} \sum_{j,k,l,m=1}^{d+r} E_{pq}E_{vu}(\partial_{jk} g_{pq})(\partial_{lm} g)\left(C_{jl}C_{km} + C_{jm}C_{kl}\right) \\ &= \sum_{p,q,v,u=1}^{d} E_{pq}E_{vu}\Big(c_{qv}c_{up} + c_{qu}c_{vp} - (\gamma\gamma^\intercal)_{up}(\gamma\gamma^\intercal)_{qv}1_{\{(p,q)\notin\mathcal{S},(v,u)\notin\mathcal{S}\}} - (\gamma\gamma^\intercal)_{qu}(\gamma\gamma^\intercal)_{vp}1_{\{(p,q)\notin\mathcal{S},(v,u)\notin\mathcal{S}\}}\Big) \\ &= 2\left(\mathrm{Tr}\left(EcEc\right) - \mathrm{Tr}\left(\mathcal{M}_{\mathcal{S}}^{-}(E)\gamma\gamma^\intercal\mathcal{M}_{\mathcal{S}}^{-}(E)\gamma\gamma^\intercal\right)\right)\end{aligned}$$

By Lemma 6, the bias correction term can be constructed using the second order derivative, which contains two terms. With respect to the second term, we have

$$
\begin{aligned}
&-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d+r}\omega^{*\intercal}\partial^2_{jk,lm}g(C)\omega^*\left(C_{jl}C_{km}+C_{jm}C_{kl}\right)\\
=&-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d+r}\sum_{p,q=1}^{d}\omega^*_p\omega^*_q\partial^2_{jk,lm}g_{pq}(C)\left(C_{jl}C_{km}+C_{jm}C_{kl}\right)\\
=&0.
\end{aligned}
$$

As to the first term, we have

$$
\begin{aligned}
&-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d+r}\sum_{p,q,u,v=1}^{d}\left(E_{qu}(F_{vp}-c^{-1}_{vp})+E_{vp}(F_{qu}-c^{-1}_{qu})\right)\partial_{jk}g_{pq}\partial_{lm}g_{uv}\left(C_{jl}C_{km}+C_{jm}C_{kl}\right)\\
=&-\frac{1}{2k_n}\sum_{p,q,u,v=1}^{d}\left(E_{qu}(F_{vp}-c^{-1}_{vp})+E_{vp}(F_{qu}-c^{-1}_{qu})\right)\sum_{j,k,l,m=1}^{d+r}\partial_{jk}g_{pq}\partial_{lm}g_{uv}\left(C_{jl}C_{km}+C_{jm}C_{kl}\right)\\
=&-\frac{1}{2k_n}\sum_{p,q,u,v=1}^{d}\left(E_{qu}(F_{vp}-c^{-1}_{vp})+E_{vp}(F_{qu}-c^{-1}_{qu})\right)\left(\mathrm{Tr}\left(\partial g_{pq}C(\partial g_{uv})^\intercal C\right)+\mathrm{Tr}\left(\partial g_{pq}C\partial g_{uv}C\right)\right)\\
=&-\frac{1}{2k_n}\sum_{p,q,u,v=1}^{d}\left(E_{qu}(F_{vp}-c^{-1}_{vp})+E_{vp}(F_{qu}-c^{-1}_{qu})\right)\\
&\quad\cdot\left(c_{qv}c_{up}+c_{qu}c_{vp}-(\gamma\gamma^\intercal)_{up}(\gamma\gamma^\intercal)_{qv}1_{\{(p,q)\notin\mathcal{S},(v,u)\notin\mathcal{S}\}}-(\gamma\gamma^\intercal)_{qu}(\gamma\gamma^\intercal)_{vp}1_{\{(p,q)\notin\mathcal{S},(v,u)\notin\mathcal{S}\}}\right)\\
=&\frac{1}{k_n}(d-d_{\mathcal{A}})\omega^{*\intercal}c\omega^*+\frac{1}{k_n}(\mathrm{Tr}[M^-_{\mathcal{S}}(E)\gamma\gamma^\intercal M^-_{\mathcal{S}}(F-c^{-1})\gamma\gamma^\intercal]\\
&\quad+\mathrm{Tr}[M_{\mathcal{S}}(E)\gamma\gamma^\intercal]\mathrm{Tr}[M_{\mathcal{S}}(F-c^{-1})\gamma\gamma^\intercal]-\mathrm{Tr}[M_{\mathcal{S}}(E)\gamma\gamma^\intercal M_{\mathcal{S}}(F-c^{-1})\gamma\gamma^\intercal]).
\end{aligned}
$$

Therefore, the final bias correction term is:

$$
\begin{aligned}
&\frac{1}{k_n}\omega^{*\intercal}\left(c-\mathrm{Diag}(c)\right)\omega^*+\frac{1}{k_n}(d-d_{\mathcal{A}})\omega^{*\intercal}c\omega^*\\
&+\frac{1}{k_n}\omega^{*\intercal}\left(\mathrm{Tr}[\gamma\gamma^\intercal(F-c^{-1})]\mathbb{I}+\gamma\gamma^\intercal(F-c^{-1})-2\mathrm{Diag}(\gamma\gamma^\intercal(F-c^{-1}))\right)\gamma\gamma^\intercal\omega^*.
\end{aligned}
$$

□

E Proof of Theorem 6

Proof. For the bias-correction term, we have already calculated

$$
-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d+r}\omega^\intercal_{i\Delta_n}\partial^2_{jk,lm}g(\widehat{C}_{i\Delta_n})\omega_{i\Delta_n}\left(\widehat{C}_{jl,i\Delta_n}\widehat{C}_{km,i\Delta_n}+\widehat{C}_{jm,i\Delta_n}\widehat{C}_{kl,i\Delta_n}\right)=0,
$$

so there is no bias. As to the asymptotic variance, we first obtain

$$\partial(\omega^{\intercal} g(C)\omega) = \begin{pmatrix} (\mathrm{Diag}(\omega))^2 & \omega(\beta^{\intercal}\omega)^{\intercal} - (\mathrm{Diag}(\omega))^2 \beta \\ \cdot & \left(\beta^{\intercal} (\mathrm{Diag}(\omega))^2 \beta\right) - (\beta^{\intercal}\omega)(\beta^{\intercal}\omega)^{\intercal} \end{pmatrix},$$

where the $(2,1)$ block is given by the transpose of the $(1,2)$ block. As a result, the asymptotic variance is given by

$$\begin{aligned} & 2\int_0^t \mathrm{Tr}\left(\omega_s^{\intercal}\partial g(C_s)\omega_s \cdot C_s \cdot \omega_s^{\intercal}\partial g(C_s)\omega_s \cdot C_s\right) ds \\ =& 2\int_0^t \left\{ \mathrm{Tr}\left[\left((\mathrm{Diag}(\omega_s))^2\gamma_s\gamma_s^{\intercal}\right)^2\right] + \mathrm{Tr}\left[\omega^{\intercal}\beta_s e_s \beta_s^{\intercal}\omega_s\omega_s^{\intercal}c_s\omega_s\right] + \mathrm{Tr}\left[\omega_s^{\intercal}\beta_s e_s\beta_s^{\intercal}\omega_s\omega_s^{\intercal}\gamma_s\gamma_s^{\intercal}\omega_s\right]\right\} ds. \end{aligned}$$

□