ABSTRACT

DINH, HOI. Essays on High-Dimensional Financial Econometrics. (Under the direction of Mehmet Caner).

This dissertation focuses on high-dimensional financial econometrics in specific settings, particularly on very large portfolios where the number of assets exceeds the time span of the portfolio. Each chapter combines rigorous theoretical development with empirical analyses to evaluate the proposed methods in the context of current literature.

In Chapter 1, I study factor models where common factors are measured with errors. I provide a novel method that combines nodewise regression with instrumental variables to obtain consistent estimates of both factor loadings and the precision matrix of the errors. By applying the Sherman-Morrison-Woodbury formula to these estimates, a consistent precision matrix for asset returns is obtained, which in turn enables us to study the Sharpe Ratio—the key metric capturing the trade-off between return and variance in financial economics. Empirical evidence demonstrates that this approach yields a higher Sharpe Ratio compared to other methods in the literature.

Chapter 2 introduces a new concept in “online” investing. Although online investment strategies have been studied previously, most have focused on small portfolios to obtain the optimal allocation for each asset. The proposed approach extends both the theoretical and empirical analysis to high-dimensional settings, offering new insights and algorithmic comparisons that advance our understanding of optimal investment strategies in complex, large-scale portfolios. In particular, I show that on “average”, the algorithm performs as well as the best-fixed strategy in hindsight, where the investor has complete knowledge of the future market outcomes.

Chapter 3 provides a new precision matrix estimator for the econometrics and statistics literature using a latent (hidden) factor setup and its application to financial metrics such as Sharpe Ratio. Our empirical findings provide the first evidence of a double ascent in the Sharpe Ratio.

Essays on High-Dimensional Financial Econometrics

by
Hoi Dinh

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina
2025

APPROVED BY:

Denis Pelletier	Ilze Kalnina
Zheng Li	Mehmet Caner Chair of Advisory Committee

DEDICATION

To my parents, Khanh and Lieu,
and my beloved wife, Hoang Anh

BIOGRAPHY

Hoi Dinh was born in 1994 in a small town in northern Vietnam called Cam Giang. He earned his Bachelor of Science degree in Economics from the Foreign Trade University in Vietnam in 2016 and his Master of Science degree in Economics from Portland State University in 2019.

He is currently pursuing a PhD in Economics at North Carolina State University under the guidance of Professor Mehmet Caner, with his doctoral research focusing on high-dimensional financial econometrics.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone who supported and guided me throughout my doctoral journey and during the completion of this dissertation.

First and foremost, I am profoundly grateful to my advisor, Dr. Caner, whose invaluable guidance, mentorship, and unwavering support have been instrumental in shaping my primary research question and inspiring new avenues of inquiry. I will always cherish our power lunches at the small shop across Nelson Hall, where we savored Chipotle Chicken Sandwiches, spent countless hours proofreading, and explored innovative research ideas. His class, the highlight of my five years at North Carolina State University, left a lasting impact on my academic journey. My first chapter was inspired by ideas from his class during my second year, while my second chapter emerged from our paper discussions in my third year. Each semester, he updated the course materials to help us understand new research trends and spark fresh ideas. Thank you, Dr. Caner, for being my advisor.

I also extend my heartfelt gratitude to my committee members—Dr. Pelletier, Dr. Kalnina, Dr. Li—for their time and invaluable insights.

I am especially grateful to my professor and dear friend, John Gallup.

Finally, I am deeply grateful to my family—my parents, Khanh and Lieu, whose unwavering love and support have carried me through my studies, and my wife, whose constant encouragement has made this journey possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

RESIDUAL-BASED NODEWISE REGRESSION IN MISMEASURED FACTOR MODELS

1.1 Introduction

This study examines the precision matrix of asset returns from factors measured with errors. We derive consistent estimates of the precision matrix of asset returns using instrumental variables (IVs) and nodewise regression in very large portfolios. Specifically, the number of assets significantly exceeds the time span. Our approach enables us to study the Sharpe Ratio and its maximum, which is the key issue in empirical asset pricing, capturing the trade-off between the returns and the risk of a portfolio. Our method combines factor models and instrumental variables with high-dimensional econometrics contributing to the financial econometrics literature.

Calculating the maximum Sharpe Ratio requires an accurately estimated precision matrix. Caner et al. (2023) provides “feasible residual-based nodewise" regression to obtain a consistent estimate of the precision matrix. This method requires consistent estimates of factor loadings in the first stage. However, measurement errors in factors make these estimates inconsistent.

Furthermore, even the well-known Fama–French factors–“small minus big” (SMB) and “high minus low” (HML)–serve as proxies for unobserved common risk factors. Since both SMB and HML are derived from returns on stock portfolios sorted by firm characteristics, they are expected to contain measurement errors. Our first contribution is to apply IVs to provide consistent factor loading estimates and then propose feasible residual-based nodewise regression to estimate the precision matrix of errors. Then, using the Morrison-Woodbury formula as in Caner et al. (2023), we get the consistent precision matrix of asset returns – our second contribution. Lastly, our empirical results show that using our method with more factors can achieve a higher Sharpe Ratio and outperform other methods in the literature.

The most relevant papers on nodewise regression and related methods include the following. Meinshausen and Bühlmann (2006) established the nodewise regression approach and derived optimal results for normally distributed data. Callot et al. (2021) uses the nodewise regression approach from Meinshausen and Bühlmann (2006) to estimates of the variance, risk, and optimal weight of a portfolio based on assets returns. However, their method relies on the key assumption of sparsity in the precision matrix of the returns, an assumption that is more restrictive and less realistic compared to Caner et al. (2023) and our method in this paper. Additionally, Callot et al. (2021) do not provide theoretical proofs for the estimation of the Sharpe Ratio. Loh and Wainwright (2011) extends the application of nodewise regression to the case of high-dimensional settings with measurement errors. However, their approach is based on the precision matrix of the observed data. To obtain the precision matrix for asset returns, it is first necessary to derive a feasible precision matrix of residuals. Therefore, their method is not directly applicable.

Meng et al. (2011) and Bai and Wang (2016) are notable contributions to the literature on measurement error in factors. Meng et al. (2011) proposes the ordinary least squares instrumental variables estimator (OLIVE), in which they use returns of other assets as instruments to estimate the factor loadings. This approach is based on the intuition that all asset returns are influenced by common factors, so one can use the information contained in other asset returns to improve the beta estimate for a specific asset. Bai and Wang (2016) makes use of the principal components method for a large number of instruments, including potentially weak IVs, which improves the efficiency of the IV approach. These papers focus on the estimation of factor loadings, so they provide no estimates of the Sharpe Ratio.

In high-dimensional econometrics, one key paper is Fan et al. (2011). The authors were the first to introduce a method for constructing the precision matrix of returns in large portfolios using factor models. However, their approach relies on the assumption of sparsity and diagonality in the error covariance matrix, whereas we use a less restrictive assumption that only requires sparsity of the precision matrix of the errors.

Regarding additional contributions, Ledoit and Wolf (2003) and Ledoit and Wolf (2004) propose a linear shrinkage estimator for the covariance matrix and demonstrate its application in portfolio optimization. In addition, several earlier papers analyze portfolios and some of the metrics, such as risk and the Sharpe Ratio. For example, Fan et al. (2011), Fan et al. (2013) propose a principal orthogonal complement thresholding estimator (POET).

The organization of this paper is as follows. Section 1.2 introduces the model setup, incorporates measurement error in the factors, outlines our assumptions, and presents the method of using instruments for estimating a feasible precision matrix for errors. Section 1.3 provides the precision matrix estimate for assets returns. Section 1.4 analyzes the consistency of the Sharpe Ratio in a portfolio with a large number of assets in three different scenarios. Section 1.5 explains the simulation set-up and compares various methods. Finally, Section 1.6 provides an exercise in an out-of-sample forecasting.

1.2 Model

In this section, we introduce the model setup and show how to use nodewise regression with instruments to obtain the precision matrix of errors. Next, we present a practical algorithm for implementing feasible nodewise regression. Finally, we show how to apply the Morrison-Woodbury formula, using the precision matrix of errors, in the factors model to derive the precision matrix of asset returns.

1.2.1 Setup

Let $y_{j,t}$ be the j^{th} asset's excess asset return at time t, and f_t^* be the $K \times 1$ vector of commonly observed factors used to explain excess return. For asset $j = 1, ..., p$ and period $t = 1, ..., n$, let

$$y_{j,t} = b_j' f_t^* + \epsilon_{j,t} \tag{1.1}$$

where b_j is a $K \times 1$ vector of factor loadings and $\epsilon_{j,t}$ is the idiosyncratic term. Rather than directly observing the true factor f_t^*, we observe f_t, which is measured with error. f_t is linked to f_t^* via a conditional distribution, that is, for $t = 1, ..., n : f_t \sim Q(.|f_t^*)$. We use a simple errors-in-variables (EIV) setup, where $f_t = f_t^* + v_t$, v_t is a $K \times 1$ vector of measurement errors, and $f_t^* \perp\!\!\!\perp v_t$. Rewriting (1.1)

$$y_{j,t} = b_j' \times (f_t - v_t) + \epsilon_{j,t} = b_j' f_t + u_{j,t} \tag{1.2}$$

where $u_{j,t} := \epsilon_{j,t} - b_j' v_t$ is the *composite error term*. In the vector form, we can rewrite (1.1):

$$
\begin{aligned}
y_j &= X^{*\prime} b_j + \epsilon_j \\
&= (X - V)' b_j + \epsilon_j = X' b_j + (\epsilon_j - V' b_j) \\
&= X' b_j + u_j
\end{aligned}
\tag{1.3}
$$

where $y_j = (y_{j,1}, ..., y_{j,n})'$ is an $n \times 1$ vector returns of asset j^{th}, $X^* = (f_1^*, ..., f_n^*)$ is a matrix of true but unobserved common factors. $X = (f_1, .., f_t ..., f_n)$ is a $K \times n$ matrix of observed but mismeasured common factors, $V = (v_1, ..., v_t, ..., v_n)$ is a $K \times n$ matrix of measurement errors, and $u_j = (u_{j,1}, ..., u_{j,t})' = \epsilon_j - V' b_j$ is an $n \times 1$ vector of composite errors for the j^{th} asset.

For each time period $t = 1, ..., n$, the returns y_t is an $p \times 1$ vector of asset returns in time t, defined as follows:

$$
y_t = B f_t + u_t \tag{1.4}
$$

where $f_t = (f_1, .., f_j ..., f_p)$ is a $K \times 1$ vector of observed common factors in time t, $u_t = \epsilon_t - B v_t = (u_{1,t} ..., u_{j,t}, ..., u_{p,t})'$ is a $p \times 1$ vector of composite errors, $\epsilon_t = (\epsilon_{1,t}, , ..., \epsilon_{p,t})'$ is a $p \times 1$ error vector, B is $p \times K$ matrix of factor loadings, and $v_t = (v_{1,t}, ..., v_{k,t})'$ is $k \times 1$ vector of measurement errors.

The $k \times k$ covariance matrix of v_t is $\Sigma_v = E[v_t v_t']$, and $p \times p$ covariance matrix of ϵ_t is $\Sigma_\epsilon = E[\epsilon_t \epsilon_t']$. Then the $p \times p$ covariance matrix of u_t is

$$
\begin{aligned}
\Sigma_n &:= E[u_t u_t'] - E[u_t] E[u_t]' \\
&= E[(\epsilon_t - B v_t)(\epsilon_t - B v_t)'] = E[\epsilon_t \epsilon_t' - B v_t \epsilon_t' - \epsilon_t v_t' B' + B v_t v_t' B'] \\
&= \Sigma_\epsilon + B \Sigma_v B'
\end{aligned}
\tag{1.5}
$$

In addition, we can express the relationship of all assets in all time periods in a more compact form:

$$
Y = BX + U \tag{1.6}
$$

where Y is a $p \times n$ matrix, B is a $p \times K$ matrix, and U is a $p \times n$ matrix of elements $u_{j,t}$. $U = \mathscr{E} + BV$ where $\mathscr{E}$ is a $p \times n$ matrix of elements $\epsilon_{j,t}$. As in Caner et al. (2023), we assume that the Eigmax(Σ_n) can grow with number of assets p, and sparsity for the precision matrix of composite errors $\Theta := \Sigma_n^{-1}$. Consider each row of Θ, for $j = 1, ..., p : \Theta_j'$ is the j^{th} row, let S_j represents the indices of the nonzero cells, for $m = 1, ..., p$

$$
S_j := \{ j : \Theta_{j,m} \neq 0 \}
$$

where $\Theta_{j,m}$ represents the j^{th} row and m^{th} element in that row. Then, we define the cardinality of the non-zero cells in the j^{th} row of precision matrix as $s_j := |S_j|$. Denote the maximum of nonzero elements across all rows $j = 1,..,p$ of the precision matrix Θ as

$$\bar{s} := \max_{1\leq j\leq p} s_j.$$

For future reference, let Y_{-j} represent all other assets except the j^{th}:

$$Y_{-j} = B_{-j}X + U_{-j} \tag{1.7}$$

where B_{-j} is a $(p-1)\times K$ submatrix of B with the j^{th} row removed, U_{-j} is a $p-1\times n$ matrix, $U_{-j} = \mathscr{E}_{-j} - B_{-j}V$.

Let $\Sigma_{n,-j,-j}$ represent a $(p-1)\times(p-1)$ subset of Σ_n with the j^{th} row and j^{th} column removed. Using the inverse formula for partitioned matrices, Yuan (2010) shows that

$$\Theta_{j,j} := (\Sigma_{n,j,j} - \Sigma_{n,j,-j}\Sigma^{-1}_{n,-j,-j}\Sigma^{-1}_{n,-j,j})^{-1} \tag{1.8}$$

and

$$\Theta'_{j,-j} := -\Theta_{j,j}\Sigma_{n,j,-j}\Sigma^{-1}_{n,-j,-j} \tag{1.9}$$

where $\Theta'_{j,-j}$ is the j^{th} row of Θ without the j^{th} element in that row.

As defined above, $u_{j,t} = \epsilon_{j,t} - b'_j v_t = \epsilon_{j,t} - B_j v_t$, $u_{-j,t}$ is the $(p-1\times 1)$ vector of composite errors in the period t without the j^{th} element in u_t. $\epsilon_{-j,t}$ is defined similarly and $u_{-j,t} = \epsilon_{-j,t} - B_{-j}v_t$. Let κ_j be the minimizing value of κ:

$$\begin{aligned} E[u_{j,t} - u'_{-j,t}\kappa]^2 &= E[(u_{j,t} - u'_{-j,t}\kappa)'(u_{j,t} - u'_{-j,t}\kappa)] \\ &= E[u^2_{j,t} - 2u_{j,t}u'_{-j,t}\kappa + \kappa' u_{-j,t}u'_{-j,t}\kappa] \end{aligned} \tag{1.10}$$

Considering each term in (1.10)

$$\begin{aligned} E[u_{j,t}u'_{-j,t}] = E[(\epsilon_{jt} - b'_j v_t)(\epsilon_{-j,t} - B_{-j}v_t)'] &= E[\epsilon_{jt}\epsilon'_{-jt} - b'_j v_t \epsilon'_{-jt} - \epsilon_{jt} v'_t B'_{-j} + b'_j v_t v'_t B'_{-j}] \\ &= E[\epsilon_{jt}\epsilon'_{-jt}] + b'_j E[v_t v'_t] B'_{-j} \\ &= \Sigma_{e,-j,j} + b'_j \Sigma_v B'_{-j} = \Sigma_{n,-j,j} \end{aligned} \tag{1.11}$$

Similarly,

$$E[u_{-jt}u'_{-jt}] = \Sigma_{e,-j,-j} + B_{-j}\Sigma_v B'_{-j} = \Sigma_{n,-j,-j} \tag{1.12}$$

and

$$E[u_{jt}^2] = E[(\epsilon_{jt} - B_j v_t)(\epsilon_{jt} - B_j v_t)'] = E[\epsilon_{jt}^2] + b'_j \Sigma_v b_j = \Sigma_{n,j,j} \tag{1.13}$$

Substituting (1.11), (1.12) and (1.13) into (1.10), and differentiating with respect to κ

$$\Sigma_{n,-j,j} - \Sigma_{n,-j,-j}\kappa = 0$$

Then we get a solution of

$$\kappa_j = \Sigma_{n,-j,-j}^{-1}\Sigma_{n,-j,j} \tag{1.14}$$

and

$$\Theta'_{j,-j} = -\Theta_{j,j}\kappa'_j \tag{1.15}$$

Defining $\eta_{j,t} := u_{j,t} - u'_{-j,t}\kappa_j$, and using the definitions of $\Sigma_{n,-j,j}$ and $\Sigma_{n,-j,-j}$, we can show that

$$\begin{aligned} E[u_{-j,t}\eta_{j,t}] &= E[u_{-j,t}u_{j,t}] - E[u_{-j,t}u'_{-j,t}\kappa_j] \\ &= \Sigma_{n,-j,j} - \Sigma_{n,-j,-j}\Sigma_{n,-j,-j}^{-1}\Sigma_{n,-j,j} = 0 \end{aligned} \tag{1.16}$$

This yields

$$u_{j,t} = u'_{-j,t}\kappa_j + \eta_{j,t} \tag{1.17}$$

We can see that $\eta_{j,t}$ and vector $u_{-j,t}$ are orthogonal for $j = 1, ..., p$ and $t = 1, ..., n$
In vector form, for each $j = 1, ..., p$, we have

$$u_j = U'_{-j}\kappa_j + \eta_j \tag{1.18}$$

where $u_j = (u_{j,1}, ..., u_{j,n})'$ is an $n \times 1$ vector, and U_{-j} is $n \times (p-1)$ matrix. Next,

$$E[u_{j,t}^2] = E[(u'_{-j,t}\kappa_j + \eta_{j,t})'(u'_{-j,t}\kappa_j + \eta_{j,t})] = \kappa'_j\Sigma_{n,-j,-j}\kappa_j + E[\eta_{j,t}^2] \tag{1.19}$$

which with (1.13) and (1.14) implies that

$$E[\eta_{jt}^2] = \Sigma_{n,j,j} - \Sigma_{n,j,-j}\Sigma_{n,-j,-j}^{-1}\Sigma_{n,-j,-j} = \frac{1}{\Theta_{j,j}} \tag{1.20}$$

Define for each $j = 1,..,p$

$$\tau_j^2 = E[\eta_{jt}^2] = \frac{1}{\Theta_{j,j}} \tag{1.21}$$

Then the diagonal terms in the precision matrix

$$\Theta_{j,j} = \frac{1}{\tau_j^2} \tag{1.22}$$

and the j^{th} row of the precision matrix without the j^{th} element, by (1.8) and (1.14) is

$$\Theta'_{j,-j} = \frac{-\kappa'_j}{\tau_j^2} \tag{1.23}$$

Define

$$\Theta = T^{-2}C \tag{1.24}$$

where

$$C := \begin{bmatrix} 1 & -\kappa_{1,2} & \dots & -\kappa_{1,p} \\ -\kappa_{2,1} & 1 & \dots & -\kappa_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ -\kappa_{p,1} & -\kappa_{p,2} & \dots & 1 \end{bmatrix}$$

and $T^{-2} := \text{diag}(\tau_1^{-2}, ..., \tau_p^{-2})$.

Use (1.18) and the sparsity assumption of the precision matrix as in Caner et al. (2023), with a sequence $\kappa_n > 0$, and for $j = 1, ..., p$

$$\tilde{\kappa}_j = \arg\min_{\kappa_j \in \mathbb{R}^{p-1}} \left[||u_j - U_j'\kappa_j||_n^2 + 2\lambda_n ||\kappa_j||_1 \right]. \tag{1.25}$$

(1.25) has a notable difference with residual-based nodewise regression in Caner et al. (2023) due to the presence of the *composite error* u_j. We now show how to turn this into feasible regression which will consistently estimate κ_j. The first step is to get consistent estimates for b_j incorrectly measured common factors. A well-established solution to the errors-in-variables problem is finding instrumental variables. Let $Z = [z_1, ...z_q]$ be a $q \times n$ matrix of instrumental variables for which $E[ZU] = 0$. Note that we allow $q > n$.

Multiplying both sides of (1.3) by Z we have:

$$Zy_j = ZX'b_j + Zu_j \tag{1.26}$$

Running OLS on equation (1.26), we have

$$\begin{aligned}
\hat{b}_j &= (XZ'ZX')^{-1}XZ'Zy_j \\
&= (XZ'ZX')^{-1}XZ'Z(X'b_j + u_j) \\
&= b_j + (XZ'ZX')^{-1}XZ'Zu_j
\end{aligned} \tag{1.27}$$

Next, define the OLS composite residual as

$$\begin{aligned}
\hat{u}_j &= y_j - X'\hat{b}_j \\
&= [X'b_j + u_j] - X'[b_j + (XZ'ZX')^{-1}XZ'Zu_j] \\
&= [I_n - X'(XZ'ZX')^{-1}XZ'Z]u_j \\
&= M_{xz}^{IV}u_j
\end{aligned} \tag{1.28}$$

where we define

$$M_{xz}^{IV} := I_n - X'(XZ'ZX')^{-1}XZ'Z$$

The estimation of $\hat{B}_{-j}$ comes from transposing equation (1.7), and multiplying both sides by Z:

$$ZY'_{-j} = ZX' \times B'_{-j} + ZU'_{-j}$$

By OLS, $\hat{B}'_{-j}$ is a $K \times (p-1)$ matrix

$$\begin{aligned}
\hat{B}'_{-j} &= (XZ'Z_jX')^{-1}XZ'ZY'_{-j} \\
&= (XZ'ZX')^{-1}XZ'Z[(X'B'_{-j} + U'_{-j}] \\
&= B'_{-j} + (XZ'ZX')^{-1}XZ'ZU'_{-j}
\end{aligned} \tag{1.29}$$

Using the transpose of (1.7), the residuals make up a $n \times (p-1)$ matrix $\hat{U}'_{-j}$

$$\begin{aligned}
\hat{U}'_{-j} &= Y'_{-j} - X'\hat{B}'_{-j} \\
&= [(X'B'_{-j} + U'_{-j}] - X'[B'_{-j} + (XZ'ZX')^{-1}XZ'ZU'_{-j}] \\
&= [I_n - X'(XZ'ZX')^{-1}XZ'Z]U'_{-j} \\
&= M^{IV}_{xz}U'_{-j} \qquad (1.30)
\end{aligned}$$

Premultiplying (1.18) by M^{IV}_{xz} and using (1.28) and (1.30),

$$\hat{u}_j = \hat{U}'_{-j}\kappa_j + \eta_{xj} \qquad (1.31)$$

where we define

$$\eta_{xj} = M^{IV}_{xz}\eta_j \qquad (1.32)$$

The feasible nodewise estimator is defined by

$$\hat{\kappa}_j = \arg\min_{\kappa_j \in \mathbb{R}^{p-1}} \left[||\hat{u}_j - \hat{U}'_{-j}\kappa_j||_n^2 + \lambda_n ||\kappa_j||_1 \right] \qquad (1.33)$$

To get $\hat{\Theta}'_j$, the j^{th} row of the estimated precision matrix of errors, we define:

$$\hat{\tau}_j^2 := \frac{\hat{u}'_j(\hat{u}_j - \hat{U}'_{-j}\hat{\kappa}_j)}{n} \qquad (1.34)$$

The j^{th} element in the j^{th} row of $\hat{\Theta}$ is

$$\hat{\Theta}_{j,j} := \frac{1}{\hat{\tau}_j^2}$$

and

$$\hat{\Theta}_{j,-j} := -\frac{1}{\hat{\tau}_j^2}\hat{\kappa}'_j$$

In the following section, we want to show that the estimator for each row $\hat{\Theta}_j$ is consistent. For convenience, we rewrite

$$\hat{\Theta}'_j = \frac{\hat{C}'_j}{\hat{\tau}_j^2} \qquad (1.35)$$

where $\hat{C}'$ is $1 \times p$ vector represents 1 in j^{th} cell, and the other cell is $-\hat{\kappa}'_j$

In a data rich environment, we assume that there are other assets, denoted by q, which are outside of our portfolio and can serve as strong instrumental variables. This idea is a similar to the OLIVE method proposed by Meng et al. (2011). Note that to address the identification issue, we need q to be larger than the number of common factors K, but p and q can be larger than n , and we allow both for q and $p \to \infty$. For the j^{th} instrument at time t, $z_{i,t}$, for $i = 1, .., q$ and time periods $t = 1, .., n$ such that:

$$z_{i,t} = \beta_i' f_t^* + e_{i,t} : \tag{1.36}$$

The main Theorem 1 shows the consistency of estimates of the precision matrix of errors and determines the rate of convergence.

1.2.2 Assumptions and Results

We now introduce the assumption required to get the consistency of the j^{th} row of the precision matrix estimator. These assumptions closely follow these in Caner et al. (2023), with additional conditions incorporated to address the use of instrumental variables.

We consider notation now. Define a as a $n \times 1$ column vector and A as a generic $d \times l$ matrix. $||a||_1$ is the l_1 norm of a, $||a||_2$ is the l_2-norm of a, and $||a||_{l_\infty}$ is the l_∞-norm of a. We define $||A||_\infty = \max_{1 \le i \le d, 1 \le j \le l} |A_{ij}|$, $||A||_{l_\infty} = \max_{1 \le i \le d} \sum_{j=1}^{l} |A_{ij}|$, $||A||_{l_1} = \max_{1 \le j \le l} \sum_{i=1}^{d} |A_{ij}|$. Let Eigmin($A$) and Eigmax($A$) represent the minimum and maximum eigenvalue of the matrix A, respectively.

Assumption 1:

(i) $\{f_t\}_{t=1}^n, \{v_t\}_{t=1}^n, \{\epsilon_t\}_{t=1}^n, \{e\}_{t=1}^n$ are strictly stationary and ergodic. Furthermore, $\{v_t\}_{t=1}^n$ and $\{\epsilon_t\}_{t=1}^n$ are independent of each other. Both v_t and ϵ_t have zero means, with variances Σ_v and Σ_ϵ, respectively, with their minimum eigenvalues bounded away from 0: Eigmin$(\Sigma_v) \ge c > 0$, and Eigmin$(\Sigma_\epsilon) \ge c > 0$, for a positive constant c. $\max_{1 \le j \le p} \mathbb{E}[\epsilon_{j,t}^2] < C < \infty$, $\max_{1 \le j \le p} \mathbb{E}[e_{j,t}^2] < C < \infty$. Finally, $\{\epsilon_t\}_{t=1}^n$ and $\{e_t\}_{t=1}^n$ are independent of each other.

(ii) For the strong mixing variables $f_t, \epsilon_t, v_t : \alpha(t) \le \exp(-C\, t^{r_0})$, for a positive constant $r_0 > 0$

Assumption 2: For all $t = 1, ..., n$ and $j = 1, ..., p$, and $k = 1, .., K$, there exist positive constants $r_1, r_2, r_3, r_4 > 0$ and another set of positive constants $b_1, b_2, b_3, b_4, s_1, s_2, s_3, s_4 > 0$ such that

(i) $P[|\epsilon_{j,t}| > s_1] \le \exp(-(s_1/b_1)^{r_1})$

(ii) $P[|\eta_{j,t}| > s_2] \le \exp(-(s_2/b_2)^{r_2})$

(iii) $P[|f_{k,t}| > s_3] \le \exp(-(s_3/b_3)^{r_3})$

(iv) $P[|v_{k,t}| > s_4] \leq \exp(-(s_4/b_4)^{r_4})$

(v) There exists $0 < \kappa_1 < 1$ such that $\kappa_1^{-1} = 3r_1^{-1} + r_0^{-1}$ and we also assume that $3r_2^{-1} + r_0^{-1} > 1$, $3r_3^{-1} + r_0^{-1} > 1$, and $3r_4^{-1} + r_0^{-1} > 1$

Define $\kappa_2^{-1} := 1.5r_1^{-1} + 1.5r_2^{-1} + r_0^{-1}$, $\kappa_3^{-1} := 1.5r_1^{-1} + 1.5r_3^{-1} + r_0^{-1}$, $\kappa_4^{-1} := 1.5r_1^{-1} + 1.5r_4^{-1} + r_0^{-1}$, and let $\kappa_{min} = \min(\kappa_1, \kappa_2, \kappa_3, \kappa_4)$

Assumption 3:

(i) $[\ln \max(p, q)]^{(2/\kappa_{min}-1)} = o(n)$

(ii) $K^2 = o(n)$

(iii) $K = o(p)$

(iv) $K = o(q)$

Assumption 4:

(i) The covariance matrix of f_t is Σ_{f_t}, such that Eigmin$[\Sigma_{f_t}] \geq c > 0$, for a positive constant c

(ii) $\max_{1\leq k\leq K} E[f_{k,t}^2] \leq c < \infty$ and $\min_{1\leq k\leq K} E[f_{k,t}^2] \geq c > 0$

(iii) $\max_{1\leq j\leq p} E[\eta_{j,t}^2] \leq c < \infty$

Assumption 5: $\bar{s}\sqrt{\frac{\max(\ln p, \ln q)}{n}} \longrightarrow 0$

Note that Assumptions 1-3 are assumptions used in Fan et al. (2011) and Caner et al. (2023). Assumption 1(a) implies cross-independent of errors used in Meng et al. (2011). In addition, Assumption 3 implies $\sqrt{\dfrac{\max(\ln p, \ln q)}{n}} = o(1)$

We now formally define λ_n. With a generic positive constant, $C > 0$

$$\lambda_n := C\sqrt{\bar{s}}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right] = o_p(1)$$

The tuning parameters λ_n involve K^3 and K^4 due to using instrumental variables in the first step and nodewise regression via factor models. We now provide consistent estimation for the rows of the precision matrix of composite errors.

Theorem 1 Under Assumption 1-5

$$||\hat{\Theta}-\Theta||_{l_\infty} := \max_{1\le j\le p}||\Theta'_j-\Theta'_j||_1 = \max_{1\le j\le p}||\Theta_j-\Theta_j||_1 = O_p(\bar{s}\lambda_n)$$

where $\hat{\Theta}_j$ and Θ_j are column representations of row vectors, $\hat{\Theta}'_j$ and Θ'_j respectively.

Proofs for this and subsequent theorems are in Appendix A.

1.3 Precision matrix estimation for asset returns

The covariance matrix of asset returns:

$$\Sigma_y := E[y_t y_t'] - E[y_t]E[y_t]' \tag{1.37}$$

From (1.1),

$$\begin{aligned}
\Sigma_y &= B\Sigma_{f_t^*}B' + \Sigma_\epsilon \\
&= B\Sigma_{f_t - v_t}B' + \Sigma_\epsilon \\
&= B\Sigma_{f_t}B' + \{\Sigma_\epsilon + B\Sigma_v B'\} \\
\Sigma_y &= B\Sigma_{f_t}B' + \Sigma_n
\end{aligned} \tag{1.38}$$

It is straightforward to show that our composite error term u_t satisfies Eigmin$(\Sigma_n) \geq C$, where C is a generic positive constant (see (A.20)). Since all conditions required for the application of the Morrison-Woodbury formula are met. By letting $\Gamma := \Sigma_y^{-1}$ and $\Theta := \Sigma_n^{-1}$, we get the precision matrix for the returns:

$$\Gamma := \Theta - \Theta B[\Sigma_{f_t}^{-1} + B'\Theta B]^{-1}B'\Theta \tag{1.39}$$

and the estimator of the precision matrix for the return is:

$$\hat{\Gamma} := \hat{\Theta} - \hat{\Theta}\hat{B}[\widehat{\Sigma_{f_t}}^{-1} + \hat{B}'\hat{\Theta}_{sym}\hat{B}]^{-1}\hat{B}'\hat{\Theta} \tag{1.40}$$

Where $\hat{\Theta}_{sym} := \frac{\hat{\Theta}+\hat{\Theta}'}{2}$ is the symmetric version of our feasible nodewise regression estimator for the precision matrix for composite errors. $\widehat{\Sigma_{f_t}} = \frac{1}{n}XX' - \frac{1}{n^2}X1_n1_n'X'$ where 1_n is $n \times 1$ column vectors of ones. Next, for convenience in analyzing each term, we rewrite the precision matrix and its estimates as follows

$$\begin{aligned}
\Gamma &= \Theta - \Theta L\Theta \\
\hat{\Gamma} &= \hat{\Theta} - \hat{\Theta}\hat{L}\hat{\Theta}
\end{aligned}$$

where

$$L := B[\Sigma_{f_t}^{-1} + B'\Theta B]^{-1}B'$$

and its estimate

$$\hat{L} := \hat{B}[\widehat{\Sigma_{f_t}}^{-1} + \hat{B}'\Theta\hat{B}]^{-1}\hat{B}'.$$

This allows us to write concisely

$$\max_{1\le j\le p} ||\hat{\Gamma}_j' - \Gamma_j'||_1 = \max_{1\le j\le p} ||(\hat{\Theta}_j' - \Theta_j') - (\hat{\Theta}_j' \hat{L} \hat{\Theta}_j' - \Theta_j' L \Theta_j')||_1 \tag{1.41}$$

In Theorem 1, we derived convergence rates for the error terms. Our objective now is to establish consistency and convergence rates for the returns precision matrix. To this end, we simplify the corresponding terms and impose the following assumption on the factor loadings.

Assumption 6: The factor loadings are assumed to satisfy the following conditions:

1. $\max_{1\le j\le p} \max_{1\le k\le K} |b_{jk}| \le C < \infty$, where $C > 0$ is a positive constant
2. $\left|\left|\frac{B'B}{p} - \Delta\right|\right|_{l_2} = o(1)$ for a $K \times K$ symmetric positive definite matrix Δ: Eigmin$(\Delta) \ge C > 0$. This assumption implies Eigmin$(\frac{B'B}{P}) \ge C > 0$

Assumption 7:

1. Eigmax(Σ_n) = Eigmax$(\Sigma_\epsilon + B\Sigma_v B') \le C r_n$, with $C > 0$, a positive constant, and $r_n \to \infty$ as $n \to \infty$, and $r_n/p \to 0$, and r_n is a positive sequence.
2. Define $l_n = r_n^2 K^{5/2} \bar{s}^{3/2} \left[K^3 \sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4 \sqrt{\frac{\ln n}{n}}\right]$, then $\bar{s} l_n \to 0$

Let $\mu := E[y]$, and $\hat{\mu} = \frac{1}{n}\sum_{t=1}^n y_t$, where y_t is $p \times 1$ vector of asset returns. Theorem 2 is the main theorem of the chapter, representing the consistent estimation of the precision matrix for asset returns.

Theorem 2

(i) Under Assumptions 1-4, and 6-7

$$\max_{1\le j\le p} ||\hat{\Gamma}_j - \Gamma_j||_1 = O_p(\bar{s} l_n) = o_p(1)$$

where $\hat{\Gamma}_j$ and Γ_j are column representations of row vectors $\hat{\Gamma}_j'$ and Γ_j' respectively.

(ii) Under Assumptions 1-6

$$||\hat{\mu} - \mu||_\infty = O_p\left[\max\left(K\sqrt{\frac{\ln n}{n}}, \sqrt{\frac{\ln p}{n}}\right)\right] = o_p(1)$$

1.3.1 A Practical Algorithm for Estimating the Precision Matrix

Given p-large number of assets and q other assets serving as instrumental variables, we provide a practical algorithm to get the precision matrix estimator for asset returns $\hat{\Gamma}$, using instrumental variables and the residual-based nodewise regression estimator $\hat{\Theta}$

1. Use equation (1.27) to form $\hat{b}_j$:

$$\hat{b}_j = (XZ'ZX')^{-1}XZ'Zy_j$$

   Where $X = (f_1, ..., f_t, ..., f_n) : K \times n$ matrix where $f_t : K \times 1$ are observed but mismeasured factors, y_j as the j^{th} asset returns $(n \times 1)$ $Z = (z_1, ..., z_t, ..., z_n) : q \times n$ matrix, where $z_t : q \times 1$ are others assets of returns, outside of the portfolio, serve as instrumental variables.

2. Use equation (1.28) to get the residual from a least squares based regression

$$\hat{u}_j = y_j - X'\hat{b}_j$$

3. Use equation (1.29) to get estimates of the transpose of factor loadings, $\hat{B}'_{-j}$, which is a $K \times p-1$ matrix.

$$\hat{B}'_{-j} = (XZ'ZX')^{-1}XZ'ZY_{-j}$$

   where $Y_{-j} : n \times (p-1)$ is the transpose matrix of asset return except the j^{th} asset.

4. Use equation (1.30) to form the transpose matrix of residuals for all asset returns except the j^{th}, $\hat{U}'_{-j}$, which is a $n \times (p-1)$ matrix

$$\hat{U}'_{-j} = Y'_{-j} - X'\hat{B}'_{-j} \tag{1.42}$$

5. Run nodewise regressions of $\hat{u}_j$ on $\hat{U}'_{-j}$ via lasso, and get λ_n from Generalized Information Criterion (GIC)

6. Use equation (1.34) to get $\hat{\tau}^2_j$, then from $\hat{\Theta}'_j$ as in (1.35)

7. Repeat steps 1-5, for $j = 1, ...p$, to get $\hat{\Theta}$

8. Form

$$\hat{B} = (XZ'ZX')^{-1}XZ'ZY$$

which is a $p \times K$ matrix of OLS estimates, where $Y : p \times n$ the matrix of all asset returns $j = 1, \ldots p$ represents all the columns of assets. Also, the covariance matrix estimate for factors

$$\widehat{\Sigma_{f_t}} = \frac{1}{n} X X' - \frac{1}{n^2} X 1_n 1'_n X'$$

where 1_n is a $n \times 1$ column vector of ones.

9. Finally, the precision matrix estimate for all asset returns

$$\hat{\Gamma} = \hat{\Theta} - \hat{\Theta} \hat{B} \left[\widehat{\Sigma_{f_t}}^{-1} + \hat{B}' \frac{\hat{\Theta} + \hat{\Theta}'}{2} \hat{B} \right]^{-1} \hat{B}' \hat{\Theta}$$

1.4 Sharpe Ratio Analysis

In this section, we provide Sharpe Ratio metrics that use the precision matrix of asset returns. These metrics have been discussed heavily in Caner et al. (2023). Since the proofs would take much space to analyze, we include only two illustrative examples. Specifically, we include the squared Sharpe Ratio metrics derived from the Global Minimum Variance (GMV) and the Markowitz Mean-Variance (MKW) portfolio.

Assumption 8: Let c be a positive constant. We impose the following assumptions:

(i) $K^3 \bar{s} l_n = o(1)$

(ii) $\frac{\mu' \Gamma \mu}{p} \geq c > 0, \frac{1_p \Gamma 1_p}{p} \geq c > 0$

Note that we replace Assumption 7(ii) with Assumption 8(i), which implies that the variance of the GMV portfolio, $\left[\frac{1_p' \Gamma 1_p}{p}\right]^{-1}$ is finite.

Global Minimum-Variance (GMV) Portfolio

To obtain the GMV, we minimize the variance of the portfolio. The portfolio is chosen according to a weight vector w whose elements sum to one. Short sales are allowed.

$$w_{GMV} = \arg\min_{w \in \mathbb{R}^p} w' \Sigma_y w, \text{subjecto} \, w' \mathbf{1}_p = 1$$

The solution is well known and given by:

$$w_{GMV} = \frac{\Sigma_y^{-1} \mathbf{1}_p}{\mathbf{1}_p' \Sigma_y^{-1} \mathbf{1}_p}$$

Substituting these weights into the Sharpe Ratio formula and normalized by the number of assets, we obtain:

$$SR = \frac{w_{GMV}' \mu}{\sqrt{w_{GMV}' \Sigma_y w_{GMV}}} = \sqrt{p} \left(\frac{\mathbf{1}_p' \Sigma_y^{-1} \mu}{p} \right) \left(\frac{1_p' \Sigma_y^{-1} \mathbf{1}_p}{p} \right)^{-1/2} \tag{1.43}$$

The estimator is

$$\widehat{SR}_{GMV} = \sqrt{p}\left(\frac{\mathbf{1}'_p\hat{\Gamma}\hat{\mu}}{p}\right)\left(\frac{\mathbf{1}'_p\hat{\Gamma}\mathbf{1}_p}{p}\right)^{-1/2}$$

The following theorem establishes the consistency of the Sharpe ratio in the GMV portfolio.

Theorem 3 Under Assumption 1-4, 6,7(i) and 8(ii) with $\left|\frac{1'_p\Gamma\mu}{p}\right| \geq C > 0$

$$\left|\frac{\widehat{SR}^2_{GMV}}{SR^2} - 1\right| = O_p(K^{3/2}\bar{s}\,l_n) = o_p(1)$$

Markowitz Mean-variance (MKW) Portfolio

In this setup, the goal is to find the portfolio with the smallest variance that achieves the expected return ρ_1.

$$w_{mkw} = \arg\min_{w\in\mathbb{R}^p}(w'\Sigma_y w), \text{ subjet to } w'1_p = 1, \text{ and } w'\mu = \rho_1$$

The formula for optional weight is

$$w_{mkw} := \left[\frac{G-\rho_1 F}{AG-F^2}\right]\left(\frac{\Sigma_y^{-1}1_p}{p}\right) + \left[\frac{\rho_1 A - F}{AD-F^2}\right]\left(\frac{\Sigma_y^{-1}\mu}{p}\right) \tag{1.44}$$

where the formulas for A, G and F are $A := \frac{1'_p\Gamma 1_p}{p}$, $G := \frac{\mu'\Gamma\mu}{p}$, and $F := \frac{1'_p\Gamma\mu}{p}$. Then, the estimators of these terms as $\hat{A} := \frac{1'_p\hat{\Gamma}1_p}{p}$, $\hat{G} := \frac{\hat{\mu}'\hat{\Gamma}\hat{\mu}}{p}$, and $\hat{F} := \frac{1'_p\hat{\Gamma}\hat{\mu}}{p}$.

We obtain the Sharp Ratio:

$$SR_{mkw} = \rho_1\sqrt{p\left(\frac{AG-F^2}{A\rho_1^2 - 2F\rho_1 + G}\right)} \tag{1.45}$$

The estimator is:

$$\widehat{SR}_{mkw} = \rho_1\sqrt{p\left(\frac{\hat{A}\hat{G}-\hat{F}^2}{\hat{A}\rho_1^2 - 2\hat{F}\rho_1 + \hat{G}}\right)} \tag{1.46}$$

Theorem 4: Under Assumption 1-4, 6, 7(i), and 8 with a condition $\left|\frac{1'_p\mu}{p}\right| \geq C > 0$, $AG - F^2 \geq C_1 > 0$

and $A\rho_1^2 - 2F\rho_1 + G \geq C_1 > 0$ with ρ_1 uniformly bounded away from zero and infinity, we have

$$\left| \frac{\widehat{SR}_{mkw}^2}{SR_{mkw}^2} - 1 \right| = O_p(K^3 \bar{s} l_n) = o_p(1)$$

Since we use the same assumptions as in Caner et al. (2023), the proofs of Theorem 3 and 4 follow the same procedure as in Caner et al. (2023)

1.5 Simulations

We aim to address two main questions through simulations. First, as described in Theorem 3 and 4, we investigate whether the estimation errors of the Sharpe Ratio (squared) using our methods decrease when $\gamma = p/n$ is fixed and both p and n jointly increase. We have two experiments for simulation: (1) where $p = n/2$, which is low dimension (LD), and (2) where $p = 3n/2$ which is high dimension (HD). Secondly, we compare the performance of our new methods with existing methods. We set the number of instrumental variables, denoted by q, either equal to the number of factors or to the number of assets. Both approaches yield very similar results; therefore, we report only the former case. The results of the latter case can be provided upon request.

1.5.1 Set up

Similar to Caner et al. (2023), for $t = 1, ..., n$, we defined the observed factors with measurement errors

$$f_t = f_t^* + v_t$$

where f_t^* is a $K \times 1$ vector of the unobserved true factors and v_t is a $K \times 1$ vector of measurement error. $f_t : K \times 1$ are drawn from $\mathcal{N}(\mu_f, \Sigma_f)$, where μ_f is a $K \times 1$ vector with each element equal to 0.5, and Σ_f is the factor covariance matrix with diagonal elements set to 1, and all off-diagonals set to 0.25. The measurement errors are drawn from $\mathcal{N}(0, \Sigma_v)$, the covariance matrix of measurement error is the diagonal matrix with all elements set to $1/100$. Then, asset returns are defined as:

$$y_t = \mu_t + B_t f_t + \epsilon_t$$

where μ_t is drawn independently from $\mathcal{N}(0.5, 1)$. The vectors $B_{t,j}$ of factor loadings for assets $j = 1, ..., p$ is drawn from $\mathcal{N}(0.5_K, I_K)$, and the matrix of loadings in time t is given by $B_t = (B_{t,1}, B_{t,2}, ..., B_{t,p})^T$. For simplicity, we set the number of factors as $K = 3$[1], and the idiosyncratic errors $\epsilon_t \sim \mathcal{N}(0_p, \Sigma_\epsilon)$ with Σ_ϵ has Toeplitz form, as in Callot et al. (2021), such that the $(i, j)^{th}$ element of Σ_ϵ is $0.25^{|i-j|}$. To generate the instrument variables z_t, for $t = 1, .., n$

$$z_t = \mu_t + B_t f_t^*$$

where μ_t and f_t^* are generated the same as in y_t.

[1]the results when $K = 5$ are similar. Additional details are available upon request

1.5.2 Metrics

We consider the metric of estimation error for the Sharpe Ratio (squared) in four different portfolios, including the Global Minimum Variance Portfolio (GMV), the Markowtiz-Mean Variance (MKW) portfolio, the maximum Sharpe Ratio portfolio (MSR), and the maximum out-of-sample Sharpe Ratio (OOS-MSR). The target return in the MKW portfolio is $\rho = 0.8\%$. We use generalized information criterion (GIC) as in Caner et al. (2023) to choose the regularization parameter λ.

The model comparisons include several approaches from the literature. The residual-based nodewise from Caner et al. (2023) is indicated by rbNW. We compare this to the POET method from Fan et al. (2013), the nonlinear shrinkage method of Ledoit and Wolf (2017)(NLS) as well as their single factor nonlinear shrinkage method (SFNLS). Our new proposed residual-based instrumental variables nodewise method is indicated as rbIV.

1.5.3 Results

Tables (1.1), (1.2), and (1.3) show the simulation results. The values in each cell show the average absolute estimation error for estimating the squared Sharpe Ratio. In each table, the first four rows are for $p = 3n/2$, and the next four columns are for $p = n/2$. The tables show the results for various sample sizes $n = 100, 200, 400$, with three common factors, $K = 3$.

Table 1.1: Estimation Error, $K = 3$, $n = 100$

	rbNW	POET	NLS	SFNLS	rbIV
MSR-HD	0.0508	**0.4453**	0.0695	0.0654	0.0508
OOS-MSR-HD	0.2105	0.6643	0.1895	0.2969	**0.2100**
GMV-SR-HD	1.0506	**0.2759**	0.8664	0.8813	1.0470
MKW-SR-HD	0.8815	**0.0905**	0.7400	0.7336	0.8799
MSR-LD	0.0777	0.4292	0.0774	0.0917	**0.0776**
OOS-MSR-LD	0.1692	0.6055	**0.1429**	0.2283	0.1681
GMV-SR-LD	0.6477	0.4922	**0.4733**	0.5595	0.6395
MKW-SR-LD	0.4705	**0.1796**	0.3302	0.2745	0.4683

Table 1.2: Estimation Error, $K = 3$, $n = 200$

	rbNW	POET	NLS	SFNLS	rbIV
MSR-HD	**0.0319**	0.4787	0.0639	0.0587	0.0320
OOS-MSR-HD	0.1751	0.6754	0.1731	0.2755	**0.1749**
GMV-SR-HD	1.0089	**0.1827**	0.9383	0.9084	1.0080
MKW-SR-HD	0.8470	**0.0545**	0.8050	0.7688	0.8467
MSR-LD	**0.0408**	0.4705	0.0640	0.0624	0.0409
OOS-MSR-LD	0.1303	0.5784	**0.1263**	0.2048	0.1303
GMV-SR-LD	0.6098	**0.2735**	0.5055	0.5233	0.6086
MKW-SR-LD	0.4459	**0.2003**	0.3813	0.3292	0.4457

Table 1.3: Estimation Error, $K = 3$, $n = 400$

	rbNW	POET	NLS	SFNLS	rbIV
MSR-HD	0.0184	0.4969	0.0832	0.0505	**0.0184**
OOS-MSR-HD	0.1233	0.6521	0.1655	0.2691	**0.1233**
GMV-SR-HD	0.9035	**0.1086**	1.0003	0.9157	0.9034
MKW-SR-HD	0.7776	**0.0438**	0.8612	0.7808	0.7776
MSR-LD	**0.0292**	0.4967	0.0597	0.0503	0.0293
OOS-MSR-LD	0.0804	0.5945	0.1167	0.2022	**0.0804**
GMV-SR-LD	0.4497	**0.1886**	0.4616	0.3952	0.4497
MKW-SR-LD	0.3852	**0.2229**	0.3955	0.3385	0.3852

First, the tables show that our method achieves low- and high-dimensional consistency, as stated in our theorems. Analyzing Table (1.1) with $K = 3$ and MSR-HD (Maximum Sharp Ratio in high dimensional where $p = 3n/2$), we observe that our rbIV has an estimation error at 0.0508 with $n = 100$, and this error declines to 0.0320 with $n = 200$, and then declines to 0.0184 with $n = 400$. To give another example, analyzing the OOS-MSR (the out-of-sample - Maximum Sharpe Ratio portfolio) with our method rbIV, in low-dimensional set up, the estimation error is 0.1681 at $p = n/2, n = 100$, then this error declines to 0.1303 at $p = n/2, n = 200$, then further declines to 0.0804 at $p = n/2, n = 400$. This indicates that as both n and p increase jointly, the estimation error decreases.

Next, we consider which method achieves the smallest estimation errors. The results are mixed for $n = 100$ and $n = 200$, but our rbIV dominates other techniques at high-dimension for OOS-MSR.

1.6 Empirical Application

The objective of the empirical application is to conduct an out-of-sample comparison among estimations. We estimate the GMV and MKW portfolios using our own and others' estimators and four commonly used financial metrics: the Sharpe Ratio (SR), average return, portfolio turnover, and portfolio variance. Similar to Caner et al. (2023) we set the return constraint ρ to 0.8% per month in MKW portfolio. We use the monthly data from January 1994 to May 2018 with out-of-sample data from January 2004 to May 2018. We selected all stocks in the S&P 500 for at least one month in the out-of-sample period and have data for the entire 1994-2018 period, resulting in 304 stocks. The out-of-sample period includes the recession of 2008. Then, we randomly selected 50 stocks from the pool that served as instrumental variables. The main focus of this application is predictions using a large portfolio, so we use all the stocks available.

1.6.1 Metrics

We apply a rolling window approach to estimate the out-of-sample Sharpe Ratio, following the procedures in Callot et al. (2021) and Caner et al. (2023). We split our total time period sample size n into an in-sample and an out-of-sample, indexed as $(1 : n_I)$ and $(n_I : n)$, respectively. Initially, we use the in-sample data to estimate the weights $\hat{w}_{n_I}$ and then compute out-of-sample returns $\hat{w}'_{n_I} y_{n_I+1}$. Then, the window is rolled forward by one to $(2 : n_I + 1)$ form a new in-sample $\hat{w}_{n_I+1}$ and calculate corresponding out-of-sample returns $\hat{w}'_{n_{I+1}} y_{n_I+2}$. We repeat this process until we reach the end of the sample.

Without transaction cost, the average return and variance for out-of-sample data are computed as follows:

$$\hat{\mu}_{os} = \frac{1}{n - n_I} \sum_{t=n_I}^{n-1} \hat{w}'_t y_{t+1} \text{ and } \hat{\sigma}^2_{y,os} = \frac{1}{n - n_I - 1} \sum_{t=n_I}^{n-1} (\hat{w}'_t y_{t+1} - \hat{\mu}_{os})$$

Sharpe Ratios are estimated as follows:

$$SR = \frac{\hat{\mu}_{os}}{\hat{\sigma}_{y,os}} \tag{1.47}$$

When transaction costs are present, the definition of the metrics is modified. Let c be the transaction cost, which we set at 10 basis points. Let $y_{P,t+1} = \hat{w}_t' y_{t+1}$ be the return of the portfolio P in period $t+1$; in the presence of transaction costs, the return is defined as

$$y_{P,t+1}^{Net} = y_{P,t+1} - c(1 + y_{P,t+1}) \sum_{j=1}^{p} |\hat{w}_{j,t+1} - \hat{w}_{j,t}^{+}|$$

where $\hat{w}_{j,t}^{+} = \frac{\hat{w}_{j,t}(1 + y_{j,t+1})}{1 + y_{P,t+1}}$, with $y_{j,t+1}$ representing the excess return added to the risk-free rate for the j^{th} asset, and $y_{P,t+1}$ representing the portfolio P excess return added to the risk-free rate. The adjustment made in $\hat{w}_{j,t}^{+}$ is because the portfolio at the end of the period has changed compared to the portfolio at the beginning of the period. Then, we define the mean and variance of the out-of-sample portfolio returns with transaction costs as

$$\hat{\mu}_{os-c} = \frac{1}{n - n_I} \sum_{t=n_I}^{n-1} y_{P,t}^{Net} \text{ and } \hat{\sigma}_{y,os-c}^2 = \frac{1}{n - n_I - 1} \sum_{t=n_I}^{n-1} (y_{P,t}^{Net} - \hat{\mu}_{os-c})^2$$

and the Sharp Ratio with transaction costs as computed as:

$$SR_c = \frac{\hat{\mu}_{os-c}}{\hat{\sigma}_{y,os-c}} \tag{1.48}$$

1.6.2 Methods

The model comparisons include several approaches from the literature. We consider the nonlinear shrinkage methods of Ledoit and Wolf (2017) (NLS) and the single-factor nonlinear shrinkage (SFNLS), the POET method from Fan et al. (2013), along with the feasible residual-based nodewise (rbNW) approach from Caner et al. (2023). The rbNW is calculated separately using a single-factor model (market as the only factor), three-factor, and five-factor Fama and French model. We also include an equally weighted (EW) portfolio for comparison. Our new method is rbIV, and we calculate this using one, three or five factors. The main difference between rbIV and rbNW is that the rbIV method uses q instrumental variables to get the factor loadings in the first step, as described in section (1.6.1). The number of other assets (q) serving as instrumental variables is greater than the number of factors k, representing over-identification. To assess the validity of all instrumental variables, we employ the Sargan-Hansen J statistic across various cases—Fama/French factors 1 factor, 3 factors, and 5 factors. The null hypothesis is: *The set of instruments is valid;* in the example below, we apply the Sargan-Hansen J statistic for Fama and French 5 factors, using 50 randomly selected instrumental variables. The null hypothesis is not rejected, it implies that all our instruments are valid.

1.6.3 Results

We report the results for both Global Minimum Variance and Markowitz Sharpe Ratio portfolios, without transaction cost in Table 1.4 and with transaction costs in Table 1.5. The in-sample $n_I = 120$ and out-of-sample period $n - n_I = 173$. In addition, we use the same data but with different out-of-sample from January 2008 to May 2018 ($n - n_I = 155$). The sub-sample yields similar results and are included in the Appendix A.

We compare two portfolios based on the Sharpe Ratio under transaction costs in Table 1.5. Our rbIV method outperforms other techniques, yielding the highest Sharpe Ratio of 0.270 within the Global Minimum Variance (GMV) Portfolio when utilizing the five Fama and French factors. The rbNW, Caner et al. (2023), with only a single factor ranks second at 0.251.

Furthermore, while Caner et al. (2023) shows that a single-factor model can maximize the Sharpe Ratio, our results indicate that incorporating additional factors enhances performance. For instance, in the GMV portfolio, our rbIV approach achieves a Sharpe Ratio of 0.248 with three factors, which further increases to 0.270 when using a five-factor model.

Table 1.4: Out-of-Sample Empirical Results (Jan 2004 – May 2018) Without Transaction Costs

	Global Minimum Portfolio			**Markowitz Portfolio**		
	Sharpe	Return	SD	Sharpe	Return	SD
EW	0.253	0.011	0.043			
SFNLS	0.242	0.007	0.029	0.236	0.007	0.029
NLS	0.249	0.007	0.029	0.245	0.007	0.029
POET	0.239	0.007	0.030	0.246	0.007	0.030
1 Factor						
rbNW	0.251	0.008	0.031	0.251	0.008	0.031
3 Factors						
rbIV	0.253	0.007	**0.029**	0.256	0.007	0.029
rbNW	0.249	0.008	0.031	0.254	0.008	0.030
5 Factors						
rbIV	**0.276**	0.008	**0.029**	0.274	0.008	0.029
rbNW	0.246	0.007	0.030	0.251	0.007	0.030

Table 1.5: Out-of-Sample Empirical Results (Jan 2004 – May 2018) With Transaction Costs

	Global Minimum			**Markowitz Portfolio**		
	Sharpe	Return	SD	Sharpe	Return	SD
EW	0.250	0.011	0.043			
SFNLS	0.237	0.007	0.029	0.236	0.007	0.029
NLS	0.245	0.007	0.029	0.245	0.007	0.029
POET	0.239	0.007	0.030	0.246	0.007	0.030
1 Factor						
rbNW	0.251	0.008	0.031	0.251	0.008	0.031
3 Factors						
rbIV	0.248	0.007	0.029	0.250	0.007	0.029
rbNW	0.246	0.008	0.031	0.251	0.008	0.030
5 Factors						
rbIV	**0.270**	0.008	**0.029**	0.266	0.008	0.029
rbNW	0.242	0.007	0.030	0.248	0.007	0.030

CHAPTER

2

ONLINE INVESTING

2.1 Introduction

In this chapter, we study "online" investing – a sequential decision-making process in which an investor can change their portfolio over time [1]– as a practical application of *online convex optimization* (*OCO*). To illustrate the main problem in online investing, consider the following example. An online investor wants to distribute his wealth among a set of assets without prior knowledge of market outcomes. The objective is to maximize the increase in his wealth over a sequence of many trading periods. This problem can be viewed as an instance of *online convex optimization* (*OCO*). We formally define *OCO* as follows. For each trading period $t = 1, \ldots, T$, an online investor chooses a point w_t in a convex set, Δ. After committing to this choice, a convex loss function $f_t : \Delta \to \mathbb{R}$ is revealed. To measure the performance of the online investor, we would like to evaluate how the player's performance is constrained by a lack of foresight. The most common way is to measure the performance of the online investor versus a best fixed strategy, chosen with the benefit of hindsight. That is, the investor has complete knowledge of the future market outcomes but is restricted to choosing a fixed strategy for all iterations. This performance measure is called *Regret* in Cover (1991) and Hazan et al. (2007), formally defined as:

[1]In contrast, "offline" investing does not allow portfolio rebalancing

$$Regret_T = \max_{w^* \in \Delta_p} \sum_{t=1}^{T} f_t(w^*) - \sum_{t=1}^{T} f_t(w_t) \tag{2.1}$$

At the beginning of period t, the online investor needs to distribute his wealth w_t among p assets [2] and probability simplex $\Delta = \{w \in \mathbb{R}_+ : \sum_{i=1}^{p} w_i = 1\}$, w^* is the best fixed strategy in hindsight. One desired, but not necessarily feasible, goal for online portfolio selection is *universality*, when the average of *Regret* for T trading periods asymptotically approaches 0. In other words, we would like regret to be sublinear as in Cover (1991) and in Hazan et al. (2007), since this implies that "on average" the algorithm performs as well as the best fixed strategy in hindsight.

Cover (1991) was the first to introduce Universal Portfolios (UP), the first algorithm that can shown to be competitive with the best constant-rebalanced portfolio (CRP) in hindsight. In the CRP strategy, the investor always rebalances their wealth to ensure a fixed proportion of investment for each stock. The key idea behind UP is to maintain a distribution over all CRP and perform a Bayesian update after observing investing every trading period. The CRP strategy is only effective for small portfolios since the bound of *Regret* depends on the number of assets. Our first contribution is to provide an approach that achieves sublinear *Regret* both for small and very large portfolios. In contrast to the standard algorithm in the literature, which is based on the Newton method as described in Hazan et al. (2007), our regret bounds are derived using the l_∞-norm rather than the l_2-norm, making the approach particularly suitable for large portfolios and high-dimensional problems.

The organization of this chapter is as follows. Section 2.2 introduces the model setup and presents the algorithm that achieves sublinear *Regret*. Section 2.3 presents the simulation. Section 2.4 provides the empirical application.

2.2 Model

2.2.1 Set up

Let p be the number of assets in an example portfolio. On every trading period t, for $t = 1, ..., T$, the investor distributes their wealth among a set of p assets without knowing the market outcome in advance, such that $w_t \in \Delta_p$ and $\Delta_p = \{w \in \mathbb{R}^p, w_t' 1_p = 1\}$. Short sales are allowed[3]. Then, $w_{j,t}$ is the weight for the j^{th} asset at period t. Define the asset's return at time $t+1$

[2] without knowing the market outcomes at period t

[3] This provision is more realistic than Cover (1991) and Hazan et al. (2007) $w \in \mathbb{R}_+^p$

$$\mu_{j,t+1} = \frac{\text{price of asset } j \text{ at time } t+1 - \text{price of asset } j \text{ at time } t}{\text{price of asset } j \text{ at time } t} - 1. \tag{2.2}$$

The *no-junk-bond* assumption introduced in Hazan et al. (2007) assumes that for all $j = 1, .., p$ and $t = 1, ..., T$, price of asset j at time t is strictly positive. The investor begins with an initial wealth W_0, and his wealth at the time period t is denoted by W_t. Without transaction costs, at the period $t+1$, define his wealth as:

$$W_{t+1} = W_t[(1_p + \mu_{t+1})' w_{t+1}] \tag{2.3}$$

Over T trading periods, the *cumulative wealth* of the investor is

$$W_t = W_0 \left[\prod_{t=1}^{T} (1_p + \mu_t)' w_t \right] \tag{2.4}$$

where W_0 is the initial wealth of the investor. The goal of investors is to maximize the *cumulative wealth*, or the ratio of W_t / W_0 . This can be attained by maximizing the logarithmic growth ratio as in Cover (1991). Rewrite(2.4)

$$\log \frac{W_t}{W_0} = \sum_{t=1}^{T} \log[(\mu_t + 1_p)' w_t] \tag{2.5}$$

The above problem can be viewed as an Online Convex Optimization (*OCO*), where the convex function as in Cover (1991) and in Hazan et al. (2007) is

$$f_t(w_t) = -\log[(\mu_t + 1_p)' w_t] \tag{2.6}$$

It is clear that the function above meets both the first-order and second-order conditions for convexity. Define *Regret* as the gap between the total loss of a strategy w_t and the best fixed weights w^* in hindsight.

$$Regret_T = \max_{w^* \in \Delta_p} \sum_{t=1}^{T} f_t(w^*) - \sum_{t=1}^{T} f_t(w_t) \tag{2.7}$$

Theorem 1: Regret is sublinear, that is, $Regret_T \leq o_p(T)$

Let us begin by considering a differentiable convex function $f : \mathbb{R}^p \to \mathbb{R}$. Then, for any two points w^* and $w_t \in \Delta_p$ we have

$$f(w_t) \geq f(w^*) + \nabla f(w^*)(w^* - w_t) \tag{2.8}$$

Combining (2.7), we get:

$$\begin{aligned} Regret_T &= \sum_{t=1}^{T} [f_t(w^*) - f_t(w_t)] \\ &\leq \sum_{t=1}^{T} \nabla f_t(w^*)^T [w^* - w_t] \\ &\leq \sum_{t=1}^{T} ||\nabla f_t(w^*)||_\infty ||w^* - w_t||_1 \end{aligned} \tag{2.9}$$

where we use Hölder's Inequality for the first inequality. As in Hazan et al. (2007), for some generic positive constant C:

$$\max_{1\leq j\leq p} \max_{1\leq t\leq T} |\mu_{j,t}| \leq C \tag{2.10}$$

Now we can get the bound:

$$\begin{aligned} ||\nabla f_t(w)||_\infty &= \max_{1\leq j\leq p} \max_{1\leq t\leq T} \left| \frac{1_p + \mu_{j,t}}{(1_p + \mu_t)^T w_t} \right| \\ &= \frac{\max_{1\leq j\leq p} \max_{1\leq t\leq T} |1 + \mu_{j,t}|}{\min_{1\leq t\leq T} |(1_p + \mu_t)^T w_t|} \\ &= \frac{\max_{1\leq j\leq p} \max_{1\leq t\leq T} |1 + \mu_{j,t}|}{\min_{1\leq t\leq T} |\sum_{j=1}^{p} w_{j,t}(1 + \mu_{j,t})|} \\ &\leq \frac{\max_{1\leq j\leq p} \max_{1\leq t\leq T} |1 + \mu_{j,t}|}{\min_{1\leq t\leq T} \min_{1\leq t\leq T}(1 + \mu_{j,t})| \sum_{j=1}^{p} w_{j,t}|} \\ &\leq \frac{\max_{1\leq t\leq T} \max_{1\leq j\leq p} |1 + \mu_{t,j}|}{\min_{1\leq t\leq T} \min_{1\leq j\leq p} |1 + \mu_{t,j}|} = \rho \leq C \end{aligned} \tag{2.11}$$

Combining with(2.9), we have:

$$Regret_T \leq \rho \sum_{t=1}^{T} ||w_t - w^*||_1 \tag{2.12}$$

We now observe that $Regret_T$ depends on the rate $||w_t - w^*||_1$. The estimated weights of the portfolio are discussed in Callot et al. (2021), but not in Caner et al. (2023). In Caner et al. (2023), the main focus is on getting consistent estimators of the Sharp Ratio in different scenarios. We will prove the consistency of $Regret_T$ when the portfolio weights are estimated. The proof is not straightforward and involves several steps.

Our second contribution is to derive the rate of $||w_t - w^*||_1$. We derive the estimation of weights for both the Global Minimum Variance (GMV) and Markowitz portfolios. The next two main theorems will provide the consistency of portfolio weights.

2.2.2 Portfolio Weights Estimation

In this subsection, we established the properties of the estimated weights of portfolios based on the residual-based nodes method in Caner et al. (2023). However, since the primary focus of this chapter is on studying the properties of *Regret*, the proofs for Theorem 2 and Theorem 3 are provided in Appendix B.

Global Minimum Variance (GMV)

We choose the weights to minimize the variance of the portfolio subject to the weights summing to one. Let Σ_y be the covariance matrix of $p-$assets returns:

$$w_{gmv} = \arg\min_{w \in \mathbb{R}^p} w'\Sigma_y w, \text{ subject to } w'\mathbf{1}_p = 1$$

The optimal weights are given by:

$$w^*_{gmv} = \frac{\Gamma 1_p}{1'_p \Gamma 1_p} \tag{2.13}$$

and estimated as:

$$w_{gmv} = \frac{\hat{\Gamma} 1_p}{1'_p \hat{\Gamma} 1_p} \tag{2.14}$$

where $\Gamma = \Sigma_y^{-1}$

Theorem 1: Based on Caner et al. (2023), under Assumptions 1-4, 7, and 8 with $\left|\frac{1'_p \Gamma \mu}{p}\right| \geq c$

$$||w_{gmv} - w^*||_1 = O_p(\bar{s}^2 l_n r_n K^{3/2}) = o_p(1) \tag{2.15}$$

Markowitz Mean Variance portfolio

The optimal weights of the Markowitz Mean-Variance (MKW) is given by, as described in Chapter 1.4 or in Caner et al. (2023) and Callot et al. (2021).

$$w_{mk} = \arg\min_{w \in \mathbb{R}^p} (w' \Sigma_y w) \text{ such that } w' 1_p = 1, \text{ and } w'\mu = \rho_1$$

Theorem 2: Under Assumptions 1-4, 6, 7 (i) in Caner et al. (2023)

$$||w_{mk} - w^*||_1 = O(K^2 \bar{s}^2 l_n^2) = o_p(1) \tag{2.16}$$

Combine (2.12) with Theorem 1 or Theorem 2, we get:

$$\frac{Regret_T}{T} = o_p(1) \tag{2.17}$$

2.3 Simulation

2.3.1 Setup

The main theorem depends on the rate of weight estimation error. Therefore, we investigate whether the estimation errors of weights using residual-based nodewise regression decreases when $\gamma = p/n$ is fixed and both p and n jointly increase. We have two experiments for simulation: (1) $p = n/2$, which is low dimension (LD), and (2) $p = 3n/2$ which is high dimension (HD). The GDP for $= 1, ..., p$ is

$$y_j = \alpha_j + \sum_{k=1}^{K} \beta_{j,k} f_k + e_j \tag{2.18}$$

where y_j is the return of asset j^{th}, and f_k is common factor k. $\beta_{j,k}$ is the individual stock sensitivity to the factor. Specifically, $\beta_{j,k}$ is drawn from $\mathcal{N}(\mu_\beta, \Sigma_\beta)$, when μ_β is a $K \times 1$ vector with each element equal to 0.5, and Σ_β is the factor covariance matrix with diagonal set to 1 and off-diagonal set to 0.25. f_k is drawn independently from $\mathcal{N}(0.5, 1)$. The error term e_j is drawn from $\mathcal{N}(0, \Sigma_u)$, where Σ_u is a $p \times p$ matrix in the *Toeplitz*(ρ) form. The entry in the i^{th} row and j^{th} column of Σ_u is defined as

$$[\Sigma_u]_{i,j} := \rho^{|i-j|} \tag{2.19}$$

where $\rho = 0.25, 0.5, 0.75$. We consider the weight estimation error of GMV and MKW portfolios. For simplicity, we consider the case where the number of factors $K = 3$. The results for $K = 5$

are very similar and can be provided on request.

2.3.2 Results

Table 2.1, 2.2, and 2.3 show the simulation results. The value in each cell shows the average absolute estimation error of the optimal weights in GMV and MKW portfolios in low-dimensional cases (LD) and high-dimensional cases (HD). The tables illustrate that our method consistently achieves the principle stated in Theorem 1. For example, in Table 2.1, with $K = 3, \rho = 0.25$ and $n = 100$, the low-dimensional Global Minimum Variance portfolio (GMV LD) weights have an estimation error of 0.28. The error decreases to 0.25 with $n = 200$, to 0.18 with $n = 400$, and further decreases to 0.15 with $n = 600$. For every portfolio in Tables 2.1, 2.2, and 2.3, the estimation error declines monotonically with the number of time periods n, for both low- and high-dimenionsal portfolios.

Table 2.1: Weight estimation error, $\rho = 0.25$

	n=100	n=200	n=400	n=600
Markowitz HD	0.39	0.34	0.27	0.21
GMV HD	0.28	0.25	0.20	0.16
Markowitz LD	0.40	0.34	0.24	0.20
GMV LD	0.28	0.25	0.18	0.15

Table 2.2: Weight estimation error, $\rho = 0.5$

	n=100	n=200	n=400	n=600
Markowitz HD	0.41	0.29	0.21	0.18
GMV HD	0.32	0.23	0.17	0.15
Markowitz LD	0.43	0.30	0.21	0.18
GMV LD	0.34	0.24	0.17	0.15

Table 2.3: Weight estimation error, $\rho = 0.75$

	n=100	n=200	n=400	n=600
Markowitz HD	0.41	0.30	0.23	0.20
GMV HD	0.36	0.27	0.21	0.18
Markowitz LD	0.55	0.37	0.26	0.22
GMV LD	0.47	0.33	0.24	0.20

2.4 Empirical Application

The objective of this empirical application is to conduct an out-of-sample competition among different estimations. We estimate the Global Minimum Variance (GMV) and Markowitz Mean-Variance (MKW) portfolios using estimators and use the *Regret* metric as defined in (2.7). The Markowitz return constraint ρ_1 is set to 0.8 % per month.

We use the same data as in Chapter 1, which includes 304 stocks from January 1994 to May 2018 with an out-of-sample period of January 2004 to May 2018. We then randomly selected 50 stocks from the pool that served as instrumental variables. To compute *Regret*, we have to calculate the best fixed strategy in hindsight, using the plug-in estimator from out-of-sample data. Let w^*_{GMV} and w^*_{MKW} be the best fixed strategy in hindsight for GMV and MKW portfolios, respectively.

2.4.1 Metrics

We apply a rolling window approach to estimate the out-of-sample Sharpe Ratio, as in Callot et al. (2021), Caner et al. (2023) and Chapter 1. The sample of length n is divided into in-sample indexed $(1 : n_I)$ and out-of-sample indexed $(n_I + 1 : n)$. Also, define the average return as $\hat{r} = \frac{1}{n}\sum_{t=1}^{n} r_t$, where $r_t : p \times 1$ the vector of assets returns. Initially, we estimate the portfolio $\hat{w}_{n_I}$ in the in-sample period, then the out-of-sample portfolio returns $\hat{w}'_{n_I} r_{n_I+1}$ and the loss function $-\ln \hat{w}'_t(1_p + r_{n_I+1})$. Then, we roll the window forward by one element $(2 : n_I + 1)$, forming a new in-sample portfolio $\hat{w}_{n_I+1}$, calculating out-of-sample portfolio returns $\hat{w}'_{n_{I+1}} r_{n_I+2}$ and the loss function $-\ln \hat{w}'_{t+1}(1_p + r_{n_I+2})$. This process is repeated until the end of the sample. The out-of-sample average *Regret* without transaction costs is

$$\textit{Regret} = \frac{1}{n - n_i - 1} \ln w^{*\prime}(1_p + \hat{r}) - \ln \hat{w}'_t(1_p + r_{n_I+1}) \tag{2.20}$$

The out-of-sample Sharp Ratio is defined as in (1.47).

2.4.2 Methods

The mode comparisons include several approaches from the literature. We consider the Node-wise (NW) method from Callot et al. (2021), along with feasible residual-based nodewise methods from Caner et al. (2023). In particular, rbNW-SF, rbNW–3F, and rbNW-5F refer to residual-based nodewise regression using a single-factor (market as the only factor), a three-factor, and a five-factor model, respectively. rbIV-3F and rbIV-5F refer to three-factor and five-factor instrumental variables residual-based node-wise regression methods, respectively, proposed in Chapter 1.

2.4.3 Results

We report the *Regret* results for Global Minimum Variance and Markowitz portfolios without transaction costs in Tables 2.4 and 2.5. The Sharpe Ratio without transactions costs in Table 2.6 The first column represents the iteration of the trading periods. We consider 4 different trading periods, specifically $T = 50, 100, 150, 173$.

First, both tables show that the Nodewise method (NW) achieves the smallest average *Regret* metric, except when $T = 50$ in the Global Minimum Variance portfolio. Analyzing $T = 173$, the nodewise method achieves the lowest *Regret* at 0.0023, compare to other methods at around 0.0046.

Secondly, the Nodewise method in Callot et al. (2021) shows a consistent decline in the average of *Regret* as the trading period lengthens, aligning with the predictions of Theorem 1. In addition to the monotonic decline of *Regret* for the Global Minimum Variance portfolio noted above, Nodewise *Regret* in the Markowitz Mean Variance portfolio, starts at 0.0709 when $T = 50$, declines to 0.0705 when $T = 100$, further decreases to 0.0681, and reaches its lowest value at 0.0680 when $T = 153$. We also observe that the Nodewise method achieves the highest return of Global Minimum Variance portfolios, with the lowest *Regret* among all methods at 0.0023. This results suggests a clear relationship: A lower *Regret* metric corresponds to higher Sharpe Ratio.

Table 2.4: Empirical Results: Regret metric in Global Minimum Variance portfolio

	rbIV-5F	rbIV-3F	rbNW-SF	rbNW-3F	rbNW-5F	Nodewise
T=50	0.0051	0.0063	**0.0039**	0.0048	0.0051	0.0048
T=100	0.0073	0.0074	0.0056	0.0065	0.0073	**0.0046**
T=150	0.0037	0.0035	0.0026	0.0030	0.0037	**0.0024**
T=173	0.0046	0.0046	0.0042	0.0044	0.0046	**0.0023**

Table 2.5: Empirical Results: Regret metric in Markowitz Mean-Variance portfolio

	rbIV-5F	rbIV-3F	rbNW-SF	rbNW-3F	rbNW-5F	Nodewise
T=50	0.0711	0.0722	**0.0698**	0.0707	0.0711	0.0709
T=100	0.0729	0.0731	0.0713	0.0721	0.0729	**0.0705**
T=150	0.0690	0.0690	0.0681	0.0683	0.0690	**0.0681**
T=173	0.0700	0.0700	0.0697	0.0698	0.0700	**0.0680**

Table 2.6: Without the transaction cost ($n - n_I = 173$)

	Global Minimum				**Markowitz Portfolio**			
	Sharpe	Return	SD	Turnover	Sharpe	Return	SD	Turnover
EW	0.253	0.011	0.043	0.049				
NW	**0.278**	**0.010**	0.036	0.129	0.274	0.010	0.035	0.165
SFNLS	0.242	0.007	0.029	0.281	0.240	0.007	0.029	0.291
NLS	0.249	0.007	0.029	0.288	0.247	0.007	0.029	0.297
POET	0.240	0.007	0.030	0.091	0.247	0.007	0.030	0.115
1 Factor								
rbNW	0.251	0.008	0.031	0.117	0.250	0.008	0.031	0.141
3 Factors								
rbIV	0.253	0.007	0.029	0.259	0.256	0.007	0.029	0.278
rbNW	0.249	0.008	0.031	0.168	0.254	0.008	0.030	0.195
5 Factors								
rbIV	0.276	0.008	0.029	0.323	0.274	0.008	0.029	0.354
rbNW	0.246	0.007	0.030	0.189	0.251	0.007	0.030	0.215

CHAPTER

3

GENERAL LINEAR CLASS OF HIGH DIMENSIONAL ESTIMATORS FOR PRECISION MATRIX ESTIMATION

3.1 Introduction

In this paper, we provide a new precision matrix estimator for the econometrics-statistics literature. This can be used in high dimensional confidence interval estimators for structural parameters, network analysis, and financial metrics such as Sharpe Ratio. We start with a new formula for population precision matrix and show that it can be tied to factor model regression. Specifically, we benefit from latent factors (hidden factors). Then, we provide a new estimator for the precision matrix formula based on our population precision matrix. We also show that both ridge-less estimation and principal components based regression belong to same general class. Our estimator can be used in high dimensional context and does not require exact sparsity assumption on outcomes as in Callot et al. (2021). Note that nodewise regression-based estimators of Callot et al. (2021) or Caner et al. (2023) crucially depend on two key points: all factors are observed, and exact sparsity of either precision matrix of outcome variables or errors are needed, respectively. Technique used in those papers cannot extend to hidden factor models since observed factor structure uses OLS estimation, which is not applicable to more complicated estimation problems such as hidden factor models. We are also able to tie our results to signal-to-noise ratio aspect, which is not available in Callot et al. (2021), Caner et al. (2023) papers. We show that to get consistency of our estimators, signal-to-noise ratio should be larger than number of factors.

One key issue is why we use a hidden factor setup. First of all, we show that precision matrix estimation is related to linear regression, and that can be related to hidden factors. Each economic or financial variable can be related to latent factors, and this can be thought of as confounding variables as in causality-related literature. Also, the true variables of interest may be measured with errors, and latent factors can represent the "true" variables rather than the reported ones, especially in emerging markets. Variables such as culture, sentiment, and market can be extremely valuable in certain macrofinancial data sets. Another key aspect of hidden factor model usage is that this is a nice tool to provide dimension reduction in high dimensional econometrics.

Note that in getting confidence intervals for high dimensional models for structural parameters, debiased estimators are used, see Caner and Kock (2018),van de Geer et al. (2014). The crucial issue in these debiased estimators is to have an estimate of the inverse of the Gram Matrix, since the sample moments of the data will be singular in high dimensions, hence sample moments cannot be inverted. Our method can be modified to use as a new estimate in such models. Even in nonlinear Generalized Linear Models with structured sparsity estimators, the sample second order partial derivative of the loss is singular, hence cannot be used in high dimensions as shown by Caner et al. (2023). Our method can be used as an alternative inverse estimate of the second order partial derivative of convex loss functions. In network analysis, also our method can be used to have neighborhood selection, and can complement approach of Meinshausen and Bühlmann (2006). In financial econometrics, a penalized version is suggested by Lettau and Pelger (2020), and financial applications are in Lettau and Pelger (2020).

In high dimensional statistics, there is an excellent paper by Fan et al. (2013) that provides an analysis of hidden factors for covariance matrix estimation. Our paper differs from Fan et al. (2013) which uses a hidden factor model but uses a thresholding based principal components estimator. There are five major differences. First, we propose a general class of estimators, which also include ridge-less regression estimator as well as adaptive principal components which we describe in the paper below. Second, our signal assumption is much weaker than the one used in Fan et al. (2013), where they assume minimum signal to grow at rate of number of assets in the portfolio. Also, the third difference, the noise assumption in Fan et al. (2013) assumes maximum eigenvalue of the covariance matrix of errors, where the matrix dimension grows with number of assets p, is constant. In our paper, we have a diverging noise assumption, reflecting the financial markets. The fourth difference is that our analysis reflects finite sample analysis, hence can tie all estimation errors to signal-to-noise ratio explicitly and shows its importance in rate of convergence for our estimates for the rows of the precision matrix of asset returns. The fifth difference is that, Fan et al. (2013) rate of convergence of estimators for the precision matrix of asset returns depend on sparsity of the covariance matrix of errors, which we do not have.

Our paper differs from Bing et al. (2021) and Bunea et al. (2022) in a fundamental way. These papers analyze general linear class of estimators in hidden factor models in linear regression model and is only interested in risk of the estimators. Their papers consider only one outcome variable and how it can be predicted with this new general class of estimators. Theorem 3 of Bing et al. (2021) specifically provides a risk upper bound and this bound depends on the magnitude of coefficients in a linear system. Our analysis, on the other hand, considers the precision matrix estimation of outcome variable in linear regression via hidden factor models (i.e. an application is large portfolio formation). We show how matrix algebra can be combined with a novel idea of estimation to provide general class of estimators for precision matrix estimation. One of the main difficulties, theoretically, is that the estimation of the main diagonal terms in the precision matrix estimation. The proof technique and ideas in Bing et al. (2021) and Bunea et al. (2022) do not simply extend to this case. So precision matrix estimation is different than the risk analysis of a single asset. We also provide an explicit connection in our estimation error upper bounds to signal-to-noise ratio. The results in subcases are different also in a risk analysis versus precision matrix estimation. Bing et al. (2021) and Bunea et al. (2022) show that in certain very high dimensional cases, ridge-less regression can perform better than adaptive principal components in terms of risk asymptotically. However, in case of estimation errors of rows of precision matrix errors (in l_2-norm) we show adaptive principal components can perform better than ridge-less regression even in very high dimensional cases. These are due to different proofs for risk versus precision matrix estimation.

As a subcase of our general results we consider two important estimation techniques for precision matrix estimation, ridge-less regression (or Generalized Least Squares, GLS) and adaptive principal components regression (adaptive PCR). We observe the double-descent phenomenon also in precision matrix estimation, for each row in l_2 norm. This is a new result and shows that double descent phenomenon is more extensive than linear regression. Let p represent the number of outcome variables, and n as the sample size. Double-descent describes estimation errors going to zero when $p/n \to 0$ when both grow, and when $p/n \to \gamma$ when γ increases towards a large constant above one. Our GLS based precision matrix estimation is tuning parameter free and so has that extra advantage.

We also compare two subcase estimators with each other. In finite samples, GLS based estimator has one less bias term compared with adaptive PCR. Asymptotically however, the ranking may change, and adaptive PCR bias terms depend on number of factors, and GLS bias depends on sample size, and since number of factors may be smaller than the sample size, adaptive PCR may be better than GLS, in terms of l_2 estimation errors for each row of the precision matrix.

3.2 Model

In our model we have p outcome variables, and they are explained by K common hidden factors. Let Y represent the $n \times p$ matrix of outcomes, each row representing the t^{th} time period, $t = 1, \cdots, n$, and $j = 1, \cdots, p$ representing the outcomes. F represents $n \times K$ matrix of hidden factors, where each row, $t = 1, \cdots, n$ represents a time period, and columns represent the hidden factors, $k = 1, \cdots, K$, A represents $K \times p$ matrix of factor loadings. So we have the following model which we will decompose a row Y into two equations immediately below in (3.2) and (3.3).

$$Y = FA + U. \tag{3.1}$$

In accordance with our analysis, we subdivide all p outcomes into j^{th} one and all the remaining ones. In that respect, start by defining each row of the outcome as $y_t : p \times 1$ (represents each row, in column form, whereas $y_t' : 1 \times p$ is the row vector form) of Y above, and $f_t : K \times 1$ represents each row (in column form) of F, and $u_t : p \times 1$ is the t^{th} row (in column form) of U. We decompose A into $a_j : K \times 1$ factor loading for outcome j, and $A_{-j} : p-1 \times K$ for all other outcomes except j. In other words $A := [a_j, A_{-j}']'$. We model y_{jt} as the j^{th} outcome variable at time period $t = 1, \cdots, n$. As an example, the j^{th} outcome here can be j^{th} asset's excess return (over risk free rate such as 3 month T-Bill in USA) at time t. We define all the outcomes except j^{th} outcome as $Y_{-jt} : p-1$ vector at time t. This may represent in the same way as in the example above, all the excess asset returns in a portfolio of p assets except the j^{th} one.

We follow the hidden (latent) factor regression model in Bunea et al. (2022) and Bing et al. (2021). Let $f_t : K \times 1$ hidden (unobserved) factors. In the t^{th} row of Y in (3.1) j^{th} outcome can be written as

$$y_{jt} = f_t' a_j + u_{jt} \tag{3.2}$$

and we write all the outcomes except from the j^{th} one, in t^{th} row as

$$Y_{-jt} = A_{-j} f_t + U_{-jt}. \tag{3.3}$$

where $a_j : K \times 1$ latent variable regression vector, and $A_{-j} : (p-1) \times K$ are the factor loading matrix for each $j = 1, \cdots, p$. We use (3.2)(3.3) in our proofs rather than (3.1), and in this way our proofs are clear and tractable. Common hidden factors affect j^{th} outcome in (3.2) as well as all remaining outcome variables in (3.3). Let u_{jt}, U_{-jt} represent the errors for the j^{th} outcome, and all other remaining outcomes respectively. In vector form, the j^{th} column of Y in (3.1) can be written as

$$y_j = F a_j + u_j, \tag{3.4}$$

where $y_j : n \times 1$, and $F : n \times K$ with each row as f_t'. Then the remainder, except j^{th} column, of Y is:

$$Y_{-j} = F A_{-j}' + U_{-j}, \tag{3.5}$$

where $Y_{-j} : n \times p - 1$ matrix, and $U_{-j} : n \times p - 1$ matrix too. Now we provide our Assumptions as used in Bunea et al. (2022).

Assumption 1

(i) y_{jt}, Y_{-jt} are iid across $t = 1, \cdots, n$ for each $j = 1, \cdots, p$

(ii) u_{jt}, U_{-jt} are zero mean random variable and vector respectively, which are iid across $t = 1, \cdots, n$, for each $j = 1, \cdots, p$

(iii) f_t are iid across $t = 1, \cdots, n$

(iv) f_t, u_{jt}, U_{-jt} are mutually independent across $t = 1, \cdots, n$ given $j = 1, \cdots, p$

Assumption 1 is a standard assumption used in Bing et al. (2021) and Bunea et al. (2022), imposes independence across $t = 1, \cdots, n$ for factors and returns.

Note that Assumption 1 (i)(ii)(iii) shows that rows of Y, U, F are iid. However, when there is independence of errors which implies $E u_{jt} U_{-jt} = 0$, which is a sparsity restriction on the covariance matrix of errors. This type of restriction can be seen in Fan et al. (2008). This amounts to a diagonal error covariance matrix. We thought about relaxing this part of the assumption. Since we pursue a general precision matrix of outcome variables in a hidden factor model, and our outcomes are still correlated through the hidden factor structure. A future extension of our work can analyze a relaxed assumption about this specific type of sparsity.

Assumption 2

(i) $\min_{1 \le j \le p} \text{Eigmin}(\Sigma_{U,-j}) \ge c > 0$, where $\text{Eigmin}(\Sigma_{U,-j})$ represents the minimum eigenvalue of the matrix $\Sigma_{U,-j}$. Also $\max_{1 \le j \le p} \sigma_j^2 = \sigma^2 \le C < \infty$, with C a positive constant, and also $\min_{1 \le j \le p} \sigma_j^2 \ge c > 0$, for a positive constant c.

(ii) Let $\delta_n := \min_{1\leq j\leq p} \|\Sigma_{U,-j}\|_2$, and δ_n is nondecreasing in n, and also $r_n := \max_{1\leq j\leq p} \|\Sigma_{U,-j}\|_2$, where r_n is nondecreasing in n and may diverge with n, but $r_n/p \to 0$, and clearly $\delta_n \leq r_n$.

Assumption 2(i) is also standard assumption in the high dimensional statistics literature, allowing non-singular noise matrix. Assumption 2(ii) allows for large noise and allows divergence in the maximum eigenvalue of the noise. This assumption is reflective of financial markets, and is used in Caner et al. (2023).

Assumption 3: For each $j = 1, \cdots, p$, $A_{-j} : p-1 \times K$ has full rank K, as well as Σ_f, and $p > K+1$. We impose $\max_{1\leq j\leq p} \|A_{-j}\|_2 = O(d_{1n})$ with $d_{1n} \to \infty$ as $n \to \infty$, and $d_{1n}/\sqrt{p} \to 0$. Also Eigmax$(\Sigma_f) \leq C < \infty$, for C a positive constant.

Assumption 3 with full rank of $\Sigma_f := E(f_t - Ef_t)(f_t - Ef_t)' : K \times K$ matrix and is used in Bing et al. (2021) and Bunea et al. (2022), reflecting $K < p-1$, where we assume the number of hidden factors to be less than number of assets minus one. We allow the spectral norm of $A_{-j} : p-1 \times K$ matrix to be diverging in n, reflective of the high dimensional nature of A_{-j}, where both dimensions are allowed to grow.

Assumption 4

(i) Define, for each $j = 1, \cdots, p$, u_{jt} has zero mean and its a $\sigma_j \gamma_e$ subgaussian random variable which means for each $j = 1, \cdots, p$, $E[exp(t_1 u_{jt})] \leq exp(t_1^2 \sigma_j^2 \gamma_e^2/2)$ for all $t_1 \in R$. Define $U_{-j,t} := \Sigma_{U,-j}^{1/2} \tilde{U}_{-j,t}$, where $\tilde{U}_{-j,t}$ ($p-1 \times 1$ vector) is γ_{wj} sub-Gaussian vector which means for any unit vector ($p-1$ dimension) v, $v'\tilde{U}_{-j,t}$ is γ_{wj} sub-Gaussian. Also we assume $E\tilde{U}_{-j,t}\tilde{U}'_{-j,t} = I_{p-1}$. Without losing any generality in the proofs we can use γ_w constant instead of changing constants for each j, and γ_{wj}.

(ii) Define $f_t := \Sigma_f^{1/2} \tilde{f}_t$, where $\tilde{f}_t : K \times 1$ is a subgaussian vector with $E[\tilde{f}_t \tilde{f}_t'] = I_K$, and $\max_{1\leq k\leq K} |Ef_{tk}| \leq C < \infty$, where $C > 0$ is a positive constant

(iii) We assume $\|A\|_\infty \leq C < \infty$.

Assumption 4 is standard assumption used in Bing et al. (2021) Bunea et al. (2022). These are needed to use high dimensional probability inequalities in quadratic form and also need to have lower bounds for the singular values of certain matrices used in the proof. We relax the non-zero factor mean assumption used in Bing et al. (2021) and Bunea et al. (2022).

Define the signal-to-noise ratio uniformly for the portfolio of the assets, before the next assumption

$$\bar{\xi} = \min_{1\leq j\leq p} \xi_j,$$

whereas in Bunea et al. (2022)

$$\xi_j := \frac{\lambda_K(A_{-j}\Sigma_f A'_{-j})}{\|\Sigma_{U,-j}\|_2},$$

with $\lambda_K(A_{-j}\Sigma_f A'_{-j})$ representing the K th largest eigenvalue of $A_{-j}\Sigma_f A'_{-j} : p-1 \times p-1$ matrix, and that is the smallest positive eigenvalue of that matrix since $p-1 > K$, and $\Sigma_f : K \times K$ matrix. We have a discussion about the role of the signal-to-noise ratio in a specific scenario. Since $A_{-j} : p-1 \times K$ for all matrices $j = 1, \cdots, p$, via Assumption 3, if the largest eigenvalue of $A'_{-j}A_{-j}$ grows at the same rate K-th largest eigenvalue, we can assume $\lambda_K(A'_{-j}A_{-j}) \asymp d_{1n}^2$ and if we impose $\max_{1\leq j\leq p} \|\Sigma_{U,-j}\|_2 = r_n$, then

$$\lambda_K(A_{-j}\Sigma_f A'_{-j}) \geq \text{Eigmin}(\Sigma_f)\lambda_K(A_{-j}A'_{-j}) \geq c\,\lambda_K(A'_{-j}A_{-j}) \asymp d_{1n}^2.$$

Then the signal-to-noise ratio is lower bounded by, with $c > 0$, a positive constant

$$\bar{\xi} \geq \frac{c\, d_{1n}^2}{r_n} \tag{3.6}$$

The lower bound of $\bar{\xi}$ may grow with low r_n, and larger d_{1n}^2 rate. On the other hand, the lower bound can converge to zero, since d_{1n}^2 may grow at a slower rate than r_n in certain cases. So we allow a large range of possibilities for the lower bound in signal-to-noise ratio.

Assumption 5 below also covers diagonal estimation of precision matrix. Note that $\bar{r}, \bar{\Psi}, \bar{\eta}$ are three random sequences that are explained in detail before Lemma 1, and they represent the complexity of the model, prediction error of technique that is used, and prediction performance of the technique, respectively. In special cases of Generalized Least Squares, and adaptive PCR, we will show non-random upper bounds for these three key terms.

Now, define two rates, r_{w1} and r_{w2}. Note that r_{w2} is the square root of the rate for off-diagonal terms in the precision matrix estimation. The faster rate r_{w1} (compared to r_{w2}) is also used in estimation for the off-diagonal terms.

$$r_{w1} := \max\left(\sqrt{\frac{\log n + \bar{r}}{n\bar{\eta}}}, \sqrt{\frac{K}{\bar{\xi}\delta_n}}, \sqrt{\frac{K}{\bar{\xi}}\frac{\bar{\Psi}}{\bar{\eta}}\frac{1}{\delta_n}}, \sqrt{\frac{K}{\bar{\xi}}\frac{1}{\bar{\eta}}}\right) \tag{3.7}$$

$$r_{w2} := \max\left(r_{w1}, \sqrt{\frac{K}{\bar{\xi}}}\right) \tag{3.8}$$

Note the only difference between r_{w1} and r_{w2} is the second term in parentheses in (3.7), and (3.8). Since δ_n is nondecreasing in n, we can tell that r_{w2} is the slower rate. Define the relation between the factor loadings of j^{th} asset with all the others as $d_{2n} := \max_{1\leq j\leq p} \|a'_j\Sigma_f A'_{-j}\|_2$.

Assumption 5

(i) $\frac{d_{1n}}{\sqrt{p}} \frac{\sqrt{\log n}}{\sqrt{K \log p}} \to 0$

(ii) $\sqrt{\frac{K^3}{n}} \to 0, \quad K \frac{\sqrt{max(\log p, \log n)}}{n^{1/2}} \to 0$

(iii) $K \sqrt{\frac{p(\log p)}{n}} r_{w1} \xrightarrow{p} 0$, and $d_{2n} r_{w2} \xrightarrow{p} 0$, and d_{2n} is nondecreasing in n

Assumption 5(i) is a technical assumption that facilitates the proof of Lemma A.1. With larger p compared to n, Assumption 3 will suffice, and this assumption will not be needed.

Assumptions 5(ii) shows that there is a tradeoff between number of factors and time span, number of factors should be much smaller than time span. Assumption 5(iii) involves stochastic terms and related to consistency and rate of convergence of the off-diagonal term estimation. Assumption 5(iii) will be analyzed individually for adaptive PCR and GLS estimators.

3.3 Linear General Class of Estimators

We propose to estimate rows of the precision matrix of asset returns with our estimator. In order to achieve this objective first, we need to start with a precision matrix formula that can be tied to hidden factor models. This is not a trivial task since it is not clear how that can be done. One of the main difficulties is that how to relate the matrix inversion to linear regression context in hidden factor model context. The difficulty is that formula for main diagonal terms and off-diagonal terms of the precision matrix is different and this type of approach tying hidden factors and general linear class of estimators with precision matrix estimation has not been done before. This is one of the contributions of our paper. Define the following parameter vector which will form a key term in precision matrix formula:

$$\alpha_j^* := \arg \min_{\alpha_j \in R^{p-1}} E[(y_{jt} - \mu_j) - (Y_{-jt} - \mu_{-j})' \alpha_j]^2$$

This definition shows that we can relate to linear regression context. Note that, the definition of α_j^* is not coming from the hidden factor model directly in (3.2), and (3.3). We also need the following definitions before the next Lemma. Let $\Sigma_f = \Sigma_f^{1/2} \Sigma_f^{1/2}$ since Σ_f is a positive definite matrix.

Set $\bar{A}_{-j} := A_{-j} \Sigma_f^{1/2}$, $\bar{a}_j := \Sigma_f^{1/2} a_j$, and

$$\bar{G}_j := I_K + \bar{A}_{-j}' \Sigma_{U,-j}^{-1} \bar{A}_{-j}$$

Next, we show how they relate to hidden factor models. Let $\Theta_{j,j}$ represent the main j^{th} diagonal term, and $\Theta_{j,-j}$ represent the j^{th} row without j^{th} element in j^{th} row of Θ.

Lemma 1: Under Assumptions 1-3, for $j = 1, \cdots, p$

(i)

$$\Theta_{j,j} = \frac{1}{\tau_j^2},$$

with

$$\tau_j^2 = (\bar{a}_j' \bar{a}_j + \sigma_j^2) - \bar{a}_j' \bar{A}_{-j}' \Sigma_{U,-j}^{-1} \bar{A}_{-j} \bar{G}_j^{-1} \bar{a}_j$$

(ii)

$$\Theta_{j,-j} = \frac{-\alpha_j^{*'}}{\tau_j^2},$$

with

$$\alpha_j^* = \Sigma_{U,-j}^{-1} \bar{A}_{-j} \bar{G}_j^{-1} \bar{a}_j$$

Remarks.

1. Combining (i)(ii), we form j^{th} row of the precision matrix formula in terms of a hidden factor model. Stacking each row will produce Θ, the precision matrix.

2. Note that as seen in (ii), the coefficients $\alpha_j^* : p-1 \times 1$ are related to scaled-factor loadings $\bar{a}_j : K \times 1$. See also that $K < p-1$ by our Assumptions, and this shows dimension reduction through hidden factors.

3. See that α_j^*, τ_j^2 formulas are in terms of hidden factor model coefficients, hence this is new compared to literature as in Caner and Kock (2018), Caner and Kock (2019), and Caner et al. (2023), where they analyze outcomes without factors, errors in an observed factor model, and regressors in a linear regression respectively.

Note that Bing et al. (2021) proposes linear regression coefficient estimation with hidden factor models and analyzes the risk of an asset in a hidden factor model. However, the precision matrix estimation or formulae is not discussed in Bing et al. (2021) and Bunea et al. (2022). The difficulty is that in a precision matrix, there are p rows, and these rows will be tied to hidden factor models, and estimation will be based on this and without imposing any identification conditions on hidden factors. So it is not obvious how and why linear regression estimators in a hidden factor model as in Bing et al. (2021) and Bunea et al. (2022) should generalize to the precision matrix estimation. With Lemma 1 and Theorem 1 below, we solve this issue. In that respect for each $j = 1, \cdots, p$, we define

$$\tilde{\alpha}_j := \hat{B}(Y_{-j}\hat{B})^+ y_j = \hat{B}(\hat{B}'Y'_{-j}Y_{-j}\hat{B})^+(\hat{B}'Y'_{-j}y_j), \tag{3.9}$$

where $(.)^+$ is the Moore-Penrose inverse of a matrix, and $\hat{B} : p-1 \times q$ with $q \leq p-1$ matrix. Dimension q will be tied to specific estimators, and immediately below, we give examples about that. Our estimator is based on analysis of precision matrix in Lemma 1 and the equality in (3.9) is by (134) of Bunea et al. (2022).

With different values of $\hat{B}$ we obtain different estimators. To see that $\hat{B} = I_{p-1}$ is the minimum norm estimator which is analyzed also after this section separately, so $q = p-1$. In case of $\hat{B} = V_{1,k_j}$ with V_{1,k_j} being the $p-1 \times k_j$ matrix where the columns are the eigenvectors of $Y'_{-j}Y_{-j}/n$ matrix corresponding to the largest k_j eigenvalues, and $q = k_j$, $j = 1, \cdots, p$. We do not subscript $\hat{B}$ with j to simplify the notation. Let us define the reciprocal of the main diagonal terms in the precision matrix [1]

$$\tilde{\tau}_j^2 := \frac{y_j'(y_j - Y_{-j}\tilde{\alpha}_j)}{n} \tag{3.10}$$

Then the j the main diagonal term estimator in the precision matrix is:

$$\tilde{\Theta}_{j,j} = 1/\tilde{\tau}_j^2 \tag{3.11}$$

and the j^{th} row of the precision matrix estimate without j^{th} element (off-diagonal terms in the row) is

$$\tilde{\Theta}_{j,-j} = -\tilde{\alpha}_j'/\tilde{\tau}_j^2 \tag{3.12}$$

The j^{th} row of the estimate, to give an example, when $j = 1$ will be $\tilde{\Theta}_1' := (\frac{1}{\tilde{\tau}_1^2}, \frac{-\tilde{\alpha}_1{}'}{\tilde{\tau}_1^2})$.

3.4 Upper Bound on l_2-norm error

In this section, our aim is to get an upper bound for

$$\max_{1 \leq j \leq p} \|\tilde{\alpha}_j - \alpha_j^*\|_2,$$

which is essential in getting consistency of estimates of rows of the precision matrix by minimum norm estimator. Then let $P_{\hat{B}}$ is the projection matrix onto the range $\hat{B}$, i.e. $P_{\hat{B}} = \hat{B}[\hat{B}'\hat{B}]^+\hat{B}' = \hat{B}\hat{B}^+$, where M^+ where denotes the Moore-Penrose inverse of a matrix M, and the last equality

[1] Note that only in case of GLS-ridgeless estimator as the subcase when $p > n$ the estimator simplifies and we benefit from that expression.

is by Moore-Penrose matrix inverse rules such as (134) of Bunea et al. (2022).

Define the rank of $Y_{-j}P_{\hat{B}}$ as

$$\hat{r}_j := rank(Y_{-j}P_{\hat{B}}) \tag{3.13}$$

This term reflects the complexity in the model, and increases the bias of the estimation of rows of the precision matrix, as we show below in Lemma 1. Define

$$\hat{\eta}_j := \sigma^2_{\hat{r}_j}(Y_{-j}P_{\hat{B}})/n \tag{3.14}$$

which is the $\hat{r}_j$- th largest singular value (squared, scaled by n) of $Y_{-j}P_{\hat{B}}$ ($n \times q$ matrix). This last term is obtained when we project Y_{-j} onto range of $\hat{B}$, and will reduce the bias of estimating the rows of the precision matrix in Lemma 1 below. Next, we define

$$\hat{\Psi}_j := \sigma^2_1(Y_{-j}M_{\hat{B}})/n \tag{3.15}$$

where $M_{\hat{B}} := I_{p-1} - P_{\hat{B}}$, and $Y_{-j}M_{\hat{B}}$ is $n \times p-1$ matrix. This term is the largest singular value of $Y_{-j}M_{\hat{B}}$ matrix, and it represents the bias stemming from using a hidden factor model to estimate Y_{-j}. This last term will increase the bias of estimating the rows of the precision matrix. Define $\max_{1\le j\le p} \hat{r}_j := \bar{r}$, $\min_{1\le j\le p} \hat{\eta}_j := \bar{\eta}$, and $\max_{1\le j\le p} \hat{\Psi}_j := \bar{\Psi}$. We have the following uniform bounds for estimators for off-diagonal terms in each row of the precision matrix

Theorem 1

(i) Under Assumptions 1-4, $K \log n \le c\, n$, with $c > 0$ a positive constant, with probability at least $1 - 2/n$

$$\max_{1\le j\le p} \|\tilde{\alpha}_j - \alpha^*_j\|_2^2 \le 2\gamma_e^2\sigma^2\left(\frac{2\log n + \bar{r}}{n\bar{\eta}}\right) + \frac{CK}{\bar{\xi}}\left(1 + \frac{1}{\delta_n} + \frac{\bar{\Psi}}{\bar{\eta}\delta_n} + \frac{r_n}{\bar{\eta}\delta_n}\right).$$

(ii) Under Assumptions 1-4, and $r_n/\delta_n \to 1$,

$$\max_{1\le j\le p} \|\tilde{\alpha}_j - \alpha^*_j\|_2^2 = O_p\left(\max(\frac{K}{\bar{\xi}}, \frac{K\bar{\Psi}}{\bar{\xi}\bar{\eta}\delta_n}, \frac{K}{\bar{\xi}\bar{\eta}}, \frac{\log n + \bar{r}}{n\bar{\eta}})\right).$$

Remarks:

1. Note that $\bar{\xi}$ plays an essential role in getting consistency. Larger signal-to-noise ratio is needed, especially if the number of factors K grow fast.

2. Noise δ_n plays a regularizing role in obtaining consistency, and this type of phenomenon is observed in achieving the optimal risk by minimum l_2-norm estimator as seen in Bunea

et al. (2022).

3. One key issue is that our Lemma 1 forms a key building block for the consistency of rows of precision matrix estimation.

4. The expression will simplify in cases of GLS and Principal Components based estimation subcases. But $\bar{r}$ represents complexity, and affects the l_2-norm error adversely through the increase in the variance of estimator. $\bar{\eta}$ on the other hand is a positive influence on the l_2 error bound, since it shows with better prediction of asset returns via hidden factors we get a lower l_2 error. The positive influence is through the reduction of the bias of the estimator. $\bar{\Psi}$ shows how badly hidden factor estimates predict asset returns, and it has adverse effect on l_2 error. This term increases the bias of the estimator.

5. The first term on the right side of Lemma 1(i) is the variance term and the second term is the bias term.

6. There is no sparsity parameter on the upper bounds and rates.

3.4.1 Principal Components Estimators

In this subsection we consider principal components regression (adaptive PCR from now on) estimators. Let us designate $rank(Y_{-j}) \geq k_j \geq 1$, where k_j is a positive integer and changes with $j = 1, \cdots, p$. The adaptive PCR estimators use $\hat{B} = V_{1,k_j}$, where V_{1,k_j} is $p-1 \times k_j$ matrix where V_{1,k_j} represents a matrix which k_j columns (eigenvectors) correspond to the largest k_j eigenvalues of $Y_{-j}' Y_{-j}/n$ for $j = 1, \cdots, p$. The adaptive PCR estimator is, for $j = 1, \cdots, p$

$$\hat{\alpha}_{PCR,j} = V_{1,k_j}(Y_{-j} V_{1,k_j})^{+} y_j,$$

where $(.)^{+}$ is Moore-Penrose inverse of a matrix.

Before simplifying the terms in the upper bound in Lemma 1 we define $\hat{\lambda}_{k_j} := \frac{1}{n}\sigma^2_{k_j}(Y_{-j})$, for example with $k_j = 1$ we have the largest singular value of Y_{-j}, which is $\sigma_1^2(Y_{-j})/n$ (squared-scaled by n). One big issue is the choice of k_j. The method to choose k_j is called the elbow method by Bing et al. (2021). (19) of Bing et al. (2021) provides

$$\hat{s}_j := \max\{k_j \geq 0 : \hat{\lambda}_{k_j} \geq C_0 \Delta_{-j}\},$$

with $C_0 > 1$ and, $c_\Delta > 1$ are positive constants above one,

$$\Delta_{-j} := c_\Delta[\|\Sigma_{U,-j}\|_2 + tr(\Sigma_{U,-j})/n],$$

where we adapt from the simple regression framework of Bing et al. (2021) to our own framework which concentrates on estimation of precision matrix of asset returns. Hence we will need results uniformly over $j = 1, \cdots, p$ which extends Bing et al. (2021) and show a feasible estimator of s_j in Remark 6 to Corollary 1. Note that $k_j = \hat{s}_j$. By definition of $\hat{s}_j$ above

$$\hat{\lambda}_{\hat{s}_{j+1}} < C_0 \Delta_{-j} \leq \hat{\lambda}_{\hat{s}_j}. \tag{3.16}$$

We show in Corollary 1 in

$$P(\max_{1 \leq j \leq p} \hat{s}_j \leq K) \to 1, \tag{3.17}$$

which will simplify the upper bound for l_2-norm error. We prove the following error bounds for adaptive PCR based estimator for the off-diagonal terms in a given row for the precision matrix.

Corollary 1:

(i) Under Assumptions 1-4 and $\log p < c'n$, with $0 < c' < 1$ a positive constant, $K \log n \leq cn$, with c a positive constant, with probability at least $1 - 2/n - exp((c'-1)n)$, and

$$\begin{aligned} \max_{1 \leq j \leq p} \|\hat{\alpha}_{PCR,j} - \alpha_j^*\|_2^2 &\leq 2\gamma_e^2 \sigma^2 \left[\frac{2 \log n + K}{n r_n} \right] \\ &+ \frac{CK}{\bar{\xi}} \left[1 + \frac{3}{\delta_n} \right]. \end{aligned}$$

(ii) Under the same Assumptions in (i)

$$\max_{1 \leq j \leq p} \|\hat{\alpha}_{PCR,j} - \alpha_j^*\|_2^2 = O_p\left(max(\frac{\log n + K}{n r_n}, \frac{K}{\bar{\xi}}) \right).$$

Remarks.

1. This is a new result for precision matrix row estimation. It clearly shows, for the first square bracketed term on (i)-(ii), which is related to variance of estimator, to go to zero we need $(\log n + K)/n r_n \to 0$ and this is easy to satisfy given $K \log n < cn$ is assumed. Increasing the number of factors can be detrimental, unless the noise r_n is large.

2. Regarding the bias term (the second square bracket in (i)) and the rate corresponding in (ii) the signal to noise ratio $\bar{\xi}$ should grow larger than K to have consistency of the estimator. Using the example (3.6) we may need $K r_n / d_{1n}^2 \to 0$.

3. We note that with $p > n$, consistent estimation of α_j^* is possible, and this is also true with $p \leq n$ as well. The price to pay is we can only have $\log p < c'n$, with $0 < c' < 1$.

4. If we fix $p/n = \gamma \in [0, \infty)$ and then we increase n and p together, and we keep γ fixed to check consistency. In $(\log n + K)/(n r_n)$ condition, r_n can grow with n and p both increasing and have a positive effect on consistency. The price here is $\log n$ rate, but that is not large compared to n rate in the denominator. In the second condition $K/\bar{\xi}$, by using the example (3.6), it is tied to $K r_n / d_{1n}^2$ rate as discussed in Remark 2 here. With $K r_n / d_{1n}^2 \to 0$ we have consistency. But a very low signal to noise ratio can prevent consistency if the number of factors are fixed.

5. Selection of the principal components is by the elbow method which depends on an infeasible lower bound. To have a feasible estimation, we follow (56) of Bing et al. (2021) or (2.7) of Bing and Wegkamp (2019). Note that in additional to Assumptions in Corollary 1 above, by assuming $tr(\Sigma_{U,-j}) \geq c'' \|\Sigma_{U,-j}\|_2 (n \wedge p)$, with $c'' > 0$ a positive constant, for each $j = 1, \cdots, p$, Proposition 8 of Bing et al. (2021) can be extended with a simple union bound and the feasible estimation of the unknown number of factors results do not change Corollary 1.

In that respect, we set, for $j = 1, \cdots, p$

$$\tilde{s}_j := \sum_{k=1}^{\bar{K}} 1_{\{\hat{\sigma}_k^2(Y_{-j}) \geq \mu_n \hat{v}_{k,j}^2\}},$$

where $\mu_n := \frac{1}{4}(n + (p-1))$, and $\hat{\sigma}_k^2(Y_{-j})$ is the k th largest singular value (squared) of $Y_{-j} : n \times p-1$ matrix, and

$$\hat{v}_{k,j}^2 := \frac{\|Y_{-j} - Y_{-j,(k)}\|_F^2}{n(p-1) - \mu_n k},$$

with $Y_{-j,(k)}$ representing the best k rank approximation of Y_{-j}, (Singular value decomposition into largest k singular values)$\|.\|_F$ is the Frobenius norm, and

$$\bar{K} := \left(\frac{\kappa}{1+\kappa} \frac{n(p-1)}{\mu_n}\right)_{fli} \wedge n \wedge p,$$

with $(.)_{fli}$ represents the floor integer function, the integer part of the real number, $\kappa > 1$ is a positive constant number.

3.4.2 GLS Estimation

In this part, we consider GLS estimators when $p - 1 < n$ first and then analyze $p > n$ case. In that respect $\hat{B} = I_{p-1}$, which simplifies the upper bound in Lemma 1.

$$\hat{\alpha}_{GLS,j} = (Y_{-j})^+ y_j,$$

for $j = 1, \cdots, p$.

Note that when $p-1 < n$, GLS estimation is equivalent to OLS, and by using (132) of Bunea et al. (2022)

$$\hat{\alpha}_{GLS,j} = (Y_{-j})^{+} y_j = (Y_{-j}' Y_{-j})^{-1} Y_{-j}' y_j.$$

Also, in the financial econometrics literature, OLS based estimation when p is large and n is large and $p-1 < n$ is known to be unstable. We show that after Corollary 2-Remark 1, indeed in an important subcase, OLS will not perform well in finite samples.

Corollary 2:(GLS: $p-1 < n$)

(i) Under Assumptions 1-4 and $\log 2p < (c_1/2)\log n$, with $c_1 > 0$, and $p \log n \leq c' n$with $c' \in (0,1)$ a positive constant, $K \log n < c n$, with c a positive constant, with probability at least $1 - 2/n - 1/n^{c_1/2}$, and

$$\begin{aligned}\max_{1 \leq j \leq p} \|\hat{\alpha}_{GLS,j} - \alpha_j^*\|_2^2 \quad &\leq 2\gamma_e^2 \sigma^2 \left[\tfrac{2\log n + p}{Cn}\right] \\ &+ \tfrac{3CK}{\bar{\xi}}\left[1 + \tfrac{1}{\delta_n} + \tfrac{r_n}{C\delta_n}\right].\end{aligned}$$

(ii) Under the same Assumptions in (i), with added $r_n/\delta_n \to 1$,

$$\max_{1 \leq j \leq p} \|\hat{\alpha}_{GLS,j} - \alpha_j^*\|_2^2 = O_p(max(\frac{\log n + p}{n}, \frac{K}{\bar{\xi}})).$$

Remarks.

1. Note that when p is fixed, n is fixed and $p/n = \gamma \in (0,1)$, where γ is a positive fraction between 0 and 1, GLS or equivalently OLS does not perform well in finite samples. This is clear from the first term on the right side in Corollary 2(i).

2. In large samples, a comparison with adaptive PCR shows that since $p > K$ by Assumption 3 we expect adaptive PCR to have a better rate in estimation error. Signal-to-noise affect both estimators in the same way by reducing the bias term. Note that variance term in adaptive PCR, Corollary 1(ii), we may have $r_n \to \infty$ in the denominator so adaptive PCR has extra advantage of having a better variance than GLS $(p-1 < n)$. Adaptive PCR will perform better when $p/n = \gamma \in (0,1)$ case unlike GLS here, as discussed in Remark 1 above in finite samples.

3. Note that there is one less bias term in upper bound in GLS compared with adaptive PCR due to $\bar{\Psi} = 0$ in GLS. This may be the only advantage in finite sample context.

Now we provide the Corollary for the GLS case when $p-1>n$. But define the effective rank as

$$re(\Sigma_{U,-j}) := \frac{tr(\Sigma_{U,-j})}{\|\Sigma_{U,-j}\|_2}$$

Corollary 3: (GLS: $p-1>n$)

(i) Under Assumptions 1-4 and $\log 2p < (c_2/2)n$, with $c_2>0, C>0$, and

$$min_{1\le j\le p} re(\Sigma_{U,-j}) > Cn,$$

for C a positive constant $K\log n < cn$, with c a positive constant, and $\tilde{U}_{-j}$ has independent entries then with probability at least $1-2/n-exp((-c_2/2)n)$, and

(ii)

$$\begin{aligned}\max_{1\le j\le p} |\hat{\alpha}_{GLS,j}-\alpha_j^*\|_2^2 &\le 2\gamma_e^2\sigma^2\left[\tfrac{2\log n+n}{Cnr_n}\right] \\ &+ \tfrac{3CK}{\xi}\left[1+\tfrac{2}{\delta_n}\right].\end{aligned}$$

(iii) Under the same Assumptions in (i)

$$\max_{1\le j\le p}\|\hat{\alpha}_{GLS,j}-\alpha_j^*\|_2^2 = O_p(max(\frac{\log n+n}{nr_n},\frac{K}{\xi}))$$

Remarks.

1. First, we want to see in the case of $p/n=\gamma\in(1,\infty)$, with n and p grow but γ is taken fixed, whether GLS is consistent or not. Clearly

   $$\frac{\log n+n}{nr_n}\to 0$$

   since r_n may grow larger. So signal-to-noise ratio should grow and also dominate the number of factors as well.

2. Note that we may see double descent in GLS for off-diagonal term estimation α_j^* in the precision. matrix. We fix n and K, then analyze the changes in p. To see that in the previous Corollary, we need $p/n\to 0$ to get estimation errors converging to zero. So smaller p in OLS (GLS $p-1<n$) case provides estimation error converging to zero. Then in minimum-norm interpolator case here (GLS with $p-1>n$), with $p/n\to\infty$ with n

fixed, since by definition r_n may only depend on dimension p, and noise r_n may grow with larger p

$$\frac{\log n + n}{n r_n} \to 0,$$

and $K/\bar{\xi} \to 0$ as discussed above since r_n, d_{1n}^2 depend on p and if d_{1n}^2 dominates r_n, we may see estimation errors converging to zero, and hence double descent in sum of all comments here.

3. We compare now GLS in this case with adaptive PCR. In GLS there is one less bias term compared with adaptive PCR ($\bar{\Psi} = 0$ in GLS) in finite sample upper bound. Since $K < n$, adaptive PCR has a smaller variance term compared with GLS. Asymptotically bias terms have the same rate $K/\bar{\xi}$. So asymptotically adaptive PCR has smaller estimation errors, but in finite samples due to one less bias term in GLS, GLS may have smaller bias, compared to adaptive PCR.

4. Adaptive PCR also may have one more theoretical advantage compared with GLS. We need the noise r_n to be increasing in n to get consistency in GLS, whereas in adaptive PCR this is not needed. But in financial markets we expect $r_n \to \infty$ so this may not be of an issue between two estimators in financial markets.

3.5 Uniform Consistency of General Estimator

We provide now our main theorem of the paper. This shows that each row of the precision matrix in general case can be estimated, uniformly over all rows, consistently through our estimator. Note that the main diagonal terms rate determine the rate of convergence. The off-diagonal terms estimation is consistent but converges at a faster rate than the main diagonal terms. Denote each row of Θ as Θ_j', which is a $1 \times p$ vector. As an example, first row can be represented as $\Theta_1' := (\frac{1}{\tau_1^2}, \frac{-\alpha_1^{*'}}{\tau_1^2})$. Let Θ_j be the column representation of the row Θ_j'. Let the row estimator be $\tilde{\Theta}_j'$, and for the first row this is $\tilde{\Theta}_1' := (\frac{1}{\tilde{\tau}^2}, \frac{-\tilde{\alpha}_1'}{\tilde{\tau}_1^2})$.

Theorem 1: Under Assumptions 1-5, with $r_n/\delta_n \to 1$

$$\begin{aligned}\max_{1 \le j \le p} \|\tilde{\Theta}_j' - \Theta_j'\|_2 &= O_p\Big(K \frac{\sqrt{p \log p}}{\sqrt{n}} r_{w1}\Big) + O_p(d_{2n} r_{w2}) \\ &+ O_p\left(\max\Big(\frac{K^{3/2}}{n^{1/2}}, K \frac{\sqrt{\max(\log p, \log n)}}{n^{1/2}}\Big)\right) = o_p(1).\end{aligned}$$

Remarks.

1. First, in the general case, we see that if $p/n \to [0,\infty)$, we may obtain consistency and the price that we pay in terms of number of assets, p is just $\sqrt{\log p}$ at the first glance, as can be seen in Assumption 5. However, we have terms such as $\bar{r}, \bar{\Psi}, \bar{\eta}$, which may depend on p, and there will be analysis of these terms in adaptive PCR, and GLS cases.

2. We see that through definition of r_{w1}, and r_{w2} one of the keys to consistent estimation is the ratio $K/\bar{\xi}$, and this ratio needs to go to zero. Therefore, we cannot have a large hidden factor system and expect consistency. The key is a tradeoff between two terms. With low signal-to-noise ratio, the number of hidden factors has to be small.

3. Note that the rate for Theorem is for all linear estimators in Section 3, except GLS when $p-1>n$, the interpolating estimator. This is explained in detail, below in Corollary 6.

4. One interesting question is what if $d_{2n}=o(1)$ which shows asymptotically factor loadings are not correlated. This may be representative of a very well balanced portfolio to diversify risk. Then by Step 3 in the proof of Lemma 2(i) and Lemma 1(ii) with r_{w2} definition, under Assumptions 1-5, we can see in the proof of Theorem 1

$$\max_{1\le j\le p} \|\tilde{\Theta}'_j - \Theta'_j\|_2 = o_p(1).$$

   Also, we may get a better rate than the one in Theorem 1. If we add Assumptions 5(ii)-(iii) to the ones in Theorem 1(i), we obtain consistency in Theorem 1(ii).

3.5.1 Adaptive PCR Estimator

In case of adaptive PCR, terms with wpa1 $\bar{r} \le K$, $\bar{\Psi}/\bar{\eta} \xrightarrow{p} 1$, $\bar{\eta} \ge \delta_n$ with wpa1. So Theorem 1 rate with $\hat{\Theta}'_{PCR,j}$ representing the estimator of j^{th} row of Θ with adaptive PCR estimator.

Corollary 4

(i) Let $p/n \to [0,\infty)$ and with Assumptions 1-5(i), and with $r_n/\delta_n \to 1$

$$\begin{aligned} \max_{1\le j\le p} \|\hat{\Theta}'_{PCR,j} - \Theta'_j\|_2 &= O_p\left(K\sqrt{\log p}\max(\frac{\sqrt{\log n + K}}{\sqrt{n r_n}}, \frac{\sqrt{K}}{\sqrt{\bar{\xi}\delta_n}}) \right) \\ &+ O_p\left(d_{2n}\max(\frac{\sqrt{\log n + K}}{\sqrt{n r_n}}, \frac{\sqrt{K}}{\sqrt{\bar{\xi}}}) \right) \\ &+ O_p\left(\max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\sqrt{\max(\log p, \log n)}}{n^{1/2}}) \right). \end{aligned}$$

(ii) Let $p/n \to [0,\infty)$ and with Assumptions 1-5, and with $r_n/\delta_n \to 1$

$$\max_{1\le j\le p} \|\hat{\Theta}'_{PCR,j} - \Theta'_j\|_2 = o_p(1).$$

Remarks.

1. The price of consistency is $\sqrt{\log p}$ by number of assets, even when $p > n$. So with $p/n \to (1,\infty)$, to get consistency we need the signal-to-noise ratio dominate the number of factors. There is extensive discussion about consistency and effect of Assumption 5 after the proof of Corollary A.1.

2. There is also no exact sparsity assumption on asset returns' errors covariance as in Fan et al. (2013) hidden factor PCR-based estimator. Also, even with known factors, recent papers imposed sparsity of precision matrix of errors as in Caner et al. (2023). Also these papers do not take into account the signal-to-noise ratio $\bar{\xi}$ directly in rate of convergence results. Their rate of convergence results crucially depends on sparsity, and if there is no sparsity, there will be no consistency when $p > n$ case. Hence our results provide a general linear class of estimators with a better performance in terms of estimation errors.

3. Note that with $d_{2n} = o(1)$, we still get consistency, and we can have a faster rate of convergence compared to Corollary 4.

3.5.2 GLS Estimators

When we consider the sub case of GLS with $p < n-1$, we set with probability approaching 1 (wpa1) $\bar{r} < p$, $\bar{\Psi} = 0$, and $\bar{\eta} \ge c$, where c is a positive constant in Theorem 1 result above. So, denote the j^{th} row of estimator of the precision matrix in GLS, $p-1 < n$ case, as $\hat{\Theta}'_{GLS,j}$

Corollary 5

(i) Set $p/n = a_{2n} \to 0$, and under Assumptions 1-5(i), and $r_n/\delta_n \to 1$

$$\begin{aligned}
\max_{1\le j\le p} \|\hat{\Theta}'_{GLS,j} - \Theta'_j\|_2 &= O_p\left(\max(K\sqrt{a_{2n} l o g p}, 1)\max(\frac{\sqrt{\log n + p}}{\sqrt{n}}, \frac{\sqrt{K}}{\sqrt{\bar{\xi}}})\right) \\
&+ O_p\left(d_{2n}\max(\frac{\sqrt{\log n + p}}{\sqrt{n}}, \frac{\sqrt{K}}{\sqrt{\bar{\xi}}})\right) \\
&+ O_p\left(\max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\sqrt{\log n}}{n^{1/2}})\right).
\end{aligned}$$

(ii) Set $p/n = a_{2n} \to 0$, and under Assumptions 1-5, and $r_n/\delta_n \to 1$

$$\max_{1\le j\le p} \|\hat{\Theta}'_{GLS,j} - \Theta'_j\|_2 = o_p(1).$$

Remarks.

- Note that unless $p/n \to 0$, it is not clear that GLS will be consistent. This is a major handicap compared with adaptive PCR, since adaptive PCR is consistent even when $p/n = \gamma \in (0,1)$. Also, we see that with $K < p$, the rate of convergence of adaptive PCR in case of $p/n \to 0$ will be better than GLS here, see also Remark 3 after Corollary A.2 . Only in finite samples GLS, may have an advantage due to nuisance parameter $\bar{\Psi} = 0$ in GLS. GLS can work well only when p is very small compared with n. There is also only no tuning parameter issue in GLS, at $p-1 < n$, the formula is equivalent to OLS.

- Next we consider what if $d_{2n} \to 0$, then we still get consistency and may have a better rate of convergence.

Next, we consider GLS estimator when $p-1 > n$. Note that in this scenario main diagonal term estimation is done differently since GLS is an interpolating estimator. We prove this in Corollary A.2. In the proof of Theorem 1, it is clear that the rate of convergence should be the slower of the main diagonal or off-diagonal terms. Then since the main diagonal term estimation is tied to Lemma A.3 rate, denoting the estimator of the j^{th} row of precision matrix as $\hat{\Theta}_{INT,j}$.

Corollary 6

(i) Under Assumptions 1-4, and $\min_{1\le j\le p} re(\Sigma_{U_{-j}}) > Cn$ and $\tilde{U}_{-j}$ matrix has independent entries, $r_n/\delta_n \to 1$

$$\begin{aligned}\max_{1\le j\le p} \|\hat{\Theta}'_{INT,j} - \Theta'_{INT,j}\|_2 &= O_p\left(\max(\frac{\sqrt{\log n + n}}{\sqrt{n r_n}}, \frac{\sqrt{K} d_{2n}}{\sqrt{\bar{\xi}}})\right) \\ &+ O_p\left(\max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\sqrt{\log p)}}{n^{1/2}})\right).\end{aligned}$$

(ii) If we add to Assumptions in part Corollary 6(i), Assumption 5(ii), $\frac{\sqrt{K}}{\sqrt{\bar{\xi}}} d_{2n} \to 0$

$$\max_{1\le j\le p} \|\hat{\Theta}'_{INT,j} - \Theta'_{INT,j}\|_2 = o_p(1).$$

3.6 Empirical Application

The objective of the empirical application is to conduct an out-of-sample comparison among estimations. We estimate the Global Minimum Variance (GMV) and Maximum Sharpe Ratio (MSR) portfolios using estimators, and use four commonly used financial metrics: the Sharpe Ratio (SR), average return, portfolio turnover, and portfolio variance. We use data from January 1995 to December 2023, with an out-of-sample period from January 2015 to December 2023. We selected all stocks included in the S&P 500 for at least one month during the out-of-sample period, ensuring the data were available for the entire 1995-2023 period, resulting in 357 stocks. The out-of-sample period includes stock market of 2022 when S&P 500 index was down by 18.1%. We use the same rolling window approach to estimate out-of-sample Sharp Ratio as in Chapter 1 or in Caner et al. (2023) and Callot et al. (2021).

3.6.1 Results

We report the results for Maximum Sharpe Ratio (MSR) portfolio, with and without transaction costs. Results for the low-dimensional cases are presented in Table 3.1 - 3.3, while high-dimensional cases are shown in Tables 3.4 - 3.5. In additionally, we use the same data but with different out-of-sample from January 2005 to December 2023 ($n_I = 120$, $n - n_I = 228$). This subsample yields similar results and is available upon request. In MSR portfolio, our PCR-based method is the best model across all 3 small-to-medium (low-dimensional) portfolios scenarios for SR with transaction cost. Our Inter-GLS achieves the highest SR for two largest size of portfolio. Let us give an example by examining three portfolio of varying sizes for MSR: a small low-dimensional portfolio, moderate-sized portfolio, and large-sized portfolio. In Table 3.1, for a portfolio of 10 stocks ($p = 10, n_I = 240$), PCR-1F has 0.2260 as the best Sharpe Ratio. In Table 3.3 of 200 stocks ($p = 200, n_I = 240$), the highest SR is MSR portfolio with our PCR-7F at 0.2682 and the second best is our PCR-3F at 0.2263. Lastly, in Table 3.5, for the largest portfolio with all 357 available stocks, MSR portfolio has the highest SR using our method Inter-GLS at 0.2203. Notably, the SR of our PCR-based methods remains stable as the portfolio size increases.

We use Ledoit and Wolf (2008) test with circular bootstrap to test the significance of the Sharpe Ratio with transaction cost-winner. Specifically,

$$H_o : SR_{winner} \leq SR_0 \text{ vs } H_a : SR_{winner} > SR_0, \tag{3.18}$$

Where SR_{winner} is the method that has the highest Sharpe Ratio in each table, testing against all other methods and significance is set at threshold of $p = 0.05$. A “ $*$ ” indicates statistical significance from the winner.

To see whether double ascent in Sharpe Ratio is possible in GLS, we analyze a figure: Figure 3.1 shows MSR Sharpe Ratio with transaction cost. In figure, we include the performance of Inter-GLS, Nodewise and PCR-3F methods. We observe a double ascent pattern in the GLS estimator's performance in terms of the Sharpe Ratio. For example, in MSR portfolio, with the in-sample window size set at $n_I = 240$, in low-dimensional case, the Sharpe Ratio starts at 0.2063. As the portfolio size increases to 200 stocks, the Sharpe Ratio declines to -0.0471. In the high-dimensional case, where the number of stocks exceeds 240, the SR increases to 0.2203 for a portfolio of 357 stocks.

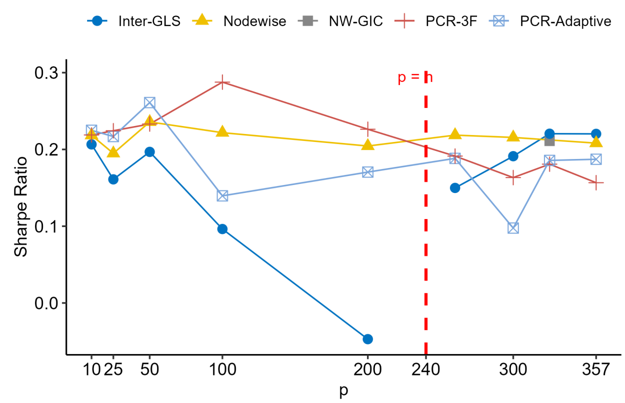

Figure 3.1: MSR Sharpe Ratio vs p

Table 3.1: Monthly Portfolio Performance of 10 Stocks, $n_I = 240$, $n - n_I = 108$

Estimator	**Maximum Sharpe Ratio**			
	Return	SD	Sharpe	Turnover
without transaction cost				
EW	0.0124	0.0647	0.1919	0.0600
RNW-GIC-SF	0.0101	0.0526	0.1913	0.0953
RNW-GIC-3F	0.0101	0.0519	0.1945	0.0889
RNW-GIC-5F	0.0102	0.0510	0.2002	0.0876
NW-GIC	0.0106	0.0505	0.2093	0.0888
POET	0.0104	0.0507	0.2051	0.0700
Ledoit-Wolf	0.0100	0.0522	0.1914	0.0927
SF-NL-LW	0.0101	0.0531	0.1895	0.0992
Inter-GLS	0.0104	0.0530	0.1957	0.0929
PCR-1F	0.0109	0.0490	**0.2221**	0.0518
PCR-3F	0.0105	0.0492	0.2142	0.0576
PCR-5F	0.0106	0.0501	0.2125	0.0948
PCR-7F	0.0105	0.0507	0.2075	0.0795
PCR-Adaptive	0.0108	0.0489	0.2211	0.0514
with transaction cost				
EW	0.0114	0.0642	0.1776	NA
RNW-GIC-SF	0.0105	0.0526	0.1993	NA
RNW-GIC-3F	0.0106	0.0518	0.2040	NA
RNW-GIC-5F	0.0107	0.0509	0.2094	NA
NW-GIC	0.0110	0.0504	0.2184	NA
POET	0.0112	0.0502	0.2235	NA
Ledoit-Wolf	0.0105	0.0521	0.2025	NA
SF-NL-LW	0.0107	0.0528	0.2027	NA
Inter-GLS	0.0109	0.0528	0.2065	NA
PCR-1F	0.0111	0.0491	**0.2260**	NA
PCR-3F	0.0108	0.0493	0.2188	NA
PCR-5F	0.0111	0.0500	0.2209	NA
PCR-7F	0.0109	0.0507	0.2160	NA
PCR-Adaptive	0.0110	0.0491	0.2249*	NA

Table 3.2: Monthly Portfolio Performance of 100 Stocks, $n_I = 240$, $n - n_I = 108$

Estimator	**Maximum Sharpe Ratio**			
	Return	SD	Sharpe	Turnover
without transaction cost				
EW	0.0107	0.0544	0.1965	0.0537
RNW-GIC-SF	0.0072	0.0460	0.1577	0.2934
RNW-GIC-3F	0.0089	0.0463	0.1912	0.3107
RNW-GIC-5F	0.0101	0.0491	0.2055	0.3604
NW-GIC	0.0097	0.0415	0.2338	0.1554
POET	0.0080	0.0462	0.1742	0.2339
Ledoit-Wolf	0.0092	0.0540	0.1711	0.5635
SF-NL-LW	0.0107	0.0523	0.2054	0.4673
Inter-GLS	0.0065	0.0646	0.1009	1.1027
PCR-1F	0.0089	0.0395	0.2265	0.1632
PCR-3F	0.0113	0.0389	**0.2900**	0.3368
PCR-5F	0.0079	0.0424	0.1862	0.6514
PCR-7F	0.0126	0.0477	0.2649	0.8913
PCR-Adaptive	0.0067	0.0429	0.1561	0.5223
with transaction cost				
EW	0.0100	0.0543	0.1844	NA
RNW-GIC-SF	0.0076	0.0457	0.1654*	NA
RNW-GIC-3F	0.0092	0.0461	0.1988*	NA
RNW-GIC-5F	0.0105	0.0486	0.2167	NA
NW-GIC	0.0092	0.0415	0.2218	NA
POET	0.0083	0.0461	0.1801*	NA
Ledoit-Wolf	0.0095	0.0536	0.1778	NA
SF-NL-LW	0.0113	0.0515	0.2197	NA
Inter-GLS	0.0062	0.0645	0.0963*	NA
PCR-1F	0.0089	0.0397	0.2244	NA
PCR-3F	0.0112	0.0390	**0.2877**	NA
PCR-5F	0.0081	0.0418	0.1939	NA
PCR-7F	0.0126	0.0471	0.2685	NA
PCR-Adaptive	0.0060	0.0431	0.1396*	NA

Table 3.3: Monthly Portfolio Performance of 200 Stocks, $n_I = 240$, $n - n_I = 108$

Estimator	Maximum Sharpe Ratio			
	Return	SD	Sharpe	Turnover
without transaction cost				
EW	0.0110	0.0547	0.2005	0.0555
RNW-GIC-SF	0.0053	0.0442	0.1205	0.2761
RNW-GIC-3F	0.0067	0.0447	0.1496	0.3171
RNW-GIC-5F	0.0072	0.0484	0.1483	0.3996
NW-GIC	0.0093	0.0429	0.2169	0.1517
POET	0.0057	0.0430	0.1336	0.1989
Ledoit-Wolf	0.0048	0.0521	0.0928	0.7320
SF-NL-LW	0.0070	0.0520	0.1354	0.6767
Inter-GLS	-0.0018	0.1168	-0.0158	6.1638
PCR-1F	0.0075	0.0376	0.2002	0.1203
PCR-3F	0.0082	0.0382	0.2151	0.2860
PCR-5F	0.0081	0.0557	0.1462	1.0298
PCR-7F	0.0128	0.0433	**0.2964**	1.0439
PCR-Adaptive	0.0081	0.0435	0.1865	0.4231
with transaction cost				
EW	0.0103	0.0546	0.1888	NA
RNW-GIC-SF	0.0056	0.0441	0.1269	NA
RNW-GIC-3F	0.0069	0.0446	0.1555	NA
RNW-GIC-5F	0.0075	0.0480	0.1564	NA
NW-GIC	0.0088	0.0429	0.2044	NA
POET	0.0059	0.0431	0.1359	NA
Ledoit-Wolf	0.0047	0.0521	0.0908	NA
SF-NL-LW	0.0073	0.0514	0.1419	NA
Inter-GLS	-0.0055	0.1166	-0.0471*	NA
PCR-1F	0.0074	0.0378	0.1956	NA
PCR-3F	0.0086	0.0379	0.2263	NA
PCR-5F	0.0069	0.0560	0.1239	NA
PCR-7F	0.0116	0.0433	**0.2682**	NA
PCR-Adaptive	0.0074	0.0436	0.1705	NA

Table 3.4: Monthly Portfolio Performance of 325 Stocks, $n_I = 240$, $n - n_I = 108$

Estimator	Maximum Sharpe Ratio			
	Return	SD	Sharpe	Turnover
without transaction cost				
EW	0.0108	0.0542	0.1988	0.0562
RNW-GIC-SF	0.0047	0.0442	0.1068	0.2982
RNW-GIC-3F	0.0060	0.0451	0.1324	0.3625
RNW-GIC-5F	0.0062	0.0474	0.1300	0.4650
NW-GIC	0.0095	0.0427	0.2228	0.1393
POET	0.0057	0.0438	0.1306	0.2105
Ledoit-Wolf	0.0023	0.0554	0.0415	0.8955
SF-NL-LW	0.0062	0.0549	0.1124	0.8921
Inter-GLS	0.0108	0.0457	**0.2356**	0.3610
PCR-1F	0.0072	0.0373	0.1933	0.1391
PCR-3F	0.0066	0.0400	0.1660	0.3270
PCR-5F	0.0064	0.0535	0.1191	1.2584
PCR-7F	0.0096	0.0449	0.2146	2.0488
PCR-Adaptive	0.0084	0.0424	0.1983	0.1841
with transaction cost				
EW	0.0101	0.0540	0.1865	NA
RNW-GIC-SF	0.0050	0.0439	0.1144	NA
RNW-GIC-3F	0.0062	0.0449	0.1385	NA
RNW-GIC-5F	0.0064	0.0470	0.1370	NA
NW-GIC	0.0090	0.0427	0.2111	NA
POET	0.0059	0.0439	0.1349	NA
Ledoit-Wolf	0.0022	0.0552	0.0404	NA
SF-NL-LW	0.0062	0.0545	0.1130	NA
Inter-GLS	0.0101	0.0458	**0.2205**	NA
PCR-1F	0.0071	0.0375	0.1898	NA
PCR-3F	0.0071	0.0393	0.1808	NA
PCR-5F	0.0048	0.0537	0.0898	NA
PCR-7F	0.0083	0.0438	0.1905	NA
PCR-Adaptive	0.0079	0.0425	0.1857	NA

Table 3.5: Monthly Portfolio Performance of 357 Stocks, $n_I = 240$, $n - n_I = 108$

Estimator	**Maximum Sharpe Ratio**			
	Return	SD	Sharpe	Turnover
without transaction cost				
EW	0.0106	0.0536	0.1985	0.0559
RNW-GIC-SF	0.0046	0.0432	0.1068	0.2857
RNW-GIC-3F	0.0057	0.0440	0.1297	0.3580
RNW-GIC-5F	0.0056	0.0461	0.1209	0.4507
NW-GIC	0.0093	0.0422	0.2202	0.1333
POET	0.0058	0.0424	0.1364	0.2009
Ledoit-Wolf	0.0015	0.0534	0.0280	0.9073
SF-NL-LW	0.0057	0.0526	0.1078	0.8847
Inter-GLS	0.0106	0.0453	**0.2338**	0.2725
PCR-1F	0.0073	0.0364	0.2004	0.1349
PCR-3F	0.0054	0.0366	0.1470	0.2798
PCR-5F	0.0056	0.0528	0.1056	1.6010
PCR-7F	0.0095	0.0506	0.1870	1.8931
PCR-Adaptive	0.0085	0.0420	0.2020	0.2608
with transaction cost				
EW	0.0099	0.0535	0.1860	NA
RNW-GIC-SF	0.0049	0.0430	0.1136	NA
RNW-GIC-3F	0.0059	0.0439	0.1346	NA
RNW-GIC-5F	0.0058	0.0458	0.1267	NA
NW-GIC	0.0088	0.0423	0.2082	NA
POET	0.0059	0.0424	0.1399	NA
Ledoit-Wolf	0.0015	0.0530	0.0281	NA
SF-NL-LW	0.0057	0.0521	0.1085	NA
Inter-GLS	0.0100	0.0454	**0.2203**	NA
PCR-1F	0.0072	0.0366	0.1961	NA
PCR-3F	0.0057	0.0364	0.1567	NA
PCR-5F	0.0047	0.0528	0.0893	NA
PCR-7F	0.0073	0.0507	0.1436	NA
PCR-Adaptive	0.0079	0.0420	0.1872	NA

REFERENCES

Abadir, K. M. and Magnus, J. R. (2005). *Matrix algebra*, volume 1. Cambridge University Press.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Bing, X., Bunea, F., Strimas-Mackey, S., and Wegkamp, M. (2021). Prediction under latent factor regression: adaptive pcr, interpolating predictors and beyond. *Journal of Machine Learning Research*, 22:1–50.

Bing, X. and Wegkamp, M. (2019). Adaptive estimation of the rank of the coefficient matrix in high dimensional multivariate response regression models. *Annals of Statistics*, 47:3157–3184.

Bunea, F., Strimas-Mackey, S., and Wegkamp, M. (2022). Interpolating predictors in high dimensional factor estimation. *Journal of Machine Learning Research*, 23:1–60.

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

Caner, M. and Kock, A. B. (2018). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Journal of Econometrics*, 203(1):143–168.

Caner, M. and Kock, A. B. (2019). High dimensional linear gmm. *arXiv.1811.08779v1*.

Caner, M., Medeiros, M., and Vasconcelos, G. F. (2023). Sharpe ratio analysis in high dimensions: Residual-based nodewise regression in factor models. *Journal of Econometrics*, 235(2):393–417.

Cover, T. M. (1991). Universal portfolios. *Mathematical finance*, 1(1):1–29.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680.

Hazan, E., Agarwal, A., and Kale, S. (2007). Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2):169–192.

Horn, R. and Johnson, C. (2013). *Matrix Analysis*. Cambridge University Press.

Ledoit, O, M. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621.

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

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

Lettau, M. and Pelger, M. (2020). Estimating latent asset pricing. *Journal of Econometrics*, 218:1–31.

Loh, P.-L. and Wainwright, M. J. (2011). High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Advances in neural information processing systems*, 24.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso.

Meng, J. G., Hu, G., and Bai, J. (2011). Olive: a simple method for estimating betas when factors are measured with error. *Journal of Financial Research*, 34(1):27–60.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research*, 11:2261–2286.

APPENDICES

APPENDIX

A

CHAPTER 1

We start with providing some useful norm inequalities. Let $A_1 : p \times K$, $A_2 : p \times q$ $A_3 : q \times K$, $B_1 : K \times K$,$B_2 : K \times q$ matrix, and generic vector $x : K \times 1$

Lemma A.1

(i) $||A_1 B_1 x||_\infty \leq K^2 ||A_1||_\infty ||B_1||_\infty ||x||_\infty$

(ii) $||A_1 B_1 A_1'||_\infty \leq K^2 ||A_1||_\infty^2 ||B_1||_\infty$

(iii) $||A_1 B_2||_\infty \leq K ||A_1||_\infty ||B_2||_\infty$

(iv) $||A_2 A_3 B_1 x||_\infty \leq q K^2 ||A_2||_\infty ||A_3||_\infty ||B_1||_\infty ||x||_\infty$

Proof of Lemma A.1

(i) and (ii) are direct from Lemma A.1 (i) and (ii) of Caner et al. (2023)

(iii) Let a'_j be the is $1\times K$ row vector of A_1, for $k=1,...,K$, b_k be $K\times 1$ column vector of B_2.

$$
\begin{aligned}
||A_1B_2||_\infty &= \max_{i\le j\le p}\max_{1\le k\le p}|a'_j b_k| \\
&\le \max_{i\le j\le p}\max_{1\le k\le p}||a_j||_1||b_k||_\infty \\
&\le \Big[\max_{i\le j\le p}||a_j||_1\Big]\Big[\max_{1\le k\le p}||b_k||_\infty\Big] \\
&= K||A_1||_\infty||B_2||_\infty \qquad \text{(A.1)}
\end{aligned}
$$

For the first inequality, we apply Hölder's inequality. For the second inequalities, we use l_1 and l_∞-norm relation.

(iv)

$$
\begin{aligned}
||A_2A_3B_1x||_\infty &\le K^2||A_2A_3||_\infty||B_1||_\infty||x||_\infty \\
&\le K^2\{q||A_2||_\infty||A_3||_\infty\}||B_1||_\infty||x||_\infty \\
&\le qK^2||A_2||_\infty||A_3||_\infty||B_1||_\infty||x||_\infty \qquad \text{(A.2)}
\end{aligned}
$$

For the first inequality, we use Lemma A.1(i), and for the second, we use Lemma A.1(ii).

Lemma A.2

This Lemma from Fan et al. (2011)

$$
\begin{aligned}
\mathbb{P}\Bigg[\max_{1\le j_1\le d_x}\max_{1\le j_2\le d_y}|\frac{1}{n}\sum_{1}^{n}X_{j_1,t}Y_{j_2,t}-X_{j_1,t}Y_{j_2,t}|>s\Bigg] \le d_xd_y\Bigg\{&n\exp\left(-\frac{(ns)^{r_4}}{C_1}\right) \\
&+\exp\left(-\frac{(n^2s^2)}{C_2(1+nC_3)}\right) \\
&+\exp\left(-\frac{(ns)^2}{C_4T}\exp\left(\frac{(ns)^{\beta(1-\beta)}}{C_5(\log ns)^\beta}\right)\right)\Bigg\}
\end{aligned}
$$

$0<\beta<1$, and β is defined as $\beta^{-1}:=1.5r_x^{-1}+1.5r_y^{-1}+r_{xy}^{-1}$

Lemma A.3

We now present an important maximal inequality applied to mismeasured factor models in a nodewise setting. Since we have a composite error term, it is necessary to extend Lemma A.3 in Caner et al. (2023) in several respects; however, we still use the standard assumptions in the literature.

(i)

$$\mathbb{P}\left[\max_{1\le j\le p}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n}\epsilon_{l,t}\epsilon_{j,t}-E\epsilon_{l,t}\epsilon_{j,t}\right|>C\sqrt{\frac{\ln p}{n}}\right]=O\left(\frac{1}{p^2}\right)$$

(ii)

$$\mathbb{P}\left[\max_{1\le k\le K}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n}v_{k,t}\epsilon_{j,t}\right|>C\sqrt{\frac{\ln p}{n}}\right]=O\left(\frac{1}{p^2}\right)$$

(iii) Let $v_{k_1,t}, v_{k_2,t}$ represents k_1, k_2 measurements of factors (elements) of the vector v_t

$$\mathbb{P}\left[\max_{1\le k_1\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n}v_{k_1,t}v_{k_2,t}-Ev_{k_1,t}v_{k_2,t}\right|>C\sqrt{\frac{\ln n}{n}}\right]=O\left(\frac{1}{n^2}\right)$$

(iv)

$$\mathbb{P}\left[\max_{1\le j\le p}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n}u_{l,t}u_{j,t}-Eu_{l,t}u_{j,t}\right|>C\left[\sqrt{\frac{\ln p}{n}}+K^2\sqrt{\frac{\ln n}{n}}+K\sqrt{\frac{\ln p}{n}}\right]\right]=O\left(\frac{1}{p^2}+\frac{1}{n^2}\right)$$

(v) Denote $U_{-j}: p-1\times n$, and $\mathscr{E}_{-j}: p-1\times n$ matrix , and let the l^{th} row and t^{th} column element of $U_{-j,l,t}$ and η_j as $n\times 1$ vector

$$\mathbb{P}\left[\max_{1\le j\le p}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n}\mathscr{E}_{-j,l,t}\eta_{j,t}-E[\mathscr{E}_{-j,l,t}\eta_{j,t}]\right|>C\sqrt{\frac{\ln p}{n}}\right]=O\left(\frac{1}{p^2}\right) \tag{A.3}$$

(vi)

$$\mathbb{P}\left[\max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}v_{k,t}\eta_{j,t}-Ev_{k,t}\eta_{j,t}\right|>C\sqrt{\bar{s}}\left(\sqrt{\frac{\ln p}{n}}+K\sqrt{\frac{\ln n}{n}}\right)\right]=O\left(\frac{1}{p^2}+\frac{1}{n^2}\right)$$

(vii)

$$\mathbb{P}\left[\max_{1\le j\le p}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n}U_{-j,l,t}\eta_{j,t}]\right|>C\left[\sqrt{\frac{\ln p}{n}}+K\sqrt{\bar{s}}\left(\sqrt{\frac{\ln p}{n}}+K\sqrt{\frac{\ln n}{n}}\right)\right]\right]=O\left(\frac{1}{p^2}+\frac{1}{n^2}\right)$$

(viii) Let $f_{k_1,t}$ represents $k_1$factors (elements) of the vector f_t ,$v_{k_2,t}$ represents measurement error of the vector f_t

$$\mathbb{P}\left[\max_{1\le k_1\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n} f_{k_1,t}v_{k_2,t} - Ef_{k_1,t}v_{k_2,t}\right| > C\sqrt{\frac{\ln n}{n}}\right] = O(\frac{1}{n^2})$$

(ix)

$$\mathbb{P}\left[\max_{1\le k\le K}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n} f_{k,t}\epsilon_{j,t}\right| > C\sqrt{\frac{\ln p}{n}}\right] = O\left(\frac{1}{p^2}\right)$$

(x)

$$\mathbb{P}\left[\max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} f_{k,t}u_{j,t} - Ef_{k,t}u_{j,t}\right| > C\left[\sqrt{\frac{\ln p}{n}} + K\sqrt{\frac{\ln n}{n}}\right]\right] = O\left(\frac{1}{p^2} + \frac{1}{n^2}\right)$$

(xi) Let $f_{k_1,t}, f_{k_2,t}$ represents k_1, k_2factors (elements) of the vector f_t

$$\mathbb{P}\left[\max_{1\le k_1\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n} f_{k_1,t}f_{k_2,t} - Ef_{k_1,t}f_{k_2,t}\right| > C\sqrt{\frac{\ln n}{n}}\right] = O\left(\frac{1}{n^2}\right)$$

(xii)

$$\mathbb{P}\left[\max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} f_{k,t}\eta_{j,t} - E[f_{k,t}\eta_{j,t}]\right| > C\sqrt{\bar{s}}(\sqrt{\frac{\ln p}{n}} + K\sqrt{\frac{\ln n}{n}})\right] = O\left(\frac{1}{p^2} + \frac{1}{n^2}\right)$$

(xiii)

$$\mathbb{P}\left[\max_{1\le k\le K}\max_{1\le j\le q}\left|\frac{1}{n}\sum_{t=1}^{n} f_{k,t}z_{j,t} - E[f_{k,t}z_{j,t}]\right| > C\left(\sqrt{\frac{\ln q}{n}} + K\sqrt{\frac{\ln n}{n}}\right)\right] = O\left(\frac{1}{q^2} + \frac{1}{n^2}\right)$$

(xiv)

$$\mathbb{P}\left[\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} e_{i,t}\epsilon_{j,t} > C\sqrt{\frac{\ln\max(p,q)}{n}}\right|\right] = O\left(\frac{1}{\max(p,q)^2}\right)$$

(xv)

$$\mathbb{P}\left[\max_{1\le i\le q}\max_{1\le k\le K}\left|\frac{1}{n}\sum_{t=1}^{n} e_{i,t}v_{k,t} > C\sqrt{\frac{\ln q}{n}}\right|\right] = O\left(\frac{1}{q^2}\right)$$

(xvi)

$$\mathbb{P}\left[\max_{1\le i\le q}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n} z_{i,t}\epsilon_{j,t}\right| > C\left[K\sqrt{\frac{\ln p}{n}}+\sqrt{\frac{\ln\max(p,q)}{n}}\right]\right] = O\left(\frac{1}{p^2}\right)$$

(xvii)

$$\mathbb{P}\left[\max_{1\le i\le q}\max_{1\le k\le K}\left|\frac{1}{n}\sum_{t=1}^{n} z_{i,t}v_{k,t}\right| \le C\left[K\sqrt{\frac{\ln n}{n}}+\sqrt{\frac{\ln q}{n}}\right]\right] = O\left(\frac{1}{n^2}+\frac{1}{1^2}\right)$$

(xviii)

$$\mathbb{P}\left[\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} z_{i,t}u_{j,t}\right| > C\left[\sqrt{\frac{\ln\max(p,q)}{n}}+K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right] = O\left(\frac{1}{q^2}+\frac{1}{n^2}\right)$$

(xix)

$$\mathbb{P}\left[\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} z_{i,t}\eta_{j,t}]\right| > C\sqrt{\bar{s}}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right] = O\left(\frac{1}{q^2}+\frac{1}{n^2}\right)$$

Proof of Lemma A.3

(i) This is Lemma A.3(i) of Caner et al. (2023), but replace notation by ϵ

(ii) the proof is similar to Lemma B1(i) of Fan et al. (2011), but replace $f_{k,t}$ with v_{kt}

(iii) This is Lemma A.3(i) of Caner et al. (2023), but replace notation by v

(iv) The proof will involve several steps and this is due to the measurement errors in factor models $f_{k,t}$. First consider in matrix form: Note $X = (f_1,...,f_n) : K \times n$, $V = (v_1,...,v_t) : K \times n$ is the measurement error matrix , $U : p \times n$: composite errors matrix , $E : p \times n$: , $B : p \times K$ matrix , $\Sigma_n : p \times p$, $\Sigma_v : K \times K$. Noted $U = \mathscr{E} - BV$

$$\begin{aligned}\frac{UU'}{n} - \Sigma_n &= \frac{(\mathscr{E}-BV)(\mathscr{E}-BV)'}{n} - [\Sigma_e + B\Sigma_v B'] \\ &= B\left[\frac{VV'}{n}-\Sigma_v\right]B' + \left[\frac{\mathscr{E}\mathscr{E}'}{n}-\Sigma_e\right] + \frac{\mathscr{E}V'B}{n} + \frac{BV\mathscr{E}'}{n}\end{aligned}$$

Then, using the triangle inequality in l_∞ norm:

$$\left\|\frac{UU'}{n}-\Sigma_n\right\|_\infty \le \left\|B\left[\frac{VV'}{n}-\Sigma_v\right]B'\right\|_\infty + \left\|\frac{\mathscr{E}\mathscr{E}'}{n}-\Sigma_e\right\|_\infty + 2\left\|\frac{BV\mathscr{E}'}{n}\right\|_\infty \tag{A.4}$$

We evaluate each term in (A.4). By Lemma A.3(iii), we have,

$$\max_{1\le k\le K}\left\|\frac{VV'}{n}-\Sigma_v\right\|_\infty \le C\sqrt{\frac{\ln n}{n}} \tag{A.5}$$

with probability at least $1-1/n^2$. By Lemma A.1(i), (A.3) and $B: p\times K, \Sigma_v : K\times K$

$$\begin{aligned}\left\|B\Big[\frac{VV'}{n}-\Sigma_v\Big]B'\right\|_\infty &\le K^2\|B\|_\infty^2\left\|\frac{VV'}{n}-\Sigma_v\right\|_\infty\\ &\le CK^2\sqrt{\frac{\ln n}{n}} \end{aligned}\tag{A.6}$$

with probability at least $1-1/n^2$. The second term in (A.3)

$$\max_{1\le j\le p}\left\|\frac{\mathscr{E}\mathscr{E}'}{n}-\Sigma_e\right\|_\infty \le C\sqrt{\frac{\ln p}{n}} \tag{A.7}$$

with probability at least $1-1/p^2$. Next, Use Lemma A.1 (iv), with $B: p\times K, V\mathscr{E}' : K\times p$:

$$\left\|\frac{BV\mathscr{E}'}{n}\right\|_\infty \le K\|B\|_\infty\left\|\frac{V\mathscr{E}'}{n}\right\|_\infty \tag{A.8}$$

Use Lemma A.3(ii), we have

$$\max_{1\le j\le p}\left\|\frac{V\mathscr{E}'}{n}\right\|_\infty \le C\sqrt{\frac{\ln p}{n}} \tag{A.9}$$

with probability at least $1-1/p^2$. With the assumption: $\max_{1\le j\le p}\max_{1\le k\le K}|b_{jk}|\le C$. Substitute (A.9) in (A.8):

$$\left\|\frac{BV\mathscr{E}'}{n}\right\|_\infty \le CK\sqrt{\frac{\ln p}{n}} \tag{A.10}$$

with probability at least $1-1/p^2$. Combine (A.6), (A.8), and (A.10) in (A.4)

$$\max_{1\le j\le p}\left\|\frac{UU'}{n}-\Sigma_n\right\|_\infty \le C\Big[K^2\sqrt{\frac{\ln n}{n}}+\sqrt{\frac{\ln p}{n}}+K\sqrt{\frac{\ln p}{n}}\Big] \tag{A.11}$$

with probability $1-O(1/n^2)-O(1/p^2)$. As a consequence,

$$\mathbb{P}\left[\max_{1\le j\le p}\max_{1\le l\le p}|\frac{1}{n}\sum_{t=1}^{n}u_{l,t}u_{j,t}-Eu_{l,t}u_{j,t}|>C[\sqrt{\ln p/n}+K^2\sqrt{\frac{\ln n}{n}}+K\sqrt{\ln p/n}]\right]=O\left(\frac{1}{p^2}+\frac{1}{n^2}\right) \tag{A.12}$$

(v) This is Lemma A.3(ii) in Caner et al. (2023), using notation of U_{-j}

$$\mathbb{P}\left[\max_{1\le j\le p}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n}\mathscr{E}_{-j,l,t}\eta_{j,t}-E[\mathscr{E}_{-j,l,t}\eta_{j,t}]\right|>C\sqrt{\frac{\ln p}{n}}\right]=O\left(\frac{1}{p^2}\right) \tag{A.13}$$

(vi) and (vii)

Using $U_{-j}=\mathscr{E}_{-j}-B_{-j}V$ where $B_{-j}: p-1\times K$ matrix, $V: K\times n$ matrix, $\eta_j: n\times 1$ vector , $U_{-j}: p-1\times n$ composite errors matrix , $\mathscr{E}_{-j}: p-1\times n$ matrix of errors

$$\begin{aligned}
\max_{1\le j\le p}\left\|\frac{U_{-j}\eta_j}{n}\right\|_\infty &= \max_{1\le j\le p}\left\|\frac{[\mathscr{E}_{-j}-B_{-j}V]\eta_j}{n}\right\|_\infty = \max_{1\le j\le p}\left\|\frac{[\mathscr{E}_{-j}-B_{-j}V]\eta_j}{n}+E[\frac{\mathscr{E}_{-j}\eta_j}{n}]-E[\frac{V\eta_j}{n}]\right\|_\infty \\
&\le \max_{1\le j\le p}\left\|\frac{\mathscr{E}_{-j}\eta_j}{n}-E[\frac{\mathscr{E}_{-j}\eta_j}{n}]\right\|_\infty+\max_{1\le j\le p}\left\|B_{-j}\left[\frac{V\eta_j}{n}-E[\frac{V\eta_j}{n}]\right]\right\|_\infty \\
&\le \max_{1\le j\le p}\left\|\frac{\mathscr{E}_{-j}\eta_j}{n}-E[\frac{\mathscr{E}_{-j}\eta_j}{n}]\right\|_\infty+\max_{1\le j\le p}\left[K\|B_{-j}\|_\infty\left\|\frac{V\eta_j}{n}-E[\frac{V\eta_j}{n}]\right\|_\infty\right]
\end{aligned} \tag{A.14}$$

where we use triangle inequality for the first inequality and $\|Bx\|_\infty\le K\|B\|_\infty\|x\|_\infty$ for the second inequality. Next, consider the second term in (A.14)

$$\left\|\frac{V\eta_j}{n}-E[\frac{V\eta_j}{n}]\right\|_\infty=\max_{1\le j\le p}\max_{1\le k\le K}\left|\frac{1}{n}\sum_{i=1}^{n}[v_{k,t}\eta_{j,t}-Ev_{k,t}\eta_{j,t}]\right| \tag{A.15}$$

First, starting with the definition of $\eta_{j,t}$

$$\begin{aligned}
\eta_{j,t}&=u_{j,t}-u'_{-j,t}\kappa_j=[u_{j,t},u'_{-j,t}]\begin{bmatrix}1\\ \kappa_{-j}\end{bmatrix}\\
&=u'_tC_j
\end{aligned} \tag{A.16}$$

Where $C_j:=\begin{bmatrix}1\\ -\kappa_j\end{bmatrix}: p\times 1.$ vector, and $u'_t=[u_{j,},u_{-j,t}]$ which is $1\times p$ row vector, definition of

$u_t = \epsilon_t - B v_t$, where $v_t : K \times 1, B : p \times K$. Next,

$$\begin{aligned}
&\max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}[v_{k,t}\eta_{j,t} - E v_{k,t}\eta_{j,t}]\right| = \max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} v_{k,t}u_t' C_j - E v_{k,t}u_t' C_j\right| \\
&\le \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}[v_{k,t}u_t' - E v_{k,t}u_t']\right\|_\infty \max_{1\le j\le p}||C_j||_1 \\
&= \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}[v_{k,t}(\epsilon_t' - v_t'B') - E v_{k,t}(\epsilon_t' - v_t'B')]\right\|_\infty \max_{1\le j\le p}||C_j||_1 \\
&= \left[\max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n} v_{k,t}\epsilon_t'\right\|_\infty + \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}[v_{k,t}v_t' - E v_{k,t}v_t']B'\right\|_\infty\right]\max_{1\le j\le p}||C_j||_1 \\
&\le \left[\max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n} v_{k,t}\epsilon_t'\right\|_\infty + \left\{K\max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n} v_{k,t}v_t' - E v_{k,t}v_t'\right\|_\infty\right\}||B'||_\infty\right]\max_{1\le j\le p}||C_j||_1 \\
&= \left[\max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} v_{k,t}\epsilon_{j,t}\right| + K\left\{\max_{1\le k\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n} v_{k,t}v_{k_2 t} - E v_{k,t}v_{k_2 t}\right|\right\}||B'||_\infty\right]\max_{1\le j\le p}||C_j||_1
\end{aligned} \tag{A.17}$$

where we use for the first equality and Hölder's inequality for the first inequality; Lemma A.1(i) for the second inequality. i.e $||Bx||_\infty \le K||B||_\infty||x||_\infty$. Next, we consider $\max_{1\le j\le p}||C_j||_1$, by definition

$$\max_{1\le j\le p}||C_j||_1 \le 1 + \max_{1\le j\le p}||\kappa_j||_1 \tag{A.18}$$

To get the bound of $\max_{1\le j\le p}||\kappa_j||_1$, we can use the relationship between $||\kappa_j||_1$, and $||\kappa_j||_2$. First.

$$\frac{\kappa_j'\Sigma_{n,-j,-j}\kappa_j}{\kappa_j'\kappa_j} \ge \text{Eigmin}(\Sigma_{n,-j,-j}) \ge \text{Eigmin}(\Sigma_n) \tag{A.19}$$

Some useful results in Abadir and Magnus (2005). First, by Exercise 8.27 (b) $B\Sigma_v B' \ge 0$ is positive semidefinite when Σ_v is positive definite. Then by Exercise 12.40(b), we have

$$\begin{aligned}
\text{Eigmin}(\Sigma_n) &= \text{Eigmin}(\Sigma_e + B\Sigma_v B) \ge \text{Eigmin}(\Sigma_e) + \text{Eigmin}(B\Sigma_v B') \\
&\ge C
\end{aligned} \tag{A.20}$$

Combine all of the above,

$$||\kappa_j||_2^2 = \kappa_j'\kappa_j \le \frac{\kappa_j'\Sigma_{n,-j,-j}\kappa_j}{\text{Eigmin}(\Sigma_n)} \le C < \infty \tag{A.21}$$

with $\max_{1 \le j \le p} E[u_{j,t}]^2 \le C < \infty$, for a generic constant C. we have

$$\max_{1\le j\le p} \|\kappa_j\|_1 \le \sqrt{\bar{s}} \max_{1\le j\le p} \|\kappa_j\|_2 \le \sqrt{s} C = O(\sqrt{\bar{s}}) \tag{A.22}$$

by (A.18), (A.22). Finally, we have

$$\max_{1\le j\le p} \|C_j\|_1 = O(\sqrt{\bar{s}}) \tag{A.23}$$

For convenience, we get the rate used later on. $\Theta_j = \frac{C_j}{\tau_j^2}$, and by (A.94)

$$\max_{1\le j\le p} \|\Theta_j\|_1 = O(\sqrt{\bar{s}}) \tag{A.24}$$

Combine Lemma A.3(ii), (A.23) in (A.17), we have

$$\mathbb{P}\left[\max_{1\le j\le p} \max_{1\le k\le K} \left| \frac{1}{n} v_{k,t}\eta_{j,t} - E v_{k,t}\eta_{j,t} \right| \le C\left[\sqrt{\frac{\bar{s}\ln p}{n}} + K\sqrt{\frac{\bar{s}\ln n}{n}}\right]\right] \ge 1 - O\left(\frac{1}{p^2}\right) - O\left(\frac{1}{n^2}\right) \tag{A.25}$$

Combine (A.14), (A.13) in (A.25)

$$\max_{1\le j\le p} \left\| \frac{U_{-j}\eta_j}{n} \right\|_\infty \le C\left[\sqrt{\frac{\ln p}{n}} + K\sqrt{\frac{\bar{s}\ln p}{n}} + K^2\sqrt{\frac{\bar{s}\ln n}{n}}\right] \tag{A.26}$$

with probability at least $1 - O(1/p^2) - O(1/n^2)$.

(viii) The follows from Theorem A.1 and Assumption 3 provides tail probability through the same algebra as in p.3346 Fan et al. (2011)

(ix) This is Lemma A.3(iii) in Caner et al. (2023), but replace notation by ϵ

(x) Start with the definition of: $u_t = \epsilon_t - B v_t$, where $u_t : p \times 1, B : p \times K, v_t : k \times 1$

$$
\begin{aligned}
&\max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}u_{j,t} - Ef_{k,t}u_{j,t}]\right| = \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}u_t' - Ef_{k,t}u_t']\right\|_\infty \\
&= \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}(\epsilon_t' - v_t'B') - Ef_{k,t}(\epsilon_t' - v_t'B')]\right\|_\infty \\
&\le \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}\epsilon_t'\right\|_\infty + \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}(f_{k,t}v_t' - Ef_{k,t}v_t')B'\right\|_\infty \\
&\le \max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}\epsilon_t'\right\|_\infty + K\max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}v_t' - Ef_{k,t}v_t']\right\|_\infty \|B'\|_\infty \\
&\le \max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}\epsilon_{j,t}\right| + K\left\{\max_{1\le k\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}v_{k_2t} - Ef_{k,t}v_{k_2t}]\right|\right\}\max_{1\le k\le K}\max_{1\le j\le p}|b_{j,k}|
\end{aligned}
\tag{A.27}
$$

where the first equality use definition, we use (A.2) i.e $\|A_1x_1\|_\infty \le K\|A_1\|_\infty\|x_1\|_\infty$ for the third inequality. Then by Lemma A.3(viii) and (ix), and assumption of factor loading $\max_{1\le k\le K}\max_{1\le j\le p}|b_{j,k}| \le C$

$$
\max_{1\le k\le K}\max_{1\le l\le p}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}u_{j,t} - Ef_{k,t}u_{j,t}]\right| \le C\left[\sqrt{\frac{\ln p}{n}} + K\sqrt{\frac{\ln n}{n}}\right] \tag{A.28}
$$

with probability at least $1 - 1/p^2 - 1/n^2$

(xi) This is Lemma A.3(iv) in Caner et al. (2023). Let $f_{k_1,t}, f_{k_2,t}$ represent k_1, k_2 factors of the vector f_t

$$
\mathbb{P}\left[\max_{1\le k_1\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n}f_{k_1,t}f_{k_2,t} - Ef_{k_1,t}f_{k_2,t}\right| > C\sqrt{\frac{\ln n}{n}}\right] = O\left(\frac{1}{n^2}\right)
$$

(xii) This is similar to Lemma A.3(vi) with the roles of f and v are interchange in proof.

$$
\mathbb{P}\left[\max_{1\le k\le K}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}\eta_{jt} - Ef_{k,t}\eta_{jt}\right| \ge C\times\bar{s}\left(\left[\sqrt{\frac{\ln p}{n}} + K\sqrt{\frac{\ln n}{n}}\right]\right)\right] = O_p\left(\frac{1}{p^2} + \frac{1}{n^2}\right) \tag{A.29}
$$

First, by and Cauchy-Schwartz inequality:

$$
\max_{1\le j\le p}\max_{1\le k\le K} E[f_{k,t}\eta_{j,t}] = O(1) \tag{A.30}
$$

then,

$$\max_{1\le j\le p}\left\|\frac{X\eta_j}{n}\right\|_\infty = C\left[\sqrt{\frac{\bar{s}\ln p}{n}}+K\sqrt{\frac{\bar{s}\ln n}{n}}\right]+C \tag{A.31}$$

with probability at least $1-1/p^2-1/n^2$

(xiii) Note $X=(f_1,f_2,..f_n)$: $K\times n$ matrix, Z is $q\times n$, β_j : $k\times 1$ vector. Let for $k=1,...,K$, f_k be j^{th} row vector of X, $k=1,..,K$. Next, by definition, $f_t* = f_t - v_t$where $f_t^* \perp\!\!\!\perp v_t$, where $v_t=(v_{1,t},...,v_{k,t})'$ is $k\times 1$ measurement vector

$$\begin{aligned}
&\max_{1\le k\le K}\left\|\frac{XZ'}{n}-E[\frac{XZ'}{n}]\right\|_\infty = \max_{1\le k\le K}\max_{1\le l\le q}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}z_{l,t}-Ef_{k,t}z_{l,t}]\right| \\
&= \max_{1\le k\le K}\max_{1\le l\le q}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}[(f_t'-v_t')\beta_l+e_{lt}])-Ef_{k,t}[(f_t'-v_t')\beta_l+e_{lt}]]\right| \\
&\le \max_{1\le k\le K}\max_{1\le l\le q}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k,t}e_{l,t}-Ef_{k,t}e_{l,t}]\right| \\
&+ \max_{1\le k\le K}\max_{1\le l\le q}\left|\frac{1}{n}\sum_{t=1}^{n}\left\{[f_{k,t}f_t'-Ef_{k,t}f_t']+[f_{k,t}v_t'-Ef_{k,t}v_t']\right\}\beta_l\right|
\end{aligned} \tag{A.32}$$

Since $\{f_t\}$ and $\{e_t\}$ are independent for all $k=1,..,K$ and $l=1,...,q$, we have $E[f_{k,t}e_{l,t}]=0$. Next, consider

$$\begin{aligned}
||f_{kt}v_t'\beta_l||_\infty &\le \Big[\max_{1\le k\le K}||f_{kt}v_t'||_1\Big]||\beta_l||_\infty \\
&\le \Big[K||f_{kt}v_t'||_\infty\Big]||\beta_l||_\infty
\end{aligned} \tag{A.33}$$

where use the Hölder's inequality and the relation between l_1 and l_∞. Combine with (A.32), to get :

$$\begin{aligned}
\max_{1\le k\le K}\left\|\frac{XZ'}{n}-E[\frac{XZ'}{n}]\right\|_\infty \le& \max_{1\le k\le K}\max_{1\le l\le q}\left|\frac{1}{n}\sum_{t=1}^{n}f_{k,t}e_{l,t}\right| \\
&+K\max_{1\le k_1\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k_1,t}f_{k_2,t}-Ef_{k_1,t}f_{k_2,t}]\right|\max_{1\le j\le q}||\beta_l||_\infty \\
&+K\max_{1\le k\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n}[f_{k_1,t}v_{k_2,t}-Ef_{k_1,t}v_{k_2,t}]\right|\max_{1\le j\le q}||\beta_l||_\infty
\end{aligned} \tag{A.34}$$

Combine Lemma A.3(ii) and (viii), and Lemma A.3(xi) in (A.34):

$$P\left(\max_{1\le k\le K}\left\|\frac{XZ'}{n}-E[\frac{XZ'}{n}]\right\|_\infty > C\left[\sqrt{\frac{\ln q}{n}}+K\sqrt{\frac{\ln n}{n}}\right]\right)=O\left(\frac{1}{q^2}+\frac{1}{n^2}\right) \tag{A.35}$$

Next,

$$\begin{aligned}
\max_{1\le k_1\le K}\max_{1\le l\le q} E[f_{k_1 t} z_{l t}] &= \max_{1\le k_1\le K}\max_{1\le l\le q} E[f_{k_1 t}(f_{k_2}-v_{k_2})b_l' + f_{k_1,t}e_{l t}] \\
&\le K \max_{1\le k_1\le K}\max_{1\le k_2\le K}\{E[f_{k_1,t}v_{k_2,t}] + E[f_{k_1,t}f_{v_2,t}]\} \max_{1\le l\le q}\max_{1\le k\le K}\|\beta_{l k}\| \\
&= O(1) \qquad \text{(A.36)}
\end{aligned}$$

By Assumption 1 and Caychy-Schwarz inequality:

$$\max_{1\le k_1\le K}\max_{1\le l\le q} E[f_{k_1 t} z_{l t}] = O(1)$$

Combine with (A.35) we have:

$$\mathbb{P}\left[\left\|\frac{XZ'}{n}\right\|_\infty \le C\left[\sqrt{\frac{\ln q}{n}} + K\sqrt{\frac{\ln n}{n}} + C\right]\right] = 1 - \left(\frac{1}{q^2}\right) - \left(\frac{1}{n^2}\right) \qquad \text{(A.37)}$$

Q.E.D

Next, we want to show the inverse of the matrix below is bounded by a constant, such that

$$\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty \le C \qquad \text{(A.38)}$$

with probability at least $1-1/q^2-1/n^2$. Note $XZ' : K \times q$ matrix, we start with

$$\begin{aligned}
\left\|\frac{1}{q}\left\{\frac{XZ'}{n}\frac{ZX'}{n} - E[\frac{XZ'}{n}\frac{ZX'}{n}]\right\}\right\|_\infty &\le \left\|\frac{1}{q}\left\{\frac{XZ'}{n}\frac{ZX'}{n} - E[\frac{XZ'}{n}]E[\frac{ZX'}{n}]\right\|_\infty \right. \qquad \text{(A.39)} \\
&+ \left\|\frac{1}{q}\left\{E[\frac{XZ'}{n}]E[\frac{ZX'}{n}] - E[\frac{XZ'}{n}\frac{ZX'}{n}]\right\}\right\|_\infty \qquad \text{(A.40)}
\end{aligned}$$

where we use the triangle inequality in l_∞-norm, consider the first term (A.39):

$$
\begin{aligned}
&\left\|\frac{1}{q}\left\{\frac{XZ'}{n}\frac{ZX'}{n}-E[\frac{XZ'}{n}]E[\frac{ZX'}{n}]\right\}\right\|_\infty \\
&\leq \left\|\frac{1}{q}\left\{\left(\frac{XZ'}{n}-E[\frac{XZ'}{n}]\right)\left(\frac{ZX'}{n}-E[\frac{ZX'}{n}]\right)+\left(\frac{XZ'}{n}-E[\frac{XZ'}{n}]\right)E[\frac{ZX'}{n}]\right.\right. \\
&\left.\left.+E\frac{XZ'}{n}\left(\frac{ZX'}{n}-E[\frac{ZX'}{n}]\right)\right\}\right\|_\infty \\
&\leq \left\|\left(\frac{XZ'}{n}-E\frac{XZ'}{n}\right)\right\|_\infty\left\|\left(\frac{XZ'}{n}-E\frac{XZ'}{n}\right)\right\|_\infty+2\left\|\left(\frac{XZ'}{n}-E\frac{XZ'}{n}\right)\right\|_\infty\left\|E\frac{ZX'}{n}\right\|_\infty \\
&\leq \left[\sqrt{\frac{\ln q}{n}}+K\sqrt{\frac{\ln n}{n}}\right]\left[\sqrt{\frac{\ln q}{n}}+K\sqrt{\frac{\ln n}{n}}+C\right] \qquad \text{(A.41)}
\end{aligned}
$$

with probability at least $1-O(1/n^2)-O(1/q^2)$, where we use triangle inequality in l_∞ for the first inequality and use Lemma A.1(iii) for that second inequality. For the rate, we use Lemma A.3(xiii), and (A.37). Consider the second term (A.40), we use definition of $z_{i,t}=\beta_i' f_t^*+e_{i,t}=\sum_{k=1}^K \beta_{k,i} f_{k,t}^*$ and $f_t=f_t^*+v_t$

$$
\begin{aligned}
\left\|\frac{1}{q}\left\{E[\frac{XZ'}{n}]E[\frac{ZX'}{n}]-E[\frac{XZ'}{n}\frac{ZX'}{n}]\right\}\right\|_\infty &= \left\|\frac{1}{n^2 q}Var(XZ')\right\|_\infty \\
&= \frac{1}{n^2 q}\max_{1\leq k\leq K}\max_{1\leq i\leq q}\sum_{t=1}^n Var(f_{k,t}z_{i,t}') \\
&= \frac{1}{n^2 q}\max_{1\leq k\leq K}\max_{1\leq i\leq q}\sum_{t=1}^n Var(f_{k,t}[f_t^*\beta_i+e_{i,t}]) \\
&\leq \frac{1}{nq}\max_{1\leq k\leq K}\max_{1\leq i\leq q} Var(f_{k,t}[f_t^*\beta_i+e_{i,t}]) \\
&\leq \frac{1}{nq}\max_{1\leq k\leq K}\max_{1\leq i\leq q} Var(f_{k,t}[\sum_{k_2=1}^K f_{k_2,t}^*\beta_{k_2,i}+e_{i,t}]) \qquad \text{(A.42)}
\end{aligned}
$$

By the definition $f_{k,t}=f_{k,t}^*+v_{k,t}$, $f_t^*\perp\!\!\!\perp v_t$

$$
\begin{aligned}
(A.42) \leq \frac{1}{nq}&\left\{K\max_{1\leq k\leq K}\max_{1\leq k_2\leq K}[Var({f_{k,t}^*}^2)+Var(f_{k,t}^*)Var(v_{k,t})]\max_{1\leq i\leq q}\max_{1\leq k\leq K}|\beta_{k,i}|^2+\right. \\
&\left.+\max_{1\leq k\leq K}\max_{1\leq i\leq q}[Var(f_{k,t}^*)Var(e_{i,t})+Var(v_{k,t})Var(e_{i,t})]\right\} \\
&= O\left(\frac{K}{nq}\right) \qquad \text{(A.43)}
\end{aligned}
$$

by (A.43) and (A.41) to form:

$$\left\| \frac{1}{q} \left\{ \frac{XZ'}{n} \frac{ZX'}{n} - E[\frac{XZ'}{n} \frac{ZX'}{n}] \right\} \right\|_{\infty} \leq \left[\sqrt{\frac{\ln q}{n}} + K \sqrt{\frac{\ln n}{n}} \right] \tag{A.44}$$

with probability at least $1 - O(1/n^2) - O(1/q^2)$. Next, we use $K\sqrt{\ln q} = o(\sqrt{n})$ and $K\sqrt{\ln n} = o(\sqrt{n})$, assume: $\text{Eigmin}(E[XZZ'X]) \geq C > 0$

$$\begin{aligned}
&\mathbb{P}\left[\left\| \frac{1}{q} \left[\frac{XZ'}{n} \frac{ZX'}{n} - E[\frac{XZ'}{n} \frac{ZX'}{n}] \right] \right\|_{l_2} \leq 0.5\text{Eigmin}\left(E \frac{1}{q} \frac{XZ'}{n} \frac{Z'X}{n} \right) \right] \\
&\geq \mathbb{P}\left[K \left\| \frac{1}{q} \left[\frac{XZ'}{n} \frac{ZX'}{n} - E[\frac{1}{q} \frac{XZ'}{n} \frac{ZX'}{n}] \right] \right\|_{\infty} \leq 0.5\text{Eigmin}\left(E \frac{1}{q} \frac{XZ'}{n} \frac{Z'X}{n} \right) \right] \\
&\geq 1 - O\left(\frac{1}{n^2} \right) - O\left(\frac{1}{q^2} \right)
\end{aligned} \tag{A.45}$$

hence, by Lemma A.1 from Fan et al. (2011):

$$\mathbb{P}\left[\text{Eigmin}\left(\frac{1}{q} \frac{XZ'}{n} \frac{ZX'}{n} \right) > \text{Eigmin}\left(\frac{1}{q} E \frac{XZ'}{n} \frac{Z'X}{n} \right) \right] \geq 1 - O\left(\frac{1}{n^2} \right) - O\left(\frac{1}{q^2} \right)$$

Next, define $Q = \left(\frac{1}{q} \frac{XZ'}{n} \frac{ZX'}{n} \right)$, using the relationship between l_2 and l_∞, we have: $||Q||_\infty \leq ||Q||_{l_2}$

$$\begin{aligned}
||Q^{-1}||_\infty \leq ||Q^{-1}||_{l_2} &= \text{Eigmax}(Q^{-1}) \\
&= \frac{1}{\text{Eigmin}(Q)} = C
\end{aligned} \tag{A.46}$$

(xiv) - (xv) We follow the Theorem A.1, and Assumption 3 provides the tail probability through Fan et al. (2011)

$$\mathbb{P}\left[\max_{1 \leq i \leq q} \max_{1 \leq j \leq p} \left| \frac{1}{n} \sum_{t=1}^{n} e_{i,t} \epsilon_{j,t} \right| > C \sqrt{\frac{\ln \max(p,q)}{n}} \right] = O\left(\frac{1}{\max(p,q)^2} \right) \tag{A.47}$$

$$\mathbb{P}\left[\max_{1 \leq i \leq q} \max_{1 \leq k \leq K} \left| \frac{1}{n} \sum_{t=1}^{n} e_{i,t} v_{k,t} \right| > C \sqrt{\frac{\ln q}{n}} \right] = O\left(\frac{1}{q^2} \right) \tag{A.48}$$

(xvi)

Note: $z_{i,t} = \beta_i' f_t^* + e_{i,t}$ $\beta_i : 1 \times K$,$f_t^* : K \times 1$, with assassination $\max_{1\le k\le K}\max_{1\le j\le q} |\beta_{i,k}| \le c$

$$\begin{aligned}
\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} z_{i,t}\epsilon_{j,t}\right| &= \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}(\beta_i' f_t^* + e_{i,t})\epsilon_{j,t}\right| \\
&= \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}(\sum_{k=1}^{K}\beta_{i,k} f_{k,t}^* + e_{i,t})\epsilon_{j,t}\right| \\
&\le K \max_{1\le k\le K}\max_{1\le i\le q}|\beta_{i,k}| \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}(f_{kt} - v_{kt})\epsilon_{j,t}\right| \\
&+ \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n} e_{i,t}\epsilon_{j,t}\right| \\
&\le C\left[K\sqrt{\frac{\ln p}{n}} + \sqrt{\frac{\ln \max(q,p)}{n}}\right]
\end{aligned}$$

with probability at least $1 - O(1/q^2) - O(1/n^2) - O(1/p^2)$, where we use Lemma A.3(ii), (ix) and (xiv) to determine the rate.

(xvii)

$$\begin{aligned}
\max_{1\le i\le q}\max_{1\le k\le K}\left|\frac{1}{n}\sum_{t=1}^{n} z_{i,t} v_{k,t}\right| &= \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}(\beta_i' f_t^* + e_{i,t})v_{k,t}\right| \\
&= \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}(\sum_{k=1}^{K}\beta_{k,i} f_{k,t}^* + e_{i,t})v_{k,t}\right| \\
&\le K \max_{1\le k\le K}\max_{1\le i\le q}|\beta_{k,i}| \max_{1\le k_1\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^{n}(f_{k_1,t} - v_{k_1,t})v_{k_2,t}\right| \\
&+ \max_{1\le k\le K}\max_{1\le i\le q}\left|\frac{1}{n}\sum_{t=1}^{n} v_{k,t} e_{i,t}\right| \\
&\le C\left[K\sqrt{\frac{\ln n}{n}} + \sqrt{\frac{\ln q}{n}}\right]
\end{aligned}$$

with probability at least $1 - O(1/q^2) - O(1/n^2)$, where we use Lemma A.3(ii), (iii) and (viii) to get the rate.

(xviii)

Start with the definition of $u_t' = \epsilon_t' - v_t' B' : v_t : K \times 1, B : p \times K, \epsilon_t : p \times 1$

$$
\begin{aligned}
\max_{1\le j\le p}\left\|\frac{Zu_j}{n}\right\|_\infty &= \max_{1\le j\le p}\left\|\frac{1}{n}\sum_{t=1}^n [z_{i,t}u_t']\right\|_\infty \\
&= \max_{1\le j\le p}\left\|\frac{1}{n}\sum_{t=1}^n [\beta_i' f_t^* + e_{it}][\epsilon_t' - v_t' B']\right\|_\infty \\
&\le \max_{1\le j\le p}\left\|\frac{1}{n}\sum_{t=1}^n e_{i,t}\epsilon_t'\right\|_\infty + \max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^n (\sum_{k=1}^K \beta_{k,i} f_{k,t}^*)\epsilon_t'\right\|_\infty \\
&+ \max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^n e_{i,t} v_t' B'\right\|_\infty + \max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^n (\sum_{k=1}^K \beta_{k,i} f_{k,t}^*) v_t' B'\right\|_\infty \\
&\le \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^n e_{i,t}\epsilon_{j,t}\right| + K\max_{1\le i\le p}\max_{1\le k\le K}|\beta_{k,i}|\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^n f_{k,t}^*\epsilon_{j,t}\right| \\
&+ \max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^n e_{i,t} v_t' B'\right\|_\infty + K\max_{1\le i\le p}\max_{1\le k\le K}|\beta_{k,i}|\max_{1\le k\le K}\left\|\frac{1}{n}\sum_{t=1}^n f_{k,t}^* v_t' B'\right\|_\infty \qquad \text{(A.49)}
\end{aligned}
$$

by Hölder's inequality and the relation between l_1, l_∞

$$
\begin{aligned}
\|f_{k,t} v_t' B'\|_\infty &\le [\|f_{k,t}^* v_t'\|_1]\|B'\|_\infty \\
&\le [K\|f_{k,t} v_t'\|_\infty]\|B'\|_\infty
\end{aligned}
$$

similarly $\|e_{i,t} v_t' B'\|_\infty \le [K\|e_{i,t} v_t'\|_\infty \|B'\|_\infty]$. Therefore,

$$
\begin{aligned}
\max_{1\le j\le p}\left\|\frac{1}{n}\sum_{t=1}^n [z_{i,t}u_t']\right\|_\infty &\le \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^n e_{i,t}\epsilon_{j,t}\right| + K\max_{1\le i\le p}\max_{1\le k\le K}|\beta_{k,i}|\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^n f_{k,t}^*\epsilon_{j,t}\right| \\
&+ K\max_{1\le i\le q}\max_{1\le k\le K}\left|\frac{1}{n}\sum_{t=1}^n [e_{i,t} v_{kt}]\right|\|B'\|_\infty \\
&+ K^2\max_{1\le i\le p}\max_{1\le k\le K}|\beta_{k,i}|\max_{1\le k_1\le K}\max_{1\le k_2\le K}\left|\frac{1}{n}\sum_{t=1}^n [(f_{k_1,t} - v_{k_1,t})v_{k_2,t}]\right|\|B'\|_\infty \\
&\le C\left[\sqrt{\frac{\ln\max(p,q)}{n}} + K\sqrt{\frac{\ln\max(p,q)}{n}} + K^2\sqrt{\frac{\ln n}{n}}\right]
\end{aligned}
$$

with probability at least $1 - o(1/q^2) - O(1/n^2) - O(1-p^2)$, where we use Lemma A.3(ii),(iii) to get the rate.

(xix)

$$
\begin{aligned}
\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}[z_{i,t}\eta_{j,t}]\right| &= \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}z_{i,t}u_t'C_j\right| \\
&\le \max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^{n}[z_{i,t}u_t']\right\|_\infty \max_{1\le j\le p}\|C_j\|_1 \\
&= \max_{1\le i\le q}\left\|\frac{1}{n}sum_{t=1}^{n}[z_{i,t}(\epsilon_t'-v_t'B')]\right\|_\infty \max_{1\le j\le p}\|C_j\|_1 \\
&= \left[\max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^{n}z_{i,t}\epsilon_t'\right\|_\infty + \max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^{n}[z_{i,t}v_t'B']\right\|_\infty\right]\max_{1\le j\le p}\|C_j\|_1 \\
&\le \left[\max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^{n}z_{i,t}\epsilon_t'\right\|_\infty + \{K\max_{1\le i\le q}\left\|\frac{1}{n}\sum_{t=1}^{n}z_{i,t}v_t'\right\|_\infty\}\|B'\|_\infty\right]\max_{1\le j\le p}\|C_j\|_1 \\
&= \left[\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}z_{i,t}\epsilon_{j,t}\right| + K\{\max_{1\le i\le q}\max_{1\le k\le K}\left|\frac{1}{n}\sum_{t=1}^{n}z_{i,t}v_{k,t}\right|\}\|B'\|_\infty\right]\max_{1\le j\le p}\|C_j\|_1
\end{aligned}
\tag{A.50}
$$

where we use Hölder's inequality for the first inequality. By Lemma A.3(vii) we have

$$
\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^{n}[z_{i,t}\eta_{j,t}]\right| \le C\sqrt{\bar{s}}\left[K\sqrt{\frac{\ln p}{n}}+\sqrt{\frac{\ln\max(p,q)}{n}}+K\left[K\sqrt{\frac{\ln n}{n}}+\sqrt{\frac{\ln q}{n}}\right]\right]
$$

with probability at least $1-O(1/q^2)-O(1/n^2)$

Now, we aim to establish an l_1 bound on the nodewise regression estimates. Following the approach in Caner et al. (2023), we define the event as follows.

$$\mathscr{A}_1 := \left\{ 2 \max_{1 \le j \le p} \left\| \frac{\eta'_{xj} \hat{U}'_{-j}}{n} \right\|_\infty \le \lambda_n \right\} \tag{A.51}$$

as in Caner and Kock (2018) we define population adaptive restricted eigenvalue condition, for $j = 1, ..., p$ and let δ_{S_j}

$$\phi^2(s_j) := \min_{\delta_{S_j} \in \mathbb{R}^{p-1}} \left\{ \frac{\delta'_j \Sigma_{n,-j,-j} \delta_j}{||\delta_{S_j}||_2^2} : \delta \in \mathbb{R}^p - \{0\}, ||\delta_{S_j^c}||_1 \le 3\sqrt{s_j} ||\delta_{S_j}||_2 \right\} \tag{A.52}$$

then. empirical version as follows:

$$\hat{\phi}^2(s_j) := \min_{\delta_{S_j} \in \mathbb{R}^{p-1}} \left\{ \frac{\delta'_j (\hat{U}_{-j} \hat{U}'_{-j}/n) \delta_j}{||\delta_{S_j}||_2^2} : \delta \in \mathbb{R}^p - \{0\}, ||\delta_{S_j^c}||_1 \le 3\sqrt{s_j} ||\delta_{S_j}||_2 \right\} \tag{A.53}$$

and for each $j = 1, ..., p$, we define the event

$$\mathscr{A}_{2j} := \{\hat{\phi}^2(s_j) \ge \frac{\phi^2(s_j)}{2}\}$$

Lemma A.4

Lemma A.4 Under $\mathscr{A}_1$, and $\mathscr{A}_{2j}$ with a sequence $\lambda_n > 0$

$$\max_{1 \le j \le p} ||\hat{\kappa}_j - \kappa_j||_1 \le \frac{24 \lambda_n \bar{s}}{\phi^2(\bar{s})} = o_p(\lambda_n \bar{s})$$

This Lemma A.4 of Caner et al. (2023)

Lemma A.5

For the proof of Lemma A.5, we use Assumptions 1-4:

$$\mathbb{P}(\mathscr{A}_1) \geq 1 - O\left(\frac{1}{p^2}\right) - O\left(\frac{1}{n^2}\right)$$

and

$$\kappa_n = C\sqrt{\bar{s}}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]$$

Proof of Lemma A.5

Using $\hat{U}'_{-j} = M_{xz}U'_{-j} = [I_n - P_{xz}]U'_{-j}$, where $P_{xz} = X'(XZ'ZX')^{-1}XZ'Z$, and $\eta_{xj} = [I_n - P_{xz}]\eta_j$. Note that U is $p \times n$ matrix and U_{-j} is the submatrix, which is U without j^{th} row.

$$\begin{aligned} \frac{\hat{U}_{-j}\eta_{xj}}{n} &= \frac{U_{-j}(I_n - P_{xz})'(I_n - P_{xz})\eta_j}{n} \\ &= \frac{U_{-j}\eta_j}{n} - \frac{U_{-j}P'_{xz}\eta_j}{n} - \frac{U_{-j}P_{xz}\eta_j}{n} + \frac{U_{-j}P'_{xz}P_{xz}\eta_j}{n} \end{aligned} \tag{A.54}$$

Next, by triangle inequality in l_∞-norm:

$$\begin{aligned} \left\|\frac{\hat{U}_{-j}\eta_{xj}}{n}\right\|_\infty &= \left\|\frac{U_{-j}\eta_j}{n}\right\|_\infty + \left\|\frac{U_{-j}P'_{xz}\eta_j}{n}\right\|_\infty + \left\|\frac{U_{-j}P_{xz}\eta_j}{n}\right\|_\infty \\ &+ \left\|\frac{U_{-j}P'_{xz}P_{xz}\eta_j}{n}\right\|_\infty \end{aligned} \tag{A.55}$$

Compared to (A.29) of Caner et al. (2023), we have two extra terms since using instruments to estimate the factor loadings and precision matrix of composite errors. Consider the first term in (A.55), by Lemma A.3(vii)

$$\max_{1\leq j\leq p}\left\|\frac{U_{-j}\eta_j}{n}\right\|_\infty \leq C\left[\sqrt{\frac{\ln p}{n}} + K\sqrt{\frac{\bar{s}\ln p}{n}} + K^2\sqrt{\frac{\bar{s}\ln n}{n}}\right] \tag{A.56}$$

with probability at least $1 - O(1/p^2) - O(1/n^2)$. The second term in (A.55), note that $U_{-j}Z'$:

$p-1\times q$ matrix; $ZX: q\times K'$ matrix,$(XZ'ZX')^{-1}: K\times K$ matrix, $X\eta_j: K\times 1$ vector

$$\begin{aligned}\max_{1\le j\le p}\left\|\frac{U_{-j}P'_{xz}\eta_j}{n}\right\|_\infty &= \left\|\frac{U_{-j}Z'ZX'(XZ'ZX')^{-1}X\eta_j}{n}\right\|_\infty\\ &= \frac{1}{q}\max_{1\le j\le p}\left\|\frac{U_{-j}Z'}{n}\frac{ZX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{X\eta_j}{n}\right\|_\infty\\ &\le K^2\left[\max_{1\le j\le p}\left\|\frac{U_{-j}Z'}{n}\right\|_\infty\right]\left[\left\|\frac{ZX'}{n}\right\|_\infty\right]\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]\left[\max_{1\le j\le p}\left\|\frac{X\eta_j}{n}\right\|_\infty\right] \qquad (A.57)\end{aligned}$$

where we use Lemma A.1 (iv) for inequality. Consider the first term in (A.57)

$$\max_{1\le j\le p}\left\|\frac{U_{-j}Z'}{n}\right\|_\infty \le \left\|\frac{UZ'}{n}\right\|_\infty = \max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^n z_{i,t}u_{j,t}\right| \qquad (A.58)$$

by Lemma A.3(xviii)

$$\begin{aligned}\mathbb{P}\left[\max_{1\le i\le q}\max_{1\le j\le p}\left|\frac{1}{n}\sum_{t=1}^n z_{i,t}u_{j,t}\right| \le C\left[\sqrt{\frac{\ln\max(p,q)}{n}}+K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right]\\ \ge 1-O\left(\frac{1}{q^2}+\frac{1}{n^2}+\frac{1}{p^2}\right) \qquad (A.59)\end{aligned}$$

combine with Lemma A.3(xii), (xiii), and (A.46)

$$\begin{aligned}\max_{1\le j\le p}\left\|\frac{U_{-j}P'_{xz}\eta_j}{n}\right\|_\infty &\le K^2\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\\ &\times\left[\sqrt{\frac{\ln q}{n}}+K\sqrt{\frac{\ln n}{n}}+C\right]\times[C]\left[\sqrt{\bar{s}}\left[\sqrt{\frac{\ln p}{n}}+K\sqrt{\frac{\ln n}{n}}\right]+C\right]\\ &= O\left(K^3\sqrt{\frac{\max(\ln p,\ln q)}{n}}+K^4\sqrt{\frac{\ln n}{n}}\right) \qquad (A.60)\end{aligned}$$

with probability at least $1-1/p^2-1/q^2-1/n^2$. Similarly, the third term in (A.55)

$$\max_{1\le j\le p}\left\|\frac{U_{-j}P_{xz}\eta_j}{n}\right\|_\infty \le K^2\left[\max_{1\le j\le p}\left\|\frac{U_{-j}X'}{n}\right\|_\infty\right]\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]\left[\left\|\frac{XZ'}{n}\right\|_\infty\right]\left[\max_{1\le j\le p}\left\|\frac{Z\eta_j}{n}\right\|_\infty\right] \qquad (A.61)$$

consider the first term in (A.61) :

$$\max_{1\le j\le p}\left\|\frac{U_{-j}X'}{n}\right\|_\infty \le \|UX'/n\|_\infty = \max_{1\le j\le p}\max_{1\le k\le K}\left|\frac{1}{n}\sum_{i=1}^n u_{j,t}f_{k,t}\right| \qquad (A.62)$$

by Lemma 3(xii), Lemma (A.3)(vi), Lemma 3(viii) to get:

$$\max_{1\le j\le p}\left\|\frac{U_{-j}P_{xz}\eta_j}{n}\right\|_\infty \le C\sqrt{\bar{s}}\left[K^3\sqrt{\frac{\max(\ln p,\ln q)}{n}}+K^4\sqrt{\frac{\ln n}{n}}\right] \tag{A.63}$$

with probability at least $1-O(1/p^2)-O(1/q^2)-O(1/n^2)$. Next, consider the last term in (A.62),

$$\begin{aligned}
&\max_{1\le j\le p}\left\|\frac{U_{-j}P'_{xz}P_{xz}\eta_j}{n}\right\|_\infty = \max_{1\le j\le p}\left\|\frac{U_{-j}[Z'ZX'(XZ'ZX')^{-1}X][X'(XZ'ZX')^{-1}XZ'Z]\eta_j}{n}\right\|_\infty\\
&=\frac{1}{q^2}\max_{1\le j\le p}\left\|\frac{U_{-j}Z'}{n}\frac{ZX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{Z_jX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{Z\eta_j}{n}\right\|_\infty\\
&\le\frac{K^2}{q^2}\max_{1\le j\le p}\left[\left\|\frac{U_{-j}Z'}{n}\frac{ZX'}{n}\right\|_\infty\right]\left[\left\|\left(\frac{1}{p}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]\left[\left\|\max_{1\le j\le p}\frac{XZ'}{n}\frac{Z\eta_j}{n}\right\|_\infty\right]\\
&\le\frac{K^2}{q^2}\left[q\max_{1\le j\le p}\left\|\frac{U_{-j}Z'}{n}\right\|_\infty\left\|\frac{ZX'}{n}\right\|_\infty\right]\left[K^2\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty^2\left\|\frac{XX'}{n}\right\|_\infty\right]\left[q\left\|\frac{XZ'}{n}\right\|_\infty\max_{1\le j\le p}\left\|\frac{Z\eta_j}{n}\right\|_\infty\right]\\
&\le K^4\left[\max_{1\le j\le p}\left\|\frac{U_{-j}Z'}{n}\right\|_\infty\right]\left[\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{p}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]^2\left[\left\|\frac{XX'}{n}\right\|_\infty\right]\left[\max_{1\le j\le p}\left\|\frac{Z\eta_j}{n}\right\|_\infty\right]
\end{aligned} \tag{A.64}$$

where we use Lemma A.1 (i), for the first inequality $||A_1B_1x||_\infty \le K^2||A_1||_\infty||B_1||_\infty||x||_\infty$, with $A_1 : p\times K$, $B_1 : K\times K$, $x : K\times 1$, then Lemma A.1 (iii) and Lemma A.1(ii) to get the second inequality. By Lemma A.3(iv), assumption $\max_{1\le k\le K}\mathbb{E}[f_{k,t}^2]\le C<\infty$, $\max_{1\le k\le K}\mathbb{E}[v_{k,t}^2]\le C<\infty$ and Cauchy-Schwarz inequality:

$$\max_{1\le k_1\le K}\max_{1\le k_2\le K}|Ef_{k_1,t}f_{k_2,t}+Ev_{k_2,t}v_{k_2,t}|=O(1)$$

combine with Lemma A.3(xi) to get:

$$\max_{1\le j\le p}\left\|\frac{XX'}{n}\right\|_\infty \le C\sqrt{\frac{\ln n}{n}}+C \tag{A.65}$$

with probability at least $1-O(1/n^2)$. Combine (A.38), Lemma A.3(xiii), (xix) to form:

$$
\begin{aligned}
\max_{1\le j\le p}\left|\frac{U_{-j}P'_{xz}P_{xz}\eta_j}{n}\right| &\le K^4\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]^2 \\
&\left[K\sqrt{\frac{\ln q}{n}}+K\sqrt{\frac{\ln n}{n}}+C\right]^2\times[C]\times\left[\sqrt{\frac{\ln n}{n}}+C\right]\left[\sqrt{\frac{\ln q}{n}}+C\right] \\
&\times\left\{\sqrt{\bar{s}}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right\} \\
&=O\left(K^4\sqrt{\bar{s}}\left[K^2\frac{\ln\max(p,n)}{n}+K^4\frac{\ln n}{n}\right]\right) \qquad \text{(A.66)}
\end{aligned}
$$

Combine (A.56), (A.60), (A.63), and (A.64) in order to form:

$$
\left\|\frac{\hat{U}_{-j}\eta_{xj}}{n}\right\|_\infty \le C\sqrt{\bar{s}}\left[K^3\sqrt{\frac{\max(\ln p,\ln q)}{n}}+K^4\sqrt{\frac{\ln n}{n}}\right] \qquad \text{(A.67)}
$$

with probability at least $1-O(1/p^2)-O(1/n^2)$.

Lemma A.6

We use Assumptions 1-5, for $j=1,\dots,p$

$$
\mathbb{P}(\mathscr{A}_{2j})\ge 1-O(1/p^2)-O(1/n^2) \qquad \text{(A.68)}
$$

Where $\mathscr{A}_{2j}:=\{\hat{\phi}^2(s_j)\ge\phi^2(s_j)/2\}$

Proof of Lemma A.6:

We follow the proof procedure from Lemma A.6 of Caner et al. (2023). For each $j=1,\dots,p$, we add and subtract $\delta'_j(U_{-j}U_{-j}/n)\delta_j$

$$
\begin{aligned}
\frac{\delta'_j\hat{U}_{-j}\hat{U}'_{-j}\delta_j}{n} &= \frac{\delta'_j\hat{U}_{-j}\hat{U}'_{-j}\delta_j}{n}-\frac{\delta'_jU_{-j}U'_{-j}\kappa_j}{n}+\frac{\delta'_jU_{-j}U'_{-j}\delta_j}{n} \\
&\ge\frac{\delta'_jU_{-j}U'_{-j}\delta_j}{n}-\left|\frac{\delta'_j\hat{U}_{-j}\hat{U}'_{-j}\delta_j}{n}-\frac{\delta'_jU_{-j}U'_{-j}\delta_j}{n}\right|
\end{aligned}
$$

Then, by adding and subtracting $\frac{\delta_j\Sigma_{n,-j,-j}\delta'_j}{n}$ on the right-hand side of the previous inequality, we obtain:

$$\frac{\delta'_j \hat{U}_{-j} \hat{U}'_{-j} \delta_j}{n} \geq \frac{\delta'_j \Sigma_{n,-j,-j} \delta_j}{n} - \left| \frac{\delta'_j U_{-j} U'_{-j} \delta_j}{n} - \frac{\delta_j \Sigma_{n,-j,-j} \delta'_j}{n} \right|$$
$$- \left| \frac{\delta'_j \hat{U}_{-j} \hat{U}'_{-j} \delta_j}{n} - \frac{\delta'_j U_{-j} U'_{-j} \delta_j}{n} \right| \tag{A.69}$$

We can apply Hölder inequality twice for the third right-side term to get the bound:

$$\left| \frac{\delta'_j \hat{U}_{-j} \hat{U}'_{-j} \delta_j}{n} - \frac{\delta'_j U_{-j} U'_{-j} \delta_j}{n} \right| \leq ||\delta_j||_1^2 \left\| \frac{\hat{U}_{-j} \hat{U}'_{-j}}{n} - \frac{U_{-j} U'_{-j}}{n} \right\|_\infty$$

Similarly, the second term of (A.69)

$$\left| \frac{\delta'_j \hat{U}_{-j} \hat{U}'_{-j} \delta_j}{n} \right| \geq \frac{\delta'_j \Sigma_{n,-j,-j} \delta_j}{n} - ||\delta_j||_1^2 \left[\left\| \frac{\hat{U}_{-j} \hat{U}_{-j}}{n} - \frac{U_{-j} U_{-j}}{n} \right\|_\infty + \left\| \frac{U_{-j} U_{-j}}{n} - \Sigma_{n,-j,-j} \right\|_\infty \right] \tag{A.70}$$

using $\hat{U}'_{-j} = M_{xz} U'_{-j}$, $M^{iv}_{xz} = I - P_{xz} = I - X'(XZ'ZX')^{-1} XZ'Z$, we start considering

$$\frac{\hat{U}_{-j} \hat{U}'_{-j}}{n} - \frac{U_{-j} U'_{-j}}{n} = \frac{U_{-j} [P'_{xz} P_{xz} - P_{xz} - P'_{xz}] U'_{-j}}{n}$$
$$= \frac{U_{-j} P'_{xz} P_{xz} U'_{-j}}{n} - \frac{U_{-j} P_{xz} U'_{-j}}{n} - \frac{U_{-j} P'_{xz} U'_{-j}}{n} \tag{A.71}$$

Next, we use triangle inequality in l_∞ norm

$$\left\| \frac{\hat{U}_{-j} \hat{U}'_{-j}}{n} - \frac{U_{-j} U'_{-j}}{n} \right\|_\infty \leq \left\| \frac{U_{-j} P'_{xz} P_{xz} U'_{-j}}{n} \right\|_\infty + \left\| \frac{U_{-j} P_{xz} U'_{-j}}{n} \right\|_\infty + \left\| \frac{U_{-j} P'_{xz} U'_{-j}}{n} \right\|_\infty$$
$$\leq \left\| \frac{U_{-j} P'_{xz} P_{xz} U'_{-j}}{n} \right\|_\infty + 2 \left\| \frac{U_{-j} P_{xz} U'_{-j}}{n} \right\|_\infty \tag{A.72}$$

We consider the first term in (A.72): Note : $U_{-j} Z'$: $p-1 \times q$ matrix , ZX' : $q \times K$ matrix ,

$XZ'ZX: K \times K$ matrix

$$\begin{aligned}
\left\|\frac{U_{-j}P_{xz}'P_{xz}U_{-j}'}{n}\right\|_\infty &= \left\|\frac{1}{q^2}\frac{U_{-j}Z'}{n}\frac{ZX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{ZU_{-j}'}{n}\right\|_\infty \\
&\leq \frac{K^2}{q^2}\left[\left\|\frac{U_{-j}Z'}{n}\frac{ZX'}{n}\right\|_\infty\right]^2\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty \\
&\leq \frac{K^2}{q^2}\left[q\max_{1\leq j\leq p}\left\|\frac{U_{-j}Z'}{n}\right\|_\infty\max_{1\leq j\leq q}\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[K^2\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty^2\max_{1\leq j\leq p}\left\|\frac{XX'}{n}\right\|_\infty\right] \\
&\leq K^4\left[\max_{1\leq j\leq p}\left\|\frac{U_{-j}Z'}{n}\right\|_\infty\right]^2\left[\max_{1\leq j\leq q}\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]^2\left[\max_{1\leq j\leq p}\left\|\frac{XX'}{n}\right\|_\infty\right] \quad \text{(A.73)}
\end{aligned}$$

where we use Lemma A.1(ii) for the first inequality: $||A_1B_1A_1'||_\infty \leq K^2||A_1||_\infty^2||B_1||_\infty$ with $A_1 = \frac{U_{-j}Z'}{n}\frac{ZX'}{n}$, $B_1 = \left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}$. Then, use Lemma A.1(iii) $||A_1B_2|| \leq q||A_1||_\infty||B_2||_\infty$, $A: p\times q, B_2: q\times$ for the first term and Lemma A.1(i) for the second term in the first inequality to get the second inequality. The rate depends on the first term and by (A.59) to form:

$$\left\|\frac{U_{-j}P_{xz}'P_{xz}U_{-j}'}{n}\right\|_\infty \leq K^4\left[\sqrt{\frac{\ln q}{n}}+K\sqrt{\frac{\ln n}{n}}\right]^2 \quad \text{(A.74)}$$

with probability at least $1-O(1/q^2)-O(1/n^2)$. Next, the second term in (A.72)

$$\begin{aligned}
\left\|\frac{U_{-j}P_{xz}U_{-j}'}{n}\right\|_\infty &= \left\|\frac{1}{q}\frac{U_{-j}X'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{ZU_{-j}'}{n}\right\|_\infty \\
&\leq \frac{K}{q}\left[\max_{1\leq j\leq p}\left\|\frac{U_{-j}X'}{n}\right\|_\infty\right]\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_{l_1}\right]\left[\max_{1\leq j\leq p}\left\|\frac{XZ'}{n}\frac{ZU_{-j}'}{n}\right\|_\infty\right] \\
&\leq \frac{K^2}{q}\left[\max_{1\leq j\leq p}\left\|\frac{U_{-j}X'}{n}\right\|_\infty\right]\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]\left[\max_{1\leq j\leq p}\left\|\frac{XZ'}{n}\frac{ZU_{-j}'}{n}\right\|_\infty\right] \\
&\leq K^2\left[\max_{1\leq j\leq p}\left\|\frac{U_{-j}Z}{n}\right\|_\infty\right]\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]\left[\left\|\frac{XZ'}{n}\right\|_\infty\right]\left[\max_{1\leq j\leq p}\left\|\frac{ZU_{-j}'}{n}\right\|_\infty\right] \quad \text{(A.75)}
\end{aligned}$$

where we use Lemma A.9(iv) in Caner for the first inequality with $A_1: p\times K$, $D_2: K\times K$, $A_1: K\times p$, we have $||A_2D_2A_1||_\infty \leq K||A_2||_\infty||D_2||_{l_1}||A_1||_\infty$. The second inequality is by seeing that transpose of l_1 matrix norm is l_∞-norm for the second term. And, use Lemma A.1(iii):$||A_1B_2||_\infty \leq q||A_1||_\infty||B_2||_\infty$, with $A_1: K\times q, B_2: q\times p$ for the last term to get the last inequality. By (A.72), and (A.46) to form:

$$\left\|\frac{U_{-j}P_{xz}U_{-j}'}{n}\right\|_\infty \le K^2\left[\sqrt{\frac{\ln q}{n}}+K\sqrt{\frac{\ln n}{n}}\right] \tag{A.76}$$

By (A.74), (A.73) and (A.76) to form:

$$\left\|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}-\frac{U_{-j}U_{-j}'}{n}\right\|_\infty \le K^4\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]^2 \tag{A.77}$$

with probability at least $1-O(1/q^2)-O(1/n^2)$. Next,

$$\left\|\frac{U_{-j}U_{-j}'}{n}-\Sigma_{n,-j,-j}\right\|_\infty \le \left\|\frac{UU'}{n}-\Sigma_n\right\|_\infty \le \left[K^2\sqrt{\frac{\ln n}{n}}+\sqrt{\frac{\ln p}{n}}+K\sqrt{\frac{\ln p}{n}}\right] \tag{A.78}$$

with probability at least $1-O(1/p^2)-O(1/n^2)$. As we are using the same population adaptive restricted eigenvalue condition as Caner et al. (2023), by (A.42) from Caner et al. (2023)

$$\frac{||\delta_j||_1^2}{||\delta_{S_j}||_2^2} \le 16s_j \tag{A.79}$$

And then, following Caner et al. (2023) set $\epsilon_n = 16s_j\left[K^4\frac{\ln q}{n}+K^6\frac{\ln n}{n}+\sqrt{\frac{\ln p}{n}}\right]$, we obtain:

$$\mathbb{P}\left[16s_j\left(\left\|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}-\frac{U_{-j}U_{-j}'}{n}\right\|_\infty+\left\|\frac{U_{-j}U_{-j}'}{n}-\Sigma_{n,-j,-j}\right\|_\infty\right)>\epsilon_n\right] \le O(1/p^2)+O(1/n^2)+O(1/q^2) \tag{A.80}$$

Since $\epsilon_n \to 0$, we have $P[\hat{\phi}^2(s_j)/2<\phi^2(s_j)/2]\to 0$.

Q.E.D

Next, we will get a low bound for $\cap_{j=1}^p \mathscr{A}_{2j}$. From the equation above:

$$\left(16s_j\left[\left\|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}-\frac{U_{-j}U_{-j}'}{n}\right\|_\infty+\left\|\frac{U_{-j}U_{-j}'}{n}-\Sigma_{n,-j,-j}\right\|_\infty\right]\right) \subseteq \{\hat{\phi}^2(s_j)\ge\phi^2(s_j)/2\}=\mathscr{A}_{2j} \tag{A.81}$$

the definition of $\Sigma_n, \Sigma_{n,-j,-j}$ and population adaptive restricted eigenvalue condition

$$
\begin{aligned}
&16s_j\left(\left\|\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}-\frac{U_{-j}U_{-j}'}{n}\right\|_\infty+\left\|\frac{U_{-j}U_{-j}'}{n}-\Sigma_{n,-j,-j}\right\|_\infty\right)\\
&\leq 16\bar{s}\left(K^4\left[\max_{1\leq j\leq p}\left\|\frac{UZ'}{n}\right\|_\infty\right]^2\left[\max_{1\leq j\leq q}\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]^2\left[\max_{1\leq j\leq p}\left\|\frac{XX'}{n}\right\|_\infty\right]\right.\\
&\quad\left.+\left\|\frac{UU'}{n}-\Sigma_n\right\|_\infty\right)\\
&\leq \phi(\bar{s})/2\leq\phi^2(s_j)/2 \qquad\qquad \text{(A.82)}
\end{aligned}
$$

So (A.82) implies that, for $j=1,\ldots,p$

$$
\begin{aligned}
&\left[16\bar{s}\left(K^4\left[\max_{1\leq j\leq p}\left\|\frac{UZ'}{n}\right\|_\infty\right]^2\left[\max_{1\leq j\leq q}\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]^2\left[\max_{1\leq j\leq p}\left\|\frac{XX'}{n}\right\|_\infty\right]+\left\|\frac{UU'}{n}-\Sigma_n\right\|_\infty\right)\right.\\
&\qquad\left.\leq\phi^2(\bar{s})/2\right]\subseteq\mathscr{A}_{2j}
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\left[16\bar{s}\left(K^4\left[\max_{1\leq j\leq p}\left\|\frac{UZ'}{n}\right\|_\infty\right]^2\left[\max_{1\leq j\leq q}\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]^2\left[\max_{1\leq j\leq p}\left\|\frac{XX'}{n}\right\|_\infty\right]+\left\|\frac{UU'}{n}-\Sigma_n\right\|_\infty\right)\right.\\
&\qquad\left.\leq\phi^2(\bar{s})/2\right]\subseteq\cap_{j=1}^p\mathscr{A}_{2j}
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\mathbb{P}\left[(\cap_{j=1}^p\mathscr{A}_{2,j})^c\right]&\leq\mathbb{P}\left[16\bar{s}\left(K^4\left[\max_{1\leq j\leq p}\left\|\frac{UZ'}{n}\right\|_\infty\right]^2\left[\max_{1\leq j\leq q}\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]^2\right.\right.\\
&\qquad\left.\left.\left[\max_{1\leq j\leq p}\left\|\frac{XX'}{n}\right\|_\infty\right]+\left\|\frac{UU'}{n}-\Sigma_n\right\|_\infty\right)\leq\phi^2(\bar{s})/2\right]\subseteq\mathscr{A}_{2j}\\
&\leq O(1/p^2)+O(1/q^2)+O(1/n^2)
\end{aligned}
$$

Lemma A.7

Under Assumptions 1-5

$$
\max_{1\leq j\leq p}||\hat{\kappa}_j-\kappa_j||_1=O_p(\lambda_n\bar{s})=o_p(1)
$$

This this Lemma A.7 of Caner et al. (2023)

In the following section, we provide proof of consistency for the estimates of the diagonal elements.

Lemma A.8

Under assumptions 1-5:

$$\max_{1\le j\le p}|\hat{\tau}_j^2-\tau_j^2| = O_p(\sqrt{\bar{s}}\lambda_n) = o_p(1)$$

Proof of Lemma 8: Start the definition of $\hat{\tau}_j^2$. For all $j=1,...,p$

$$\hat{\tau}_j^2 := \frac{1}{n}\hat{u}_j'(\hat{u}_j - \hat{U}_{-j}'\hat{\kappa}_j)$$

and $\tau_j^2 = E[\eta_{j,t}^2]$, with $\eta_{j,t} := u_{j,t} - u_{-j,t}'\kappa_j$ and $\eta_j := (\eta_{j,1},..,\eta_{j,n})' : n\times 1$ vector $\eta_{xj} = M_{XZ}^{IV}\eta_j$; where $M_{xz}^{iv} = I - P_{xz} = I - X'(XZ'ZX')^{-1}XZ'Z$, use the fact that $\hat{u}_j = \hat{U}_{-j}'\kappa_j + \eta_{xj}$

$$\begin{aligned}\hat{\tau}_j^2 &= \frac{(\hat{U}_{-j}'\kappa_j + \eta_{xj})'(\eta_{xj} - \hat{U}_{-j}'(\hat{\kappa}_j - \kappa_j))}{n}\\ &= \frac{\eta_{xj}'\eta_{xj}}{n} - \frac{\eta_{xj}'\hat{U}_{-j}'(\hat{\kappa}_j-\kappa_j)}{n} + \frac{\kappa_j'\hat{U}_{-j}'\eta_{xj}}{n} - \frac{\kappa_j'\hat{U}_{-j}\hat{U}_{-j}'(\hat{\kappa}_j-\kappa_j)}{n}\end{aligned}$$

by the triangle inequality, we get:

$$\begin{aligned}\max_{1\le j\le p}|\hat{\tau}_j^2-\tau_j^2| \le \max_{1\le j\le p}\left|\frac{\eta_{xj}'\eta_{xj}}{n}-\tau_j^2\right| - \max_{1\le j\le p}\left|\frac{\eta_{xj}'\hat{U}_{-j}'(\hat{\kappa}_j-\kappa_j)}{n}\right|\\ + \max_{1\le j\le p}\left|\frac{\kappa_j'\hat{U}_{-j}'\eta_{xj}}{n}\right| - \max_{1\le j\le p}\left|\frac{\kappa_j'\hat{U}_{-j}\hat{U}_{-j}'(\hat{\kappa}_j-\kappa_j)}{n}\right|\end{aligned} \tag{A.83}$$

Consider each term in (A.83) carefully. by definition of M_{xz} we have:

$$\begin{aligned}\frac{\eta_{xj}'\eta_{xj}}{n} = \frac{\eta_j'M_{xz}'M_{xz}\eta_j}{n} &= \frac{\eta_j'[I_n - P_{xz} - P_{xz}' + P_{xz}'P_{xz}]\eta_j'}{n}\\ &= \frac{\eta_j'\eta_j'}{n} - \frac{\eta_j'P_{xz}\eta_j}{n} - \frac{\eta_j'P_{xz}'\eta_j}{n} + \frac{\eta_j'P_{xz}'P_{xz}\eta_j}{n}\end{aligned}$$

by the triangle inequality and $\max_{1\le j\le p}\left|\frac{\eta_j'P_{xz}\eta_j}{n}\right| = \max_{1\le j\le p}\left|\frac{\eta_j'P_{xz}'\eta_j}{n}\right|$, the first term in (A.83)

$$\max_{1\le j\le p}\left|\frac{\eta'_{xj}\eta_{xj}}{n}-\tau_j^2\right|\le\max_{1\le j\le p}\left|\frac{\eta'_j\eta_j}{n}-\tau_j^2\right|+2\max_{1\le j\le p}\left|\frac{\eta'_jP_{xz}\eta_j}{n}\right|$$
$$+\max_{1\le j\le p}\left|\frac{\eta'_jP'_{xz}P_{xz}\eta_j}{n}\right| \tag{A.84}$$

Then, we can get the rate

$$\max_{1\le j\le p}\left|\frac{\eta'_j\eta_j}{n}-\tau_j^2\right|=O_p\left(\sqrt{\frac{\ln p}{n}}\right) \tag{A.85}$$

Then note that: $X\eta_j$: $K\times 1$ vector: and $\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)$: $K\times K$, matrix and XZ' : $K\times q$ and $Z\eta_j : q\times 1$.

$$\begin{aligned}
\max_{1\le j\le p}\left|\frac{\eta'_jP_{xz}\eta_j}{n}\right| &= \max_{1\le j\le p}\left|\frac{\eta'_jX'(XZ'ZX')^{-1}XZ'Z\eta_j}{n}\right|\\
&=\frac{1}{q}\max_{1\le j\le p}\left|\frac{\eta'_jX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{Z\eta_j}{n}\right|\\
&\le\frac{1}{q}\left[\max_{1\le j\le p}\left\|\frac{\eta'_jX'}{n}\right\|_1\right]\max_{1\le j\le p}\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{Z\eta_j}{n}\right\|_\infty\\
&\le\frac{1}{q}\left[K\max_{1\le j\le p}\left\|\frac{\eta'_jX'}{n}\right\|_\infty\right]\left[K\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]\left[\max_{1\le j\le p}\left\|\frac{XZ'}{n}\frac{Z\eta_j}{n}\right\|_\infty\right]\\
&\le\frac{1}{q}\left[K\max_{1\le j\le p}\left\|\frac{\eta'_jX'}{n}\right\|_\infty\right]\left[K\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\left[q\left\|\frac{XZ'}{n}\right\|_\infty\max_{1\le j\le p}\left\|\frac{Z\eta_j}{n}\right\|_\infty\right]\right.\\
&\le K^2\left[\max_{1\le j\le p}\left\|\frac{\eta'_jX'}{n}\right\|_\infty\right]\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]\left[\left\|\frac{XZ'}{n}\right\|_\infty\right]\left[\max_{1\le j\le p}\left\|\frac{Z\eta_j}{n}\right\|_\infty\right]
\end{aligned} \tag{A.86}$$

where we use Hölder's inequality for the first inequality, (A.3) $||B_1x||_\infty\le K||B_1||_\infty||x||_\infty$ and norm inequality between l_1 and l_∞ for the second inequality, (A.2) for the third inequality. $||A_1x_1||_\infty\le q||A_1||_\infty||x_1||_\infty$ with $A_1 : K\times q$, $x_1 : q\times 1$ By Lemma A.3(v), Lemma. A.3(vi), Lemma A.3(viii) we get:

$$\max_{1\le j\le p}\left|\frac{\eta'_jP_{xz}\eta_j}{n}\right|=O_p\left(K^2\sqrt{\bar{s}}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right) \tag{A.87}$$

with probability at least $1-O(1/q^2)-O(1/n^2)$ Next, consider the third term in (A.83)

$$\begin{aligned}
\max_{1\le j\le q}\left|\frac{\eta_j' P_{xz}' P_{xz}\eta_j}{n}\right| &= \frac{1}{q^2}\max_{1\le j\le p}\left|\frac{\eta_j' Z'}{n}\frac{ZX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{Z\eta_j}{n}\right| \\
&\le \frac{K^2}{q^2}\left[\left\|\frac{\eta_j' Z'}{n}\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XX'}{n}\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right] \\
&\le \frac{K^2}{q^2}\left[q\left\|\frac{\eta_j' Z'}{n}\right\|_\infty^2\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|K^2\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\left\|\frac{XX'}{n}\right\|_\infty\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right] \\
&\le K^4\max_{1\le j\le q}\left[\left\|\frac{\eta_j' Z'}{n}\right\|_\infty\right]^2\left[\max_{1\le j\le p}\left\|\frac{ZX'}{n}\right\|_\infty\right]^2\left[\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\right]^2\left[\max_{1\le k\le K}\left\|\frac{XX'}{n}\right\|_\infty\right]
\end{aligned}$$

Where we use (A.55) in Caner for the first inequality, $|xAx|\le[K\|x\|_\infty]^2\|A\|_\infty$ where $x: K\times 1$ vector , $A: K\times K$ matrix. Use lemma A.1 (ii) for the second inequality. By Lemma A.3(xiii), (xix), (A.65), and (A.46) to form :

$$\max_{1\le j\le q}\left|\frac{\eta_j' P_{xz}' P_{xz}\eta_j}{n}\right| = O_p\left(K^4\bar{s}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]^2\right) \tag{A.88}$$

Combine (A.85) (A.87),(A.88) into have the first term on the right side of (A.52)

$$\begin{aligned}
\max_{1\le j\le p}\left|\frac{\eta_{xj}'\eta_{xj}}{n}-\tau_j^2\right| &= \max_{1\le j\le p}\left|\frac{\eta_j'\eta_j}{n}-\tau_j^2\right|+2\max_{1\le j\le p}\left|\frac{\eta_j' P_{xz}\eta_j}{n}\right| \\
&\quad+\max_{1\le j\le p}\left|\frac{\eta_j' P_{xz}' P_{xz}\eta_j}{n}\right| \\
&= O_P\left(K^2\sqrt{\bar{s}}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right)
\end{aligned} \tag{A.89}$$

See that by Lemma A. (with probability approaching one) and (A.14) that

$$\left\|\frac{\eta_{xj}'\hat{U}_{-j}'}{n}\right\|_\infty \le \lambda_n/2$$

Consider the second term on the right side of (A.83):

$$\begin{aligned}
\max_{1\le j\le p}\left|\frac{\eta_{xj}'\hat{U}_{-j}'(\hat{\kappa}_j-\kappa_j)}{n}\right| &\le \max_{1\le j\le p}\left\|\frac{\eta_{xj}'\hat{U}_{-j}'}{n}\right\|_\infty \max_{1\le j\le p}\|(\hat{\kappa}_j-\kappa_j)\|_1 \\
&\le O_p(\lambda_n)O_p(\bar{s}\lambda_n) = O_p(\bar{s}\lambda_n^2)
\end{aligned} \tag{A.90}$$

Consider the third term on the right side of (A.83) , where we use Hölder's inequality:

$$\max_{1\le j\le p}\left|\frac{\kappa_j'\hat{U}_{-j}'\eta_{xj}}{n}\right| \le \Big[\max_{1\le j\le p}\|\kappa_j\|_1\Big]\Big[\max_{1\le j\le p}\left\|\frac{\hat{U}_{-j}\eta_{xj}}{n}\right\|_\infty\Big]$$
$$= O(\bar{s}^{1/2})O_p(\lambda_n) = O_p(\bar{s}^{1/2}\lambda_n) \tag{A.91}$$

Last we consider the fourth term on the right side of (A.83). To get a better rate, we start with the Karush-kuhn-Tucker (KKT) conditions. The following $p-1$ equations form the KKT

$$\lambda_n\hat{\kappa}_j + \frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}\hat{\kappa}_j - \frac{\hat{U}_{-j}\hat{u}_{-j}}{n} = 0_{p-1}$$

Where $\hat{\kappa}_j$ is sub-differential and explained in more detail in p.160 of Caner and Kock (2018), which replaces the gradient in non-differential penalties. Also for all $j = 1,...,p$; $\|\hat{\kappa}_j\|_\infty \le 1$ Use $\hat{u}_{-j} = \hat{U}_{-j}'\kappa_j + \eta_{xj}$, rewrite KKT as

$$\lambda_n\hat{\kappa}_j + \frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}\hat{\kappa}_j - \frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}\kappa_j - \frac{\hat{U}_{-j}\eta_{xj}}{n} = 0_{p-1}$$

Rewrite the above equations as

$$\frac{\hat{U}_{-j}\hat{U}_{-j}'}{n}(\hat{\kappa}_j - \kappa_j) = \frac{\hat{U}_{-j}\eta_{xj}}{n} - \lambda_n\hat{\kappa}_j$$

Then by the triangle inequality

$$\left\|\frac{\hat{U}_{-j}\hat{U}_{-j}}{n}(\hat{\kappa}_j - \kappa_j)\right\|_\infty \le \left\|\frac{\hat{U}_{-j}\eta_{xj}}{n}\right\|_\infty + \lambda_n\|\hat{\kappa}_j\|_\infty$$
$$\le \left\|\frac{\hat{U}_{-j}\eta_{xj}}{n}\right\|_\infty + \lambda_n$$
$$= O_p(\lambda_n) \tag{A.92}$$

Then the fourth term on the right side of (A.83)

$$\max_{1\le j\le p}\left|\frac{\kappa_j'\hat{U}_{-j}\hat{U}_{-j}'(\hat{\kappa}_j - \kappa_j)}{n}\right| \le \max_{1\le j\le p}\||\kappa_j\|_1 \max_{1\le j\le p}\left\|\frac{\hat{U}_{-j}\hat{U}_{-j}'(\hat{\kappa}_j - \kappa_j)}{n}\right\|_\infty$$
$$= O(\bar{s}^{1/2}\lambda_n) \tag{A.93}$$

Where we use Hölder's Inequality and (A.92). Now get to the rate of Lemma A.8, we need to compare the rate of all four terms in (A.83). Among (A.89), (A.90), (A.91) and (A.93) , we can see that (A.91) and (A.93) are the slowest among the four terms on the right side of (A.83) and Assumptions to get the desired result.

Q.E.D

Theorem 1

Theorem 1: Under Assumptions 1-5:

$$||\hat{\Theta}-\Theta||_{l_\infty} := \max_{1\leq j\leq p} ||\hat{\Theta}'_j-\Theta'_j||_1 = \max_{1\leq j\leq p} ||\hat{\Theta}_j-\Theta_j||_1 = O_p\left(\bar{s}\lambda_n\right)$$

Proof of Theorem 1

Since $\Theta := \Sigma_n^{-1}$, by definition of τ_j^2:

$$\tau_j^2 = \frac{1}{\Theta_{j,j}} \geq \frac{1}{\text{Eigmax}(\Theta)} = \text{Eigmin}(\Sigma_n) \geq c > 0 \tag{A.94}$$

we know that $\min_{1\leq j\leq p} \tau_j^2 \geq C$, for a generic constant C. Then,

$$\min_{1\leq j\leq p} \hat{\tau}_j^2 = \min_{1\leq j\leq p} |\hat{\tau}_j^2-\tau_j^2+\tau_j^2| \geq \min_{1\leq j} \tau_j^2 - \max_{1\leq j\leq p} |\hat{\tau}_j^2-\tau_j^2| \tag{A.95}$$

by Lemma A.8, $\min_{1\leq j\leq p} \hat{\tau}_j^2$ is bounded away from zero with probability approach 1. Then

$$\max_{1\leq j\leq p} \left|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right| = \max_{1\leq j\leq p} \frac{|\hat{\tau}_j^2-\tau_j^2|}{\hat{\tau}_j^2\tau_j^2} = O_p(\sqrt{\bar{s}}\lambda_n) = o_p(1) \tag{A.96}$$

where we use Lemma A.8, (A.94), and (A.95). Finally, we want to get the rate:

$$
\begin{aligned}
\max_{1\le j\le p} ||\hat{\Theta}_j - \Theta_j||_1 &= \max_{1\le j\le p} \left|\left| \frac{\hat{C}_j}{\hat{\tau}_j^2} - \frac{C_j}{\tau_j^2} \right|\right|_1 \\
&\le \max_{1\le j\le p} \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| + \max_{1\le j\le p} \left|\left| \frac{\hat{\kappa}_j}{\hat{\tau}_j^2} - \frac{\kappa_j}{\tau_j^2} \right|\right|_1 \\
&= \max_{1\le j\le p} \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| + \max_{1\le j\le p} \left|\left| \frac{\hat{\kappa}_j}{\hat{\tau}_j^2} - \frac{\kappa_j}{\hat{\tau}_j^2} + \frac{\kappa_j}{\hat{\tau}_j^2} - \frac{\kappa_j}{\tau_j^2} \right|\right|_1 \\
&= \max_{1\le j\le p} \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| + \max_{1\le j\le p} \frac{||\hat{\kappa}_j - \kappa_j||_1}{\hat{\tau}_j^2} \\
&+ \max_{1\le j\le p} ||\kappa_j||_1 \max_{1\le j\le p} \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| \\
&= O_p(\bar{s}\lambda_n) = o_p(1)
\end{aligned}
$$

where we use (A.96), Lemma A.7, and (A.22) for the rates, Assumption 5 for the last equality.

For the proof of Theorem 2, we use Lemma A.9 for norm inequalities in Caner et al. (2023). We include it here without proof; details can be found in the original paper.

Lemma A.9

Define generic matrices $A_1 : p \times K$, $A_2 : K \times p$, $D_1 : K \times K$, $D_2 : p \times p$, also define a row vector $x' : 1 \times p$, and also define $p \times p$ matrices A_3, D_3

1. $||A_1 D_1 A_2||_{l_\infty} \leq p K^{1/2} ||A_1||_{l_\infty} ||A_2||_\infty ||D_1||_{l_2}$
2. $||A_2 D_2 A_1||_\infty \leq p ||A_2||_\infty ||A_1||_\infty ||D_2||_{l_\infty}$
3. $||x' A_3 D_3||_1 \leq ||x||_1 ||D_3||_{l_\infty} ||A_3||_{l_\infty}$
4. $||A_2 D_2 A_1||_\infty \leq p ||A_2||_\infty ||A_1||_\infty ||D_2||_{l_1}$

Lemma A.10

Proof of Lemma 10: Under Assumption 1-6

(i) $||\hat{B} - B||_\infty = O_P\left(K \left[K \sqrt{\frac{\ln \max(p,q)}{n}} + K^2 \sqrt{\frac{\ln n}{n}} \right] \right)$

(ii) $||\hat{B} - B||_{l_\infty} = O_p\left(K^{3/2} \left[K \sqrt{\frac{\ln \max(p,q)}{n}} + K^2 \sqrt{\frac{\ln n}{n}} \right] \right)$

(iii) $||B||_{l_\infty} = O_p(K)$

(iv) $||\hat{B}||_{l_\infty} = O_p(K)$

(v) $||B'||_{l_\infty} = ||B||_1 = O_p(p)$

Proof of Lemma A.10

(i) Note that $X : K \times n$ matrix, $X = [f_1, ..., f_t ..., f_n]$, where $f_t : K \times 1$ vector, $u_j : n \times 1$ vector, $XZ' : K \times q$ matrix, $ZX' : q \times K$, $Zu_j : q \times 1$, $XZ'ZX' : K \times K$ matrix. Compare to Caner et al. (2023): $\hat{b}_j - b_j = \left(\frac{XX'}{n} \right)^{-1} \frac{X u_j}{n}$ for $j = 1, ..., p$, with

$$\hat{b}_j - b_j = (XZ'ZX')^{-1} XZ'Z u_j = \frac{1}{q} \left(\frac{1}{q} \frac{XZ'}{n} \frac{ZX'}{n} \right)^{-1} \frac{XZ'}{n} \frac{Z u_j}{n}$$

We consider,

$$\begin{aligned}
\|\hat{B}-B\|_\infty &= \max_{1\le j\le p}\|\hat{b}_j'-b_j'\|_\infty = \max_{1\le j\le p}\|\hat{b}_j-b_j\|_\infty \\
&= \frac{1}{q}\max_{1\le j\le p}\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{Zu_j}{n}\right\|_\infty \\
&\le \left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\right\|_\infty \max_{1\le j\le p}\left\|\frac{Zu_j}{n}\right\|_\infty \\
&\le K\left\|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right\|_\infty\left\|\frac{XZ'}{n}\right\|_\infty \max_{1\le j\le p}\left\|\frac{Zu_j}{n}\right\|_\infty \\
&= K[C]O_p(1)O\left(K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right) \\
&= O\left(K\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right) \qquad \text{(A.97)}
\end{aligned}$$

where we use Lemma A.2 for the first inequality $\|Ax\|_\infty \le q\|A\|_\infty\|x\|_\infty$, with $A=\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}$, $\frac{XZ'}{n}$: $K\times q$ matrix, and $x=\frac{Zu_j}{n}$: $q\times 1$ vector, use Lemma A.1 (iii) for the second inequality, and use (A.35) for the rate.

(ii)

$$
\begin{aligned}
||\hat{B}-B||_{l_\infty} &= \max_{1\le j\le p}||\hat{b}_j'-b_j'||_1 = \max_{1\le j\le p}||\hat{b}_j-b_j||_1 \\
&= \max_{1\le j\le p}\left|\left|(XZ'ZX')^{-1}XZ'Zu_j\right|\right|_1 \\
&= \frac{1}{q}\max_{1\le j\le p}\left|\left|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\frac{XZ'}{n}\frac{Zu_j}{n}\right|\right|_1 \\
&\le \frac{1}{q}\left[\left|\left|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right|\right|_{l_1}\right]\left[\max_{1\le j\le p}\left|\left|\frac{XZ'}{n}\frac{Zu_j}{n}\right|\right|_1\right] \\
&\le \frac{1}{q}\left[\left|\left|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right|\right|_{l_1}\right]\left[K\max_{1\le j\le p}\left|\left|\frac{XZ'}{n}\frac{Zu_j}{n}\right|\right|_\infty\right] \\
&\le \frac{1}{q}\left[K^{1/2}\left|\left|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right|\right|_{l_2}\right]\left[Kq\left|\left|\frac{XZ'}{n}\right|\right|_\infty\max_{1\le j\le q}\left|\left|\frac{Zu_j}{n}\right|\right|_\infty\right] \\
&= K^{3/2}\left[\left|\left|\left(\frac{1}{q}\frac{XZ'}{n}\frac{ZX'}{n}\right)^{-1}\right|\right|_{l_2}\right]\left[\left|\left|\frac{XZ'}{n}\right|\right|_\infty\max_{1\le j\le q}\left|\left|\frac{Zu_j}{n}\right|\right|_\infty\right] \\
&= K^{3/2}[C]O_p(1)\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\Bigg) \\
&= O\left(K^{3/2}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right) \qquad \text{(A.98)}
\end{aligned}
$$

where the first inequality, we use $||Ax||_1\le||A||_{l_1}||x||_1$, for a generic matrix A, and generic vector x. The second inequality uses the norm inequality between l_1 and l_∞ norms (i.e $||x||_1\le dim(x)||x||_\infty$, $\dim(x)$: the dimension of the vector x). The third inequality we use the upper bound of l_1 induced matrix norm in terms of the spectral norm, as in p.365 of Horn and Johnson (2013) and use (A.2) which is $||Ax||_\infty\le q||A||_\infty||x||_\infty$ where $A: K\times q$, $x: q\times 1$ vector. We use (A.46) for the rate

(iii)

$$
||B||_{l_\infty} = \max_{1\le j\le p}||b_j'||_1 = O_p(K) \qquad \text{(A.99)}
$$

by Assumption 6 that $\max_{1\le j\le p}\max_{1\le k\le K}|b_{jk}|\le C$, for a positive constant C and uniformly over $j=1,...,p$, and $k=1,...,K$

(iv) Using the results with assumption 7, we have

$$
||\hat{B}||_{l_\infty} = O_p(K) \qquad \text{(A.100)}
$$

(v)

$$||B||_{l_1} = \max_{1\le k\le K} ||b_k||_1 = O(p) \tag{A.101}$$

Lemma A.11

1. $\text{Eigmin}(B'\Theta B) \ge \frac{cp}{Cr_n}$
2. $\left|\left|\left([\Sigma_{f_t}]^{-1} + B'\Theta B\right)^{-1}\right|\right|_{l_2} = O\left(\frac{r_n}{p}\right)$

Proof of Lemma A.11

(i) This is Lemma A.11 (i) in Caner et al. (2023)

(ii) Using Assumption 4

$$\left|\left|\left([\Sigma_{f_t}]^{-1} + B'\Theta B\right)^{-1}\right|\right|_{l_2} = \text{Eigmax}([\Sigma_{f_t}]^{-1} + B'\Theta B)^{-1} = \frac{1}{\text{Eigmin}([\Sigma_{f_t}]^{-1} + B'\Theta B)} \tag{A.102}$$

Where we use the properties of the spectral norm and $\text{Eigmax}(Q) \ge \text{Eigmin}(Q)$ as $Q \ge 0$

$$\begin{aligned}\text{Eigmin}[[cov f_t]^{-1} + B'\Theta B] &\ge \text{Eigmin}(B'\Theta B)\\ &\ge \frac{cp}{Cr_n}\end{aligned} \tag{A.103}$$

Since for invertible matrix G, we have $\text{Eigmax}(G^{-1}) = \dfrac{1}{\text{Eigmin}(G)}$, combine with(A.102), We have desired result. **Q.E.D**

Starting from this section, we use the symmetrized version $\hat{\Theta}_{sym} := \frac{\hat{\Theta}+\hat{\Theta}'}{2}$

Lemma A.12

Under Assumptions 1-4,6-7

1. $||\hat{B}'\hat{\Theta}_{sym}\hat{B} - B'\Theta B||_{l_2} = O_p\left(pK\bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right)$
2.

$$||\hat{G}||_{l_2} = O_p\left(\frac{r_n}{p}\right)$$

with $\hat{G} := ([\widehat{cov(f_t)}]^{-1} + \hat{B}'\hat{\Theta}_{sym}\hat{B})^{-1}$

Proof of Lemma A.12

(i) We start with simple adding and subtracting $\hat{B} = (\hat{B} - B) + B$, $\hat{\Theta}_{sym} = (\hat{\Theta}_{sym} - \Theta) + \Theta$ and triangle inequality

$$||\hat{B}'\hat{\Theta}_{sym}\hat{B} - B'\Theta B||_\infty \leq ||\Big[(\hat{B} - B) + B\Big]'\Big[(\hat{\Theta}_{sym} - \Theta) + \Theta\Big]\Big[(\hat{B} - B) + B\Big] - B'\Theta B||_\infty \qquad \text{(A.104)}$$

$$\begin{aligned} ||\hat{B}'\hat{\Theta}_{sym}\hat{B} - B'\Theta B||_\infty &\leq ||(\hat{B} - B)'(\hat{\Theta}_{sym} - \Theta)(\hat{B} - B)||_\infty \\ &+ 2||(\hat{B} - B)'(\hat{\Theta}_{sym} - \Theta)B||_\infty + ||(\hat{B} - B)\Theta(\hat{B} - B)||_\infty \\ &+ ||B'(\hat{\Theta}_{sym} - \Theta)B||_\infty + 2||(\hat{B} - B)'\Theta B||_\infty \end{aligned} \qquad \text{(A.105)}$$

Analyze each term in (A.105) and by Lemma A.9 (ii),(iv)

$$\begin{aligned} ||(\hat{B} - B)'(\hat{\Theta}_{sym} - \Theta)(\hat{B} - B)||_\infty &\leq \frac{1}{2}||(\hat{B} - B)'(\hat{\Theta} - \Theta)(\hat{B} - B)||_\infty \\ &+ \frac{1}{2}||(\hat{B} - B)'(\hat{\Theta}' - \Theta)(\hat{B} - B)||_\infty \\ &\leq \frac{p}{2}||\hat{B} - B||_\infty^2 ||\hat{\Theta} - \Theta||_{l_\infty} + \frac{p}{2}||\hat{B} - B||_\infty^2 ||\hat{\Theta}' - \Theta||_{l_1} \\ &= p\left[O_p\left(K\left[K\sqrt{\frac{\ln \max(p,q)}{n}} + K^2\sqrt{\frac{\ln n}{n}}\right]\right)\right]^2 O_p\left(\bar{s}\lambda_n\right) \end{aligned} \qquad \text{(A.106)}$$

Where we use Lemma A.10(i) and Theorem 1 for the rate with

$$||\hat{\Theta} - \Theta||_{l_\infty} = ||\hat{\Theta}' - \Theta||_{l_1} = \max_{1 \leq j \leq p} ||\hat{\Theta}_j - \Theta_j||_1 \qquad \text{(A.107)}$$

Since l_1- norm of transpose of $\hat{\Theta}$ involves row of $\hat{\Theta}$ (hence columns of $\hat{\Theta}'$). Next, analyze the second term in (A.105), use the definition of $\hat{\Theta}_{sym} = \frac{\hat{\Theta} + \hat{\Theta}'}{2}$ and via triangle inequality:

$$\begin{aligned} ||(\hat{B} - B)'(\hat{\Theta}_{sym} - \Theta)B||_\infty &\leq \frac{1}{2}||(\hat{B} - B)'(\hat{\Theta} - \Theta)B||_\infty \\ &+ \frac{1}{2}||(\hat{B} - B)'(\hat{\Theta}' - \Theta)B||_\infty \\ &\leq \frac{p}{2}||\hat{B} - B||_\infty ||B||_\infty ||\hat{\Theta} - \Theta||_{l_\infty} \\ &+ \frac{p}{2}||\hat{B} - B||_\infty ||B||_\infty ||\hat{\Theta}' - \Theta||_{l_1} \\ &= p\, O_p\left(K\left[K\sqrt{\frac{\ln \max(p,q)}{n}} + K^2\sqrt{\frac{\ln n}{n}}\right]\right) O(1) O_p(\bar{s}\lambda_n) \end{aligned} \qquad \text{(A.108)}$$

We use Lemma A.9 (ii) (iv) for the second inequalities, and Assumption 6, Theorem 1, and

(A.107) for the rates. Next, consider the third term in(A.105)

$$
\begin{aligned}
||(\hat{B}-B)\Theta(\hat{B}-B)||_\infty &\leq p||\hat{B}-B||_\infty^2 ||\Theta||_{l_\infty} \\
&= p\left[O_p\left(K\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right)\right]^2 O(\bar{s}^{1/2}) \quad \text{(A.109)}
\end{aligned}
$$

We use Lemma A.9(ii) for the first inequality, Lemma 10 (i) , with

$$
||\Theta||_{l_\infty} := \max_{1\leq j\leq p} ||\Theta_j'||_1 = \max_{1\leq j\leq p} ||\Theta_j||_1 = O(\bar{s}^{1/2}) \quad \text{(A.110)}
$$

Next, consider the fourth term in (A.105), use the symmetry of $\hat{\Theta}_{sym}$ and via triangle inequality:

$$
\begin{aligned}
||B'(\hat{\Theta}_{sym}-\Theta)B||_\infty &\leq \frac{1}{2}||B'(\hat{\Theta}-\Theta)B||_\infty + \frac{1}{2}||B'(\hat{\Theta}'-\Theta)B||_\infty \\
&\leq \frac{p}{2}||B||_\infty^2 ||\hat{\Theta}-\Theta||_{l_\infty} + \frac{p}{2}||B||_\infty^2 ||\hat{\Theta}'-\Theta||_{l_1} \\
&= p[O(1)]^2 O_p(\bar{s}\lambda_n) \quad \text{(A.111)}
\end{aligned}
$$

We use Lemma A.9(ii)(iv) for the inequality, use Theorem 1 and Assumption 6. for the rates. Next, now analyze the fifth term in (A.105)

$$
\begin{aligned}
||(\hat{B}-B)'\Theta B||_\infty &\leq p||\hat{B}-B||_\infty ||B||_\infty ||\Theta||_{l_\infty} \\
&= p\,O_p\left(K\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right)O(1)O(\bar{s}^{1/2}) \quad \text{(A.112)}
\end{aligned}
$$

where we use Lemma A.9(ii) for the inequality, use Lemma A.10(i), (A.122), and Assumption 6 for the rates. Compare five rates from (A.106), (A.108), (A.109), (A.111) and (A.112), with Assumption 5 that $K^6\dfrac{\max(\ln p,\ln q)}{n}=o_p(1)$, $K^8\frac{\ln n}{n}=o_p(1)$, the slowest rate is the maximum of (A.112) and (A.111). Then,

$$
||\hat{B}'\hat{\Theta}_{sym}\hat{B}-B'\Theta B||_\infty = O_p\left(p\max\left(\bar{s}\lambda_n, \bar{s}^{1/2}K\left(K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right)\right)\right)
$$

use $\lambda_n := C\,\bar{s}^{1/2}\left[K^3\sqrt{\frac{\max(\ln p,\ln q)}{n}}+K^4\sqrt{\frac{\ln n}{n}}\right]$, we can get the final rate

$$||\hat{B}'\hat{\Theta}_{sym}\hat{B} - B'\Theta B||_\infty = O_p\left(p\bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right)$$

Then by norm inequality tying spectral norm to $||.||_\infty$ norm in p.365 of Horn and Johnson (2013), and since $\hat{B}'\hat{\Theta}_{sym}\hat{B} - B'\Theta B$ is $K \times K$ matrix

$$\begin{aligned} ||\hat{B}'\hat{\Theta}_{sym}\hat{B} - B'\Theta B||_{l_2} &\leq K||\hat{B}'\hat{\Theta}_{sym}\hat{B} - B'\Theta B||_\infty \\ &= O_p\left(pK\bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right) \end{aligned} \tag{A.113}$$

Q.E.D

(ii) Since $\widehat{\Sigma_{f_t}}^{-1}, (\Sigma_{f_t})^{-1}$, does not invlove the precision matrix estimator, we proceed as in Fan et all. (2011), Lemma B5(ii) B.20 of Fan et al. (2011) provide

$$||\widehat{\Sigma_{f_t}}^{-1} - \Sigma_{f_t}^{-1}||_{l_2} = O_p\left(\sqrt{\frac{\ln n}{n}}\right)$$

Using the above function and (A.113) via triangle inequality, we get

$$\begin{aligned} ||(\Sigma_{f_t}^{-1} + \hat{B}\hat{\Theta}_{sym}\hat{B}) - (\Sigma_{f_t}^{-1} + B'\Theta B)||_{l_2} & \\ \leq ||\Sigma_{f_t}^{-1}(\Sigma_{f_t}^{-1}||_{l_2} + ||\hat{B}\hat{\Theta}_{sym}\hat{B} - B'\Theta B||_{l_2} & \\ = O_p\left(pK\bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right) & \end{aligned} \tag{A.114}$$

Since By Assumption 7: $\bar{s}l_n = \bar{s}r_n^2 K\bar{s}^{3/2}\left(\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right) = o_p(1)$, therefore, $r_n K\bar{s}^{3/2}\left(\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}}\right.\right.$ $\left.\left.K^4\sqrt{\frac{\ln n}{n}}\right]\right) = o_p(1)$. So,

$$pK\bar{s}^{3/2}\left(\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right) = o_p\left(\frac{p}{r_n}\right) \tag{A.115}$$

Therefore, we have the rate of (A.129) :

$$||(\Sigma_{f_t}^{-1} + \hat{B}\hat{\Theta}_{sym}\hat{B}) - (\Sigma_{f_t}^{-1} + B'\Theta B)||_{l_2} = o_p\left(\frac{p}{r_n}\right) \tag{A.116}$$

Then using Lemma A.1(i) of Fan and Li (2001) which is for A be an $m \times m$ random matrix, B is an $m \times m$ deterministic matrix and if there exists a positive sequence $\{c_t\}_{T=1}^{\infty}$ such that $\text{Eigmin}(B) > c_t$ Then,

$$P(\text{Eigmin}(A) \geq 0.5c_t) \geq P(||A - B||_{l_2} \leq 0.5c_T)$$

Let $A = [\Sigma_{f_t}]^{-1} + B'\Theta B$, $B = [\hat{\Sigma}_{f_t}]^{-1} + \hat{B}'\hat{\Theta}_{sym}\hat{B}$, with (A.103) and (A.116)

$$\text{Eigmin}([\hat{\Sigma}_{f_t}]^{-1} + \hat{B}'\hat{\Theta}_{sym}\hat{B}) \geq \frac{cp}{Cr_n}$$

with probability approaching 1 with $r_n << p$ as in Assumption 7. Note that for invertible matrix A, $||A^{-1}||_{l_2} = \text{Eigmax}(A^{-1}) = \dfrac{1}{\text{Egimin}(A)}$:

$$||[\hat{\Sigma}_{f_t}]^{-1} + \hat{B}'\hat{\Theta}_{sym}\hat{B})^{-1}||_{l_2} = O_p\left(\frac{r_n}{p}\right)$$

Q.E.D

We restate the definition of major terms that are used:

$$G := (\Sigma_{f_t}^{-1} + B'\Theta B)^{-1} \tag{A.117}$$

$$\hat{G} := ([\hat{\Sigma}_{f_t}]^{-1} + \hat{B}'\hat{\Theta}_{sym}\hat{B})^{-1} \tag{A.118}$$

Next, note that

$$\hat{L} := \hat{B}\hat{G}\hat{B}' \, ; \, L := BGB' \tag{A.119}$$

and,

$$l_n := r_n^2 K^{5/2} \bar{s}^{3/2} \left[K^3 \sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4 \sqrt{\frac{\ln n}{n}} \right] \tag{A.120}$$

We have the following lemma that leads to the proof of Theorem 2.

Lemma A.13

$$\|\hat{L} - L\|_{l_\infty} = O_p(l_n) = o_p(1)$$

Proof of Lemma A.13: Start with adding and subtracting and via triangle inequality

$$\begin{aligned} \|\hat{L} - L\|_{l_\infty} \leq & \|(\hat{B} - B)\hat{G}(\hat{B} - B)'\|_{l_\infty} + \|(\hat{B} - B)\hat{G}B'\|_{l_\infty} \\ & + \|B\hat{G}(\hat{B} - B)'\|_{l_\infty} + \|B\hat{G}B' - BGB'\|_{l_\infty} \end{aligned} \tag{A.121}$$

Consider the first term in (A.121):

$$\begin{aligned} & \|(\hat{B} - B)\hat{G}(\hat{B} - B)'\|_{l_\infty} \leq pK^{1/2}\|\hat{B} - B\|_{l_\infty}\|\hat{G}\|_{l_2}\|\hat{B} - B\|_\infty \\ & = pK^{1/2} O_p\left(K^{3/2}\left[K\sqrt{\frac{\ln \max(p,q)}{n}} + K^2\sqrt{\frac{\ln n}{n}} \right]\right) O_p\left(\frac{r_n}{p}\right) O_P\left(K\left[K\sqrt{\frac{\ln \max(p,q)}{n}} + K^2\sqrt{\frac{\ln n}{n}} \right]\right) \end{aligned} \tag{A.122}$$

We use Lemma A.9(i) for the first inequality, and Lemma A.10(i), (ii), and Lemma A.12(ii) for the rates. Next consider the second term in (A.121)

$$
\begin{aligned}
\|(\hat{B}-B)\hat{G}B'\|_{l_\infty} &\le pK^{1/2}\|\hat{B}-B\|_{l_\infty}\|\hat{G}\|_{l_2}\|B\|_\infty \\
&\le pK^{1/2}O_p\left(K^{3/2}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right)O_p\left(\frac{r_n}{p}\right)O_p(1) \\
&= O_p\left(r_nK^2\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right) \qquad \text{(A.123)}
\end{aligned}
$$

We use Lemma A.9(i) for the first inequality, and Lemma A.10(ii), Lemma A.12(ii), and Assumption 6: $\max_{1\le j\le p}\max_{1\le k\le K}|b_{j,k}|\le C<\infty$ shows that factor loadings are uniformity bounded away from infinity. Next, consider the third term in (A.121)

$$
\begin{aligned}
\|B\hat{G}(\hat{B}-B)'\|_{l_\infty} &\le pK^{1/2}\|B\|_{l_\infty}\|\hat{G}\|_{l_2}\|\hat{B}-B\|_\infty \\
&= pK^{1/2}O_p(K)O_p\left(\frac{r_n}{p}\right)O_p\left(K\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right) \\
&= O\left(r_nK^{3/2}\left[K\sqrt{\frac{\ln\max(p,q)}{n}}+K^2\sqrt{\frac{\ln n}{n}}\right]\right) \qquad \text{(A.124)}
\end{aligned}
$$

We use Lemma A.9(i) for the first inequality, Lemma A.10(i), Lemma A.10(iv) and Lemma A.12(ii) for the rate. Next, we analyze the fourth term of (A.121)

$$
\|B(\hat{G}-G)B'\|_{l_\infty} \le pK^{1/2}\|B\|_{l_\infty}\|\hat{G}-G\|_{l_2}\|B\|_\infty \qquad \text{(A.125)}
$$

where we use Lemma A.9(i) for the first inequality. Using definition of $\hat{G}$ and G from (A.117) and (A.118) by submultiplicativity of l_2 matrix norm (spectral norm)

$$
\begin{aligned}
\|\hat{G}-G\|_{l_2} &= \|\hat{G}[(\widehat{\Sigma_{f_t}}^{-1}+\hat{B}'\hat{\Theta}_{sym}\hat{B})-[\Sigma_{f_t}]^{-1}+B'\Theta B))]G\|_{l_2} \\
&\le \|\hat{G}\|_{l_2}\|G\|_{l_2}\|\widehat{\Sigma_{f_t}}^{-1}+\hat{B}'\hat{\Theta}_{sym}\hat{B})-\Sigma_{f_t}^{-1}+B'\Theta B)\|_{l_2} \\
&= O_p\left(\frac{r_n}{p}\right)O_p\left(\frac{r_n}{p}\right)O_p\left(pK\bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p,\ln q)}{n}}+K^4\sqrt{\frac{\ln n}{n}}\right]\right) \\
&= O_p\left(\frac{r_n^2}{p}K\bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p,\ln q)}{n}}+K^4\sqrt{\frac{\ln n}{n}}\right]\right) \qquad \text{(A.126)}
\end{aligned}
$$

We use Lemma A.12, and Lemma A.11 for the rates. Then, via Lemma A.10

$$
\begin{aligned}
||B(\hat{G}-G)B'||_{l_\infty} &= pK^{1/2}O_p(K)O_p\left(r_n^2 p^{-1} K \bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right) \\
&= O\left(r_n^2 K^{3/2} \bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right) \qquad \text{(A.127)}
\end{aligned}
$$

Since the last rate is the slowest among all on the right side of (A.121) we have the desired results.

$$
\begin{aligned}
||\hat{L}-L||_{l_\infty} &= O\left(r_n^2 K^{3/2} \bar{s}^{3/2}\left[K^3\sqrt{\frac{\max(\ln p, \ln q)}{n}} + K^4\sqrt{\frac{\ln n}{n}}\right]\right) \\
&= O_p(l_n) = o_p(1)
\end{aligned}
$$

Q.E.D

Theorem 2

Proof of Theorem 2

Although the rate of our estimated error precision matrix differs from that in Caner et al. (2023). the proof for the estimated precision matrix for asset returns is very similar. We start with:

$$\max_{1\le j\le p} ||\hat{\kappa}_j - \kappa_j||_1 = \max_{1\le j\le p} ||\hat{\kappa}'_j - \kappa'_j||_1 \le \max_{1\le j\le p} ||\Theta'_j - \Theta_j||_1 + \max_{1\le j\le p} ||\hat{\Theta}_j \hat{L}\hat{\Theta} - \Theta'_j L\Theta||_1 \tag{A.128}$$

Then the remainder of the proof once can refer to Theorem 2 (i) of Caner et al. (2023)

$$\max_{1\le j\le p} ||\hat{\Theta}'_j \hat{L}\hat{\Theta} - \Theta'_j L\Theta||_1 = O_p(\bar{s}\, l_n) = o_p(1) \tag{A.129}$$

$$\max_{1\le j\le p} ||\hat{\Theta}'_j \hat{L}\hat{\Theta} - \Theta'_j L\Theta||_1 = O_p(\bar{s}\, l_n) = o_p(1) \tag{A.130}$$

(ii)

Since

$$y_t = B(f_t - v_t) + \epsilon_t = B f_t + (\epsilon_t - B v_t)$$

as in Fan et al. (2008) with y_t, u_t begin $p \times 1$ vector of asset returns, and composite errors respectively at time $t = 1, ..., n$. We have:

$$\mu = E[y_t] = E[B(f_t - v_t) + \epsilon_t] = BE[f_t] \tag{A.131}$$

T hen,

$$\hat{\mu} - \mu = B\left[\frac{1}{n}\sum_{i=1}^{n}(f_t - E[f_t])\right] + \frac{1}{n}\sum_{i=1}^{n}\epsilon_t - B\frac{1}{n}\sum_{i=1}^{n} v_t$$

By Assumption 1. consider

$$
\begin{aligned}
||\hat{\mu}-\mu||_{\infty} &\leq \left\|B\left[\frac{1}{n}\sum_{i=1}^{n}(f_t - E[f_t])\right]\right\|_{\infty} + \left\|B\frac{1}{n}\sum_{i=1}^{n}v_t\right\|_{\infty} + \left\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_t\right\|_{\infty} \\
&\leq ||B||_{l_\infty}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}(f_t - E[f_t])\right\|_{\infty} + \left[||B||_{l_\infty}\left\|\frac{1}{n}\sum_{i=1}^{n}v_t\right\|_{\infty}\right] + \left[\left\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_t\right\|_{\infty}\right]\right. \\
&= O(K)O_p\left(\sqrt{\frac{\ln(n)}{n}}\right) + O_p\left(\sqrt{\frac{\ln p}{n}}\right) \\
&= O_p\left[\max\left(K\sqrt{\frac{\ln(n)}{n}}, \sqrt{\frac{\ln p}{n}}\right)\right]
\end{aligned}
$$

by proof of Lemma A.1(i), here we have $||Ax||_{\infty} \leq ||A||_{l_\infty}||x||_{\infty}$ for a generic vector x and matrix A. Then, by Lemma A.10(iii) and Theorem A.1 to get the rate.

Q.E.D

Theorem 3 and 4 can be found in Caner et al. (2023)

Empirical Results: Out-of-sample January 2008-2018

In-sample: January 1994-Dec 2007 ($n_I = 168$), Out-of-sample January 2008 - May 2018 ($n - n_I = 155$)

Table A.1: Out-of-Sample Empirical Results (Jan 2008 – May 2018) Without Transaction Costs($n - n_I = 155$)

	Global Minimum				Markowitz Portfolio			
	Sharpe	Return	SD	Turnover	Sharpe	Return	SD	Turnover
EW	0.234	0.011	0.049	0.050				
NW	0.254	0.010	0.041	0.132	0.244	0.010	0.042	0.172
SFNLS	0.241	0.008	0.032	0.300	**0.245**	0.008	0.032	0.309
NLS	0.228	0.007	0.033	0.311	0.236	0.008	0.033	0.320
POET	0.206	0.007	0.033	0.100	0.202	0.007	0.034	0.115
1 Factor								
rbNW	0.224	0.008	0.034	0.124	0.218	0.007	0.034	0.137
3 Factors								
rbIV	0.208	0.007	0.033	0.279	0.209	0.007	0.033	0.318
rbNW	0.214	0.007	0.034	0.177	0.227	0.008	0.034	0.201
5 Factors								
rbIV	**0.259**	0.009	0.033	0.600	0.017	0.004	0.229	9.276
rbNW	0.227	0.007	0.033	0.199	0.235	0.008	0.032	0.221

Table A.2: Out-of-Sample Empirical Results (Jan 2008 – May 2018) With Transaction Costs ($n - n_I = 155$)

	Global Minimum			**Markowitz Portfolio**		
	Sharpe	Return	SD	Sharpe	Return	SD
EW	0.232	0.011	0.049			
NW	0.252	0.010	0.041	0.243	0.010	0.042
SFNLS	0.242	0.008	0.032	**0.248**	0.008	0.032
NLS	0.234	0.008	0.033	0.242	0.008	0.033
POET	0.209	0.007	0.033	0.204	0.007	0.034
1 Factor						
rbNW	0.227	0.008	0.034	0.222	0.008	0.034
3 Factors						
rbIV	0.208	0.007	0.033	0.207	0.007	0.033
rbNW	0.215	0.007	0.034	0.228	0.008	0.034
5 Factors						
rbIV	**0.256**	0.009	0.034	0.008	0.002	0.211
rbNW	0.228	0.007	0.033	0.222	0.008	0.034

APPENDIX

B

CHAPTER 2

Proof of Theorem 1

$$||w_{gmv} - w^*||_1 = \left\|\frac{\hat{\Gamma}1_p}{1_p'\hat{\Gamma}1_p} - \frac{\Gamma 1_p}{1_p'\Gamma 1_p}\right\|_1 \tag{B.1}$$

Define : $A = \frac{1_p'\Gamma 1_p}{p}$ and $\hat{A} := \frac{1_p'\hat{\Gamma}1_p}{p}$, we rewrite (B.1)

$$||w_{gmv} - w^*||_1 = \left\|\frac{\Gamma 1_p/p}{A} - \frac{\hat{\Gamma}1_p/p}{\hat{A}}\right\|_1 = \frac{||(\Gamma 1_p/p)\hat{A} - (\hat{\Gamma}1_p/p)A||_1}{|A\hat{A}|} \tag{B.2}$$

We start with the denominator:

$$|A\hat{A}| = |A^2 + A(\hat{A} - A)| \geq A^2 - A|\hat{A} - A| \tag{B.3}$$

by (A.131) of Caner et al. (2023), we get

$$|\hat{A} - A| = O_p(\bar{s}\,l_n) = o_p(1) \tag{B.4}$$

by Lemma B.5 of Caner et al. (2023), we have $|A| \leq c$ for some positive constant $c < \infty$. Next, combine (B.4) in (B.3):

$$|\hat{A}A| \geq c^2 - o_p(1) \tag{B.5}$$

by adding and subtracting $(\Gamma 1_p/p)A$ in the numerator of (B.2)

$$\begin{aligned}
||(\Gamma 1_p/p)\hat{A} - (\hat{\Gamma} 1_p/p)A||_1 &= ||(\Gamma 1_p/p)\hat{A} - (\Gamma 1_p/p)A + (\Gamma 1_p/p)A - (\hat{\Gamma} 1_p/p)A||_1 \\
&= \frac{1}{p}||\Gamma 1_p \hat{A} - \Gamma 1_p A + \Gamma 1_p A - \hat{\Gamma} 1_p A||_1 \\
&\leq \frac{1}{p}[||\Gamma 1_p(\hat{A} - A) + (\Gamma - \hat{\Gamma})1_p A||_1] \\
&= \frac{1}{p}[||\Gamma 1_p(\hat{A} - A)||_1 + ||(\Gamma - \hat{\Gamma})1_p A||_1]
\end{aligned} \tag{B.6}$$

Consider the first term in (B.6)

$$\begin{aligned}
||\Gamma 1_p(\hat{A} - A)||_1 \leq |\hat{A} - A|||\Gamma 1_p/p||_1 &\leq |\hat{A} - A|||\Gamma||_{l_1} \\
&= |\hat{A} - A|||\Gamma||_{l_\infty}
\end{aligned} \tag{B.7}$$

This final equality in B.7 is due to the symmetry of Γ. Next, consider the second term in (B.6)

$$\begin{aligned}
\frac{1}{p}||(\Gamma - \hat{\Gamma})1_p A||_1 &= A||(\hat{\Gamma} - \Gamma)1_p/p||_1 \\
&\leq A||(\hat{\Gamma} - \Gamma)||_{l_1}||1_p/p||_1 = A||\hat{\Gamma}' - \Gamma'||_{l_\infty} \\
&= A \max_{1 \leq j \leq p} ||\hat{\Gamma}_j - \Gamma_j||_1
\end{aligned} \tag{B.8}$$

The rate of the second term from Theorem 2 of Caner et al. (2023) is

$$\max_{1 \leq j \leq p} ||\hat{\Gamma}_j - \Gamma_j||_1 = O_p(\bar{s}\, l_n) = o_p(1) \tag{B.9}$$

Combining (B.6) through (B.9),

$$||(\hat{\Gamma} 1_p/p)A - (\Gamma 1_p/p)\hat{A}||_1 = O_p(\bar{s}^2 l_n r_n K^{3/2}) \tag{B.10}$$

Then by (B.5) and(B.10) we get the rate

$$\frac{||(\Gamma 1_p/p)\hat{A} - (\hat{\Gamma} 1_p/p)A||_1}{|A\hat{A}|} = O_p(\bar{s}\, l_n r_n K^{3/2}). \tag{B.11}$$

Finally by (B.2),

$$||w_{gmv} - w^*||_1 = O_p(\bar{s}\, l_n\, r_n K^{3/2}) = o_p(1). \tag{B.12}$$

Q.E.D

Proof of Theorem 2

Following the notation of Caner et al. (2023), define A, F, and D as $A := \frac{1_p' \Gamma 1_p}{p}$, $F := \frac{1_p' \Gamma \mu}{p}$, $D := \frac{\mu' \Gamma \mu}{p}$, with $\Gamma := \Sigma_y^{-1}$. We define the estimators of these terms as $\hat{A} := \frac{1_p' \hat{\Gamma} 1_p}{p}$,$F := \frac{1_p' \hat{\Gamma} \hat{\mu}}{p}$, $D := \frac{\hat{\mu}' \hat{\Gamma} \hat{\mu}}{p}$. Let the scalar $z = AD - F^2$, by (A.123) in Caner et al. (2023), we get $z \leq |AD| = O(K)$.

From Lemma B.5 in Caner et al. (2023), we have (i) $A = O(1)$, (ii)$|D| = O(K)$, (iii)$|F| = O(K^{1/2})$. The optimal solution for the weights in Markowtiz-Mean Variance (MKW) is defined as:

$$w_{mk} := \frac{1}{AD - F^2}\left[D\Sigma_y^{-1} 1_p/p - F\Sigma_y^{-1}\mu/p\right] + \frac{1}{AD - F^2}\rho_1\left[A\Sigma_y^{-1}\mu/p - F\Sigma_y^{-1} 1/p\right] \tag{B.13}$$

Define $g := \frac{D\Sigma_y^{-1} 1_p}{p} - \frac{F\Sigma_y^{-1}\mu}{p}$, and similarly, $\hat{g} := \frac{\hat{D}\hat{\Sigma}_y^{-1} 1_p}{p} - \frac{\hat{F}\hat{\Sigma}_y^{-1}\mu}{p}$ $v := \frac{A\Sigma_y^{-1}\hat{\mu}}{p} - \frac{F\Sigma_y^{-1} 1_p}{p}$, $\hat{v} := \frac{\hat{A}\hat{\Sigma}_y^{-1}\mu}{p} - \frac{\hat{F}\hat{\Sigma}_y^{-1} 1_p}{p}$

The estimator of w_{mk} is

$$\hat{w}_{mk} := \frac{1}{\hat{z}}\left[\hat{D}\hat{\Sigma}_y^{-1} 1_p/p - \hat{F}\hat{\Sigma}_y^{-1}\mu/p\right] + \frac{1}{\hat{z}}\rho_1\left[\hat{F}\hat{\Sigma}_y^{-1} 1_p/p - \hat{F}\hat{\Sigma}_y^{-1}\hat{\mu}/p\right] \tag{B.14}$$

So

$$||w_{mk} - \hat{w}_{mk}||_1 = \left\|\left(\frac{g}{z} + \rho_1\frac{v}{z}\right) - \left(\frac{\hat{g}}{\hat{z}} - \rho_1\frac{\hat{v}}{\hat{z}}\right)\right\|_1.$$

By the triangle inequality in l_1-norm

$$
\begin{aligned}
||w_{mk} - \hat{w}_{mk}||_1 &= \left\| \left(\frac{g}{z} + \rho_1 \frac{v}{z} \right) - \left(\frac{\hat{g}}{\hat{z}} - \rho_1 \frac{\hat{v}}{\hat{z}} \right) \right\|_1 \\
&\leq \left\| \left(\frac{g}{z} - \frac{\hat{g}}{\hat{z}} \right) + \rho_1 \left(\frac{v}{z} - \frac{\hat{v}}{\hat{z}} \right) \right\|_1 \\
&\leq \left\| \frac{g}{z} - \frac{\hat{g}}{\hat{z}} \right\|_1 + \rho_1 \left\| \frac{v}{z} - \frac{\hat{v}}{\hat{z}} \right\|_1 \\
&\leq \left\| \frac{g\hat{z} - \hat{g}z}{z\hat{z}} \right\|_1 + \rho \left\| \frac{v\hat{z} - \hat{v}z}{z\hat{z}} \right\|_1 \\
&= \frac{1}{|z\hat{z}|} \left\{ ||g\hat{z} - \hat{g}z||_1 + \rho_1 ||v\hat{z} - \hat{v}z||_1 \right\}
\end{aligned}
\tag{B.15}
$$

Starting with the denominator in (B.15), by (A.22) and (A.123) of Caner et al. (2023), $z \geq C_1 > 0$, and $|z| = O(K)$

$$
\begin{aligned}
|z\hat{z}| = |z[(\hat{z} - z) + z]| &\geq z^2 - z|(\hat{z} - z)| \\
&\leq z^2 - z|(\hat{z} - z)| \\
&= O(K^2) - O(K)O(K^2 \bar{s} l_n) \\
&= O(K^2) - o_p(1)
\end{aligned}
\tag{B.16}
$$

The rate of $|\hat{z} - z| = O(K^2 \bar{s} l_n)$ from (A.126) in Caner et al. (2023). Consider the first term in (B.15),

$$
\begin{aligned}
||g\hat{z} - \hat{g}z||_1 &= ||g\hat{z} - gz + gz - \hat{g}z||_1 \\
&= ||g(\hat{z} - z) + (g - \hat{g})z||_1 \\
&\leq ||g(\hat{z} - z)||_1 + ||(g - \hat{g})z||_1
\end{aligned}
\tag{B.17}
$$

Analyzing the first term in (B.17),

$$\begin{aligned}
||g(\hat{z}-z)||_1 &= ||g||_1|\hat{z}-z| \\
&\le |\hat{z}-z|||D\Gamma 1_p/p - F\Gamma\mu/p||_1 \\
&\le |\hat{z}-z|\bigg[||D\Gamma 1_p/p||_1 + |F\Gamma\mu/p||_1\bigg] \\
&\le |\hat{z}-z|\bigg[||D\Gamma 1_p/p||_1 + |F\Gamma\mu/p||_1\bigg] \\
&= |\hat{z}-z|\bigg[|D|||\Gamma 1_p/p||_1 + |F||\Gamma\mu/p||_1]\bigg]. \qquad \text{(B.18)}
\end{aligned}$$

Consider each term in (B.18)

$$\begin{aligned}
||\Gamma 1_p/p||_1 &\le |||\Gamma||_{l_1} \\
&= ||\Gamma||_{l_\infty} \qquad \text{(B.19)}
\end{aligned}$$

by p.345 of Horn and Johnson (2013) for the inequality, then by the symmetry of Γ

$$\begin{aligned}
||\Gamma||_{l_\infty} &= ||\Omega - \Omega L\Omega||_{l_\infty} \\
&\le ||\Omega||_{l_\infty} + ||\Omega||_{l_\infty}||L||_{l_\infty}||\Omega||_{l_\infty} \\
&= O(\sqrt{s}) + O(\sqrt{s})O(r_n K^3/2)O(\sqrt{s}) = O(\bar{s}\, r_n K^{3/2}) \qquad \text{(B.20)}
\end{aligned}$$

where the rates are from $(A.96)-(A.97)$ in Caner et al. (2023). Now, the second term in (B.18)

$$\begin{aligned}
\frac{1}{p}||\Gamma\mu||_1 &\le ||\Gamma||_{l_1}||\mu||_\infty \\
&= ||\Gamma||_{l_\infty}||\mu||_\infty \\
&= O(\bar{s}\, r_n K^{3/2})O(K) = O(\bar{s}\, r_n K^{5/2}) \qquad \text{(B.21)}
\end{aligned}$$

by p.345 of Horn and Johnson (2013) for the inequality, then by the symmetry of Γ, relation between l_1 and l_∞ the rate from (B.8) in Caner et al. (2023), $||\mu|| = O(K)$. Then by (B.20),(B.21), and (A.126) from Caner et al. (2023),

$$\begin{aligned}
||g(\hat{z}-z)||_1 &= O(K^2\bar{s}\, l_n)\bigg[O(K)O(\bar{s}\, r_n K^{3/2}) + O(K^{1/2})O(\bar{s}\, r_n K^{5/2})\bigg] \\
&= O(K^2\bar{s}\, l_n)O(K^3\bar{s}\, r_n) = O(K^5\bar{s}^2 l_n) \qquad \text{(B.22)}
\end{aligned}$$

where the rate comes from Lemma B.5 in Caner et al. (2023). Next, consider the second term in

(B.17)

$$
\begin{aligned}
||(g-\hat{g})z||_1 &= |z|||g-\hat{g}||_1 \\
&= |z|||(D\Sigma_y^{-1}1_p - F\Sigma_y^{-1}\mu) - (\hat{D}\hat{\Sigma}_y^{-1}1_p - \hat{F}\hat{\Sigma}_y^{-1}\hat{\mu})||_1/p \\
&= |z|||(D\Gamma 1_p - \hat{D}\hat{\Gamma}1_p)||_1/p \\
&+ |z|||F\Gamma\mu - \hat{F}\hat{\Gamma}\hat{\mu}||_1/p
\end{aligned} \tag{B.23}
$$

Consider the first term in (B.23),

$$
\frac{1}{p}||D\Gamma 1_p - \hat{D}\hat{\Gamma}1_p||_1 \leq \frac{1}{p}||D(\Gamma-\hat{\Gamma})1_p||_1 \tag{B.24}
$$

$$
+||(D-\hat{D})(\hat{\Gamma}-\Gamma)1_p)||_1 \tag{B.25}
$$

$$
+||(D-\hat{D})\Gamma 1_p||_1 \tag{B.26}
$$

Now, consider the right-hand side term in (B.24),

$$
\begin{aligned}
\frac{1}{p}||D(\Gamma-\hat{\Gamma})1_p||_1 &= \frac{1}{p}|D|||(\Gamma-\hat{\Gamma})1_p||_1 \\
&\leq |D|||\Gamma-\hat{\Gamma}||_{l_1} = |D|\max_{1\leq j\leq p}||\Gamma_j - \hat{\Gamma}_j||_1 \\
&= O(K)O(\bar{s}l_n) = O(K\bar{s}l_n)
\end{aligned} \tag{B.27}
$$

where we use p.345 of Horn and Johnson (2013) for the first inequality and the rate from Lemma B.5 and Theorem 2(i) of Caner et al. (2023). Next, analyzing the second term of (B.25)

$$
\begin{aligned}
\frac{1}{p}||(D-\hat{D})(\hat{\Gamma}-\Gamma)1_p||_1 &= |(D-\hat{D})|\ ||(\hat{\Gamma}-\Gamma)1_p||_1 \\
&\leq |D-\hat{D}|\ ||\max_{1\leq j\leq p}|\Gamma_j - \hat{\Gamma}_j||_1 \\
&= O(K^2\bar{s}l_n)O(\bar{s}l_n) = O(K^2\bar{s}^2 l_n^2)
\end{aligned} \tag{B.28}
$$

The rate of $|D-D| = O_p(K^2\bar{s}l_n)$ and the second rate from Lemma B.4 of Caner et al. (2023). Next, consider the last term of (B.26)

$$
\begin{aligned}
||(D-\hat{D})\Gamma 1_p||_1 &= |D-\hat{D}|\ ||\Gamma 1_p||_1 \\
&\leq |D-\hat{D}|\ ||\Gamma||_{l_1} \\
&= O(K^2\bar{s}l_n)O(\bar{s}r_n K^{3/2})
\end{aligned} \tag{B.29}
$$

Combine (B.27), (B.28), and (B.29)

$$||D\Gamma 1_p - \hat{D}\hat{\Gamma} 1_p||/p_1 = O(K^2 \bar{s}^2 l_n^2) \tag{B.30}$$

Next, we consider the second term in (B.23), by simply adding and subtracting $\hat{F} = (\hat{F} - F) + F$. Similarly, $\hat{\Gamma} = (\hat{\Gamma} - \Gamma) + \Gamma$, and $\hat{\mu} = (\hat{\mu} - \mu) + \mu$.

$$\begin{aligned}
||F\Gamma\mu - \hat{F}\hat{\Gamma}\hat{\mu}||_1 &\leq ||F\Gamma\mu - [(\hat{F} - F) + F][(\hat{\Gamma} - \Gamma) + \Gamma][(\hat{\mu} - \mu) + \mu]||_1 \\
&\leq ||(\hat{F} - F)[(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu) + \Gamma(\hat{\mu} - \mu) + (\hat{\Gamma} - \Gamma)\mu + \Gamma\mu]||_1 \qquad \text{(B.31)} \\
&+ ||F[(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu) + \Gamma(\hat{\mu} - \mu) + (\hat{\Gamma} - \Gamma)\mu]||_1 \qquad \text{(B.32)}
\end{aligned}$$

First, we consider some terms that will be used often.

$$\begin{aligned}
||(\hat{\Gamma} - \Gamma)(\hat{\mu} - \mu)/p||_1 &\leq ||(\hat{\Gamma} - \Gamma)||_{l_1} ||\hat{\mu} - \mu||_\infty \\
&= \max_{1 \leq j \leq p} ||\Gamma_j - \hat{\Gamma}_j||_1 ||\hat{\mu} - \mu||_\infty \\
&= O(\bar{s} l_n) O\left(\max(K\sqrt{\frac{\ln(n)}{n}}, \sqrt{\frac{\ln p}{n}} \right) \qquad \text{(B.33)}
\end{aligned}$$

using p.345 of Horn and Johnson (2013) for the first inequality, $||\hat{\mu} - \mu||_1 \leq p||\hat{\mu} - \mu||_\infty$, and the rates from Theorem 1(i) and Theorem 2 (ii) of Caner 2023. Next,

$$\begin{aligned}
||\Gamma(\hat{\mu} - \mu)/p||_1 &\leq ||\Gamma||_{l_1} ||\mu - \mu||_\infty \\
&= O(\bar{s} r_n K^{3/2}) O\left(\max(K\sqrt{\frac{\ln(n)}{n}}, \sqrt{\frac{\ln p}{n}} \right) \qquad \text{(B.34)}
\end{aligned}$$

The rate is from (B.7) and Theorem 2 of Caner 2023. Next,

$$\begin{aligned}
||(\hat{\Gamma} - \Gamma)\mu/p||_1 &\leq ||\hat{\Gamma} - \Gamma||_{l_1} ||\mu||_\infty \\
&= \max_{1 \leq j \leq p} ||\Gamma_j - \hat{\Gamma}_j||_1 ||\mu||_\infty \\
&= O(\bar{s} l_n) O(K) = O(K \bar{s} l_n) \qquad \text{(B.35)}
\end{aligned}$$

using p.345 of Horn and Johnson (2013) for the first inequality, and $||\mu||_1 \leq p|\mu||_\infty$. Next, similarly,

$$\begin{aligned}
||\Gamma\mu/p||_1 &\leq \frac{1}{p}||\Gamma||_{l_1}||\mu||_1 \\
&= ||\Gamma||_{l_\infty}||\mu||_\infty \\
&= O(\bar{s}r_n K^{3/2})O(K) = O(\bar{s}r_n K^{5/2}) \qquad \text{(B.36)}
\end{aligned}$$

Combine (B.33)-(B.36) and $|\hat{F}-F| = O(K\bar{s}l_n)$, under Assumption 1-7(i) of Caner et al. (2023), we have

$$||F\Gamma\mu - \hat{F}\hat{\Gamma}\hat{\mu}||_1 = O(K^{1/2})O(K^2\bar{s}^2 l_n^2) \qquad \text{(B.37)}$$

Combine all above, we get the rate of the first term in (B.15),

$$||g\hat{z} - \hat{g}z||_1 = O(K)O(K^{5/2}\bar{s}^2 l_n^2) = O(K^{7/2}\bar{s}^2 l_n^2) \qquad \text{(B.38)}$$

Next, we analyze the second term in (B.15),

$$\begin{aligned}
|v\hat{z} - \hat{v}z|_1 &= |v\hat{z} - vz + vz - \hat{v}z|_1 \\
&\leq |v(\hat{z}-z)|_1 + |(v-\hat{v})z|_1 \qquad \text{(B.39)}
\end{aligned}$$

Considering each term in (B.39) and using the triangle inequality, we get

$$\begin{aligned}
|v(\hat{z}-z)|_1 &= |\hat{z}-z||v|_1 \\
&\leq |\hat{z}-z|\Big[|A\Gamma\mu/p - F\Gamma 1_p/p|_1\Big] \\
&\leq |\hat{z}-z|\Big[|A\Gamma\mu/p|_1 + |F\Gamma 1_p/p|_1\Big]. \qquad \text{(B.40)}
\end{aligned}$$

The first term in (B.40) has the rate

$$\begin{aligned}
||A\Gamma\mu/p||_1 &\leq |A|||\Gamma\mu/p||_1 \\
&= O(1)O(\bar{s}r_n K^{5/2}) \qquad \text{(B.41)}
\end{aligned}$$

Next, the second term in (B.40),

$$\begin{aligned}
||F\Gamma 1_p/p||_1 &\leq |F|||\Gamma 1_p/p||_1 \leq |F|||\Gamma||_{l_1} \\
&= O(K^{1/2})O(\bar{s}r_n K^{3/2}) \qquad \text{(B.42)}
\end{aligned}$$

Combining (B.41) and (B.42),

$$|v(\hat{z}-z)|_1 = O(K^2 \bar{s}\, l_n) O(\bar{s}\, r_n K^3) \tag{B.43}$$

The rate $|\hat{z}-z| = O(K^2 \bar{s}\, l_n)$ from (A.126) of Caner et al. (2023).

The second term in (B.39),

$$\begin{aligned}
||(v-\hat{v})z||_1 &= |z|\left[\left(||A\Gamma\mu/p + F\Gamma 1_p/p\right) - \left(\hat{A}\hat{\Gamma}\hat{\mu}/p + \hat{F}\hat{\Gamma}1_p/p\right)||_1\right] \\
&= |z|\frac{1}{p}\left[||(A\Gamma\mu - \hat{A}\hat{F}\hat{\mu}) + (F\Gamma 1_p - \hat{F}\hat{\Gamma}1_p)||_1\right] \\
&\le |z|\frac{1}{p}\left[||(A\Gamma\mu - \hat{A}\hat{F}\hat{\mu})||_1 + ||(F\Gamma 1_p - \hat{F}\hat{\Gamma}1_p||_1\right] \qquad \text{(B.44)}
\end{aligned}$$

Considering each term in (B.44), we start with simply adding and subtracting $\hat{A} = (\hat{A}-A)+A$, $\hat{\Gamma} = (\hat{\Gamma}-\Gamma)+\Gamma$, $\hat{\mu} = (\hat{\mu}-\mu)+\mu$

$$\begin{aligned}
||A\Gamma\mu - \hat{A}\hat{\Gamma}\hat{\mu}||_1/p &= ||A\Gamma\mu - [(\hat{A}-A)+A][(\hat{\Gamma}-\Gamma)+\Gamma][(\hat{\mu}-\mu)+\mu]||_1/p \\
&\le ||(\hat{A}-A)[(\hat{\Gamma}-\Gamma)(\hat{\mu}-\mu)+\Gamma(\hat{\mu}-\mu)+(\hat{\Gamma}-\Gamma)\mu+\Gamma\mu]||_1/p \qquad \text{(B.45)} \\
&+ ||A[(\hat{\Gamma}-\Gamma)(\hat{\mu}-\mu)+\Gamma(\hat{\mu}-\mu)+(\hat{\Gamma}-\Gamma)\mu]||_1/p \qquad \text{(B.46)}
\end{aligned}$$

Analyzing (B.45),

$$\begin{aligned}
&||(\hat{A}-A)[(\hat{\Gamma}-\Gamma)(\hat{\mu}-\mu)+\Gamma(\hat{\mu}-\mu)+(\hat{\Gamma}-\Gamma)\mu+\Gamma\mu]||_1 \\
&\le |\hat{A}-A|\frac{1}{p}\left\{||(\hat{\Gamma}-\Gamma)(\hat{\mu}-\mu)||_1 + ||\Gamma(\hat{\mu}-\mu)||_1 + ||(\hat{\Gamma}-\Gamma)\mu||_1 + ||\Gamma\mu||_1\right\} \qquad \text{(B.47)}
\end{aligned}$$

The rate of $|\hat{A}-A| = O(\bar{s}\, l_n)$ from (A.131) of Caner et al. (2023). Similarly,

$$\begin{aligned}
&|A[(\hat{\Gamma}-\Gamma)(\hat{\mu}-\mu)+\Gamma(\hat{\mu}-\mu)+(\hat{\Gamma}-\Gamma)\mu]|_1 \le \\
&\le |A|\left([||(\hat{\Gamma}-\Gamma)(\hat{\mu}-\mu|)+\Gamma(\hat{\mu}-\mu)+(\hat{\Gamma}-\Gamma)\mu]||_1\right) \\
&\le |A|\left(||(\hat{\Gamma}-\Gamma)(\hat{\mu}-\mu||)_1 + ||\Gamma(\hat{\mu}-\mu)||_1 + ||(\hat{\Gamma}-\Gamma)\mu||_1\right) \qquad \text{(B.48)}
\end{aligned}$$

where we use $A = O(1)$ and the definition of l_n. Combining (B.47), and (B.48),

$$||A\Gamma\mu - \hat{A}\hat{\Gamma}\hat{\mu}||_1 = O(K^2 \bar{s}^2 l_n^2) \tag{B.49}$$

Next, considering the second term of (B.44),

$$||F\Gamma 1_p - \hat{F}\hat{\Gamma}1_p||_1/p = ||F\Gamma 1_p - [(\hat{F}-F)+F][(\hat{\Gamma}-\Gamma)+\Gamma]1_p||_1/p$$
$$\leq ||(\hat{F}-F)(\hat{\Gamma}-\Gamma)1_p||_1/p \quad \text{(B.50)}$$
$$+||(\hat{F}-F)\Gamma 1_p||_1/p \quad \text{(B.51)}$$
$$+||F(\hat{\Gamma}-\Gamma)||_1/p \quad \text{(B.52)}$$

Note that using (A.112): $|\hat{F}-F| = O(K\bar{s}l_n)$, and Lemma B.5 $|F| = O(K^{1/2})$ of Caner et al. (2023), (B.50) is

$$||(\hat{F}-F)(\hat{\Gamma}-\Gamma)1_p||_1 = |\hat{F}-F|\ ||(\hat{\Gamma}-\Gamma)1_p||_1/p. \quad \text{(B.53)}$$

The second term (B.51) is

$$||(\hat{F}-F)\Gamma 1_p||_1/p = |\hat{F}-F|\ ||\Gamma 1_p||_1/p \quad \text{(B.54)}$$

The third term (B.52) is

$$||F(\hat{\Gamma}-\Gamma)||_1/p = |F|\ ||\hat{\Gamma}-\Gamma||_1/p. \quad \text{(B.55)}$$

Combining (B.53), (B.54), (B.55), and the definition of l_n, the final rate comes from the slower rate of the first term. Then,

$$||F\Gamma 1_p - \hat{F}\hat{\Gamma}1_p||_1/p = O(K\bar{s}l_n)O(\bar{s}l_n) \quad \text{(B.56)}$$

by (B.44), (B.49), and (B.57), we get

$$|(v-\hat{v})z|_1 = O(K)O(K^2\bar{s}^2l_n^2) = O(K^3\bar{s}^2l_n^2) \quad \text{(B.57)}$$

where the rate of $|z| = O(K)$ from A.123 of Caner et al. (2023) . Finally, combining (B.15) with (B.22) and (B.57),

$$||\hat{w}_{mk} - \hat{w}_{mk}||_1 = \left|\frac{1}{|z\hat{z}|}\right\{||y\hat{z}-\hat{y}z||_1 + \rho_1||v\hat{z}-\hat{v}z||_1\Big\}$$
$$= \frac{O(K^{7/2}\bar{s}^2l_n^2)}{O(K^2)} = O(K^{3/2}\bar{s}^2l_n^2) \quad \text{(B.58)}$$

Comparing the rate between (B.12) and (B.58), the rate of the weighted estimation of the Markowitz Mean Variance portfolio is faster than for the Global Minimum Variance portfolio.