Submitted to *Econometrica*

GENERAL LINEAR CLASS OF HIGH DIMENSIONAL ESTIMATORS FOR PRECISION MATRIX ESTIMATION

MEHMET CANER
Department of Economics, North Carolina State University

AGOSTINO CAPPONI
Department of Industrial Engineering and Operations Research, Columbia University

MIHAILO STOJNIC
Department of Industrial Engineering and Operations Research, Columbia University

HOI DINH
Department of Economics, North Carolina State University

KEYWORDS: Interpolation, Double Descent, latent space, precision matrix estimation, factor models.

1. INTRODUCTION

In this paper, we provide a new estimator for precision matrices of large-possibly diverging number of observed variables. This can be used in high dimensional confidence interval estimators for structural parameters in high dimensional regressions, network analysis, and financial metrics such as Sharpe Ratio.

We start by providing a formula for population precision matrix that can tie each variable to the other variables in a system. Then each variable is modeled as linear function of common but hidden factors. In this sense, hidden factors act as confounders. This approach

Mehmet Caner: mcaner@ncsu.edu
Agostino Capponi:
Mihailo Stojnic:
Hoi Dinh:

is entirely new. It merges matrix algebra with hidden factors to provide this precision matrix formula.

Then we provide a new estimator based on this new precision matrix formula which we develop. We also show that both ridge-less estimation and a new adaptive principal components based regression belong to same general class of precision matrix estimators based on linear-hidden factor models. Our estimator can be used in high dimensional context and does not require exact sparsity assumption on outcomes as in Callot et al. (2021). So we preserve the links between different variables, and do not assign zero linkages as in nodewise regression. Note that nodewise regression based estimators of Callot et al. (2021) or Caner et al. (2023) crucially depends 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. Our method can be seen as a complementary approach to nodewise regression of Meinshausen and Bühlmann (2006). Nodewise regression uses an l_1 norm based penalization to form the rows of the precision matrix. Doing so, the underlying assumption is the imposition of sparsity in the precision matrix which may be interpreted as a strong assumption. In our paper, we will not follow that approach instead, each of the outcome variables, will be tied to a lower dimensional latent factors. We will allow number of factors to diverge with the sample size.

One key issue is that 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 as we prove in Appendix B. Each economic or financial variable can be related to latent factors and this can be thought as confounding variables as in causality related literature such as in Kaltenpoth and Vreeke (2023), Angrist and Pischke (2009), Imbens and Rubin (2015). 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. Variable such as culture, sentiment, market can be extremely valuable in certain macro, financial data sets. These hidden factors in a linear model can be used to ex-

plain international stock returns such as in Singapore, Malaysia, USA as seen in Brooks and Del Negro (2004). Another key aspect of hidden factor model usage is, this is a nice tool to provide dimension reduction in high dimensional econometrics. This is due to the number of outcomes are being more numerous than the underlying hidden factors. In our paper, we let both dimensions of outcomes as well as number of hidden factors to grow. The role and significance of latent factor models and the estimation of their key components, factors and loadings, have been largely extended and further linked to asset pricing in Lettau and Pelger (2020a), causal inference and estimation of counterfactuals or treatment effects in Athey et al. (2019, 2021), Yan and Wainwright (2024), Kallus et al. (2018), recommenders systems in Agarwal et al. (2023), macroeconomic forecasting in Stock and Watson (2002a,b, 2016), matrix completion and missing panel data in Xiong and Pelger (2023), Cahan et al. (2023), Bai and Ng (2021), Duan et al. (2024), Bryzgalova et al. (2023). Chen et al. (2021) discusses how latent factors can model important features of network data such as reciprocity, degree heterogeneity, and homophily. They apply these methods to the estimation of gravity equations of trade between countries, showing how unobserved latent factors significantly influence observed trade volumes, even after controlling for traditional gravity variables like distance. Excellent theoretical and practical algorithmic results obtained in all these works usually serve as a strong indication that there is indeed a very strong rationale for utilization of factor models. Somewhat paradoxically, while we will rely on similar factor modeling concepts and underlying intuitions that they bring along, we will be able to design excellent precision matrix estimators without actually focusing on directly estimating factors themselves.

Our precision matrix estimate can be used to form debiased estimators for the parameters in a high dimensional regression. 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 (2023). Our method can be used as an alternative inverse estimate of the second or-

der partial derivative of convex loss functions. In network analysis, also our method can be used to have neighborhood selection, and as mentioned above can complement approach of Meinshausen and Bühlmann (2006). In financial econometrics, a penalized version is suggested by Lettau and Pelger (2020a), and financial applications are in Lettau and Pelger (2020b).

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 six 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, 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. We show in our paper that to have consistency of our precision matrix estimator, we need signal-to-noise ratio to dominate the number of factors. The fifth difference is that, Fan et al. (2013) rate of convergence of estimators for the precision matrix of asset returns uses estimation of hidden factors, where we do not use. There are no identification assumptions in our paper such as orthogonality of factor loadings. Also, our paper does not use any identification assumptions on the latent factors. The sixth and a key difference is that our approach also can show consistency when our number of variables is constant but the sample size is growing.

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 latent factor models to provide a new precision matrix formula. Then a novel idea of estimation is to use a linear linkage between outcome variables to obtain precision matrix estimator based on the matrix algebraic new formula. 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), Bunea et al. (2022) do not simply extend to this case. So precision matrix estimation is different than the risk analysis of a single outcome-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), 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. in GLS. In our empirical exercise in stock market in U.S., we observe double ascent phenomenon of Sharpe Ratio with this estimator which is an outcome of double descent in the precision matrix estimation errors. This is a new result and shows that double descent phenomenon is more extensive than linear regression of a single outcome variable. Since, ridgeless regression does not use any tuning parameter, this is a nice finding where we can form large number of assets and still perform well in terms of Sharpe Ratio with a simple closed form solution estimator. To be specific, 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.

Principal component (PC) centered approaches have been shown to enjoy favorable analytical properties when used for handling factor models Fan et al. (2011),Fan et al. (2013), Bai and Ng (2002, 2006), particularly in scenarios where the number of factors is predicated to be limited. Consequently, they have been traditionally viewed as the main algorithmic cornerstone and seemingly an unavoidable tool when it comes to any form of hidden factors associated estimation, prediction, or inference. On the other hand, the classical least squares techniques reemerged as an alternative over the the last few years, bringing into the spotlight as a very prominent algorithmic feature the so-called double-descent phenomenon (such a phenomenon is also tightly connected to generalization properties of neural networks and was empirically observed in Spigler et al. (2019), Belkin et al. (2019) and theoretically substantiated in e.g., Muthukumar et al. (2020), Belkin et al. (2020), Hastie et al. (2022)). As will be clear through the presentation, both of these classical algorithmic concepts – PC and least-squares – play a significant role in the design of the entire precision matrix estimation procedure that is the main topic of the present paper.

We consider notation now. Define C as a generic positive constant. Note that transpose of any matrix $A : d \times l$ is represented by A', and transpose of a column vector, a is represented by a' unless it is stated otherwise for notational convenience. Let $\|A\|_2$ represent the spectral norm of matrix A, and $\|a\|_2$ is the Euclidean norm of a vector a. Also we denote $\|A\|_F$ as the Frobenius norm of A. Let $\|A\|_{l_\infty} := \max_{1\leq j\leq d} \|A'_{-j}\|_1$ be the maximum row sum norm, with $A'_{-j} : 1 \times l$ as the j th row among d rows in matrix A. Denote the k th largest singular value of A matrix as $\sigma_k(A)$ with $k = 1$ being the largest singular value. Let A^+ represent the Moore-Penrose inverse of matrix A. Let $\lambda_k(M)$ represent the k th largest eigenvalue of $M : d \times d$, a square matrix, except $k = 1, k = d$. However, for ease of understanding, the largest eigenvalue of M is denoted as $Eigmax(M)$ and the minimum eigenvalue of M is $Eigmin(M)$. $tr(M)$ represents the trace of matrix M. Let $a \asymp b$ show the same order for a, b random variables.

2. POSSIBLE EXAMPLES

Here in this section we briefly introduce our precision matrix estimator's usage in three different cases of interest in economics and finance. Before proceeding, this is also a contribution to statistics, since we form a new precision matrix estimator of outcomes which does not depend on sparsity of the precision matrix explicitly, and estimation errors can be tied to signal-to-noise ratio explicitly.

In financial econometrics and finance, one of the key issues is large portfolio formation, and how to do this. High dimensional portfolio formation when $p > n$ is done using different precision matrix estimator of asset returns. Literature used factor model based covariance matrix of asset returns and then inverted it like in Fan et al. (2013), or shrinkage based estimation of covariance matrix of asset returns and inverting this by Ledoit and Wolf (2017), and nodewise based regression as in Callot et al. (2021) or nodewise regression combined with observed factors as in Caner et al. (2023). One of the key metrics in practice in finance is Sharpe Ratio since it can measure return-risk ratio for a given portfolio. Our estimator can be used to estimate that ratio consistently even with $p > n$, and can perform well.

The second possibility is by a simple modification of our method we can obtain estimate of inverse of Gram matrix, and replace relaxed inverse that is used to get a high dimensional confidence interval for linear regression parameters as in van de Geer et al. (2014). We believe that the new estimator can be used in causal high dimensional inference. Techniques such as relaxed inverse estimate depends crucially on the sparsity of the inverse of Gram matrix, which we will not impose. Our idea can also form and change the debiased high dimensional estimators used in van de Geer et al. (2014) and Caner and Kock (2018). We discuss a formula and how this can happen in Section 6.3.

The third possibility is the network analysis, and noting that nodewise regression can be tied to network analysis as shown in Meinshausen and Bühlmann (2006), our estimator has the same network structure but without imposing sparsity on the links between two nodes. We discuss briefly our method in Section 6.3.

3. 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 of Y into two equations immediately below in (2)(3).

$$Y = FA + U. \tag{1}$$

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 (1) j th outcome can be written as

$$y_{jt} = f_t' a_j + u_{jt}, \tag{2}$$

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

$$Y_{-jt} = A_{-j} f_t + U_{-jt}. \tag{3}$$

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 (2)(3) in our proofs rather than (1), and in this way our proofs are clear and tractable. Common hidden factors affect j th outcome in (2) as well as all remaining outcome variables in (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 (1) can be written as

$$y_j = Fa_j + u_j, \tag{4}$$

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

$$Y_{-j} = FA_{-j}' + U_{-j}, \tag{5}$$

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). We also benefit from the representation in (4)(5) in our proofs.

Define the variance of j th error as $var(u_{jt}) = \sigma_j^2 \geq c > 0$,where c is a positive constant at time t, for $j = 1 \cdots, p$. Define the variance matrix of errors except the j th one as $var(U_{-jt}) = \Sigma_{U,-j} : p-1 \times p-1$, for $j = 1, \cdots, p$ at time t. So if $j = 3$ for example σ_3^2 is the third asset error variance and $\Sigma_{U,-3}$ represents the covariance matrix of errors for all assets except the third one at time t. Define the covariance matrix of factors as $\Sigma_f := cov(f_t) : K \times K$ at time t. Define $y_t = (y_{jt}, Y_{-jt}')'$ is t th row of Y, and $\mu_j := Ey_{jt} : 1 \times 1$, $\mu_{-j} := EY_{-jt} : p-1 \times 1$, and $\mu := Ey_t : p \times 1$ at time t. Let the covariance matrix of outcomes be $\Sigma := E(y_t - \mu)(y_t - \mu)'$. at time t. Assumption 1 below will show iid nature of our data across $t = 1, \cdots, n$. Define the precision matrix as $\Theta := \Sigma^{-1}$.

3.1. *Assumptions*

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 (i)(ii)(iii) shows that rows of Y, U, F are iiid. However, when there is independence of errors which implies $Eu_{jt}U_{-jt} = 0$. But this means $Eu_t u_t' = I_p$, with $u_t = (u_{jt}, U_{-jt}')'$. This 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, with our extensive proofs we do not pursue this in our paper. A future extension of our work can analyze a relaxed assumption about this specific type of sparsity.

ASSUMPTION 2: *(i).* $\min_{1\leq j\leq p} Eigmin(\Sigma_{U,-j}) \geq c > 0$, *where* $Eigmin(\Sigma_{U,-j})$ *represents the minimum eigenvalue of the matrix* $\Sigma_{U,-j}$. *Also* $\max_{1\leq j\leq p} \sigma_j^2 = \sigma^2 \leq C < \infty$, *with* C *a positive constant, and also* $\min_{1\leq j\leq p} \sigma_j^2 \geq c > 0$, *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 *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), 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* $j = 1, \cdots, p$ $Eexp(t_1 u_{jt}) \leq exp(t_1^2 \sigma_j^2 \gamma_e^2/2)$ *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* $\max_{1 \leq k \leq K} |E f_{tk}| \leq C < \infty$*, 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 needed 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

$$\bar{\xi} = \min_{1 \leq j \leq p} \xi_j, \tag{6}$$

where as in Bunea et al. (2022)

$$\xi_j := \frac{\lambda_K(A_{-j} \Sigma_f A_{-j}')}{\|\Sigma_{U,-j}\|_2}, \tag{7}$$

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

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

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

$$\bar{\xi} \geq \frac{c d_{1n}^2}{r_n}. \tag{8}$$

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}, 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.

$$r_{w1} := \max\left(\sqrt{\frac{logn+\bar{r}}{n\bar{\eta}}}, \sqrt{\frac{K}{\bar{\xi}\delta_n}}, \sqrt{\frac{K}{\bar{\xi}}\frac{\bar{\Psi}}{\bar{\eta}}\frac{1}{\delta_n}}, \sqrt{\frac{K}{\bar{\xi}}\frac{1}{\bar{\eta}}}\right). \tag{9}$$

$$r_{w2} := \max\left(r_{w1}, \sqrt{\frac{K}{\bar{\xi}}}\right). \tag{10}$$

Note the only difference between r_{w1}, r_{w2} is the second term in parentheses in (9)(10), 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).* $\frac{d_{1n}}{\sqrt{p}}\frac{\sqrt{logn}}{\sqrt{Klogp}} \to 0$.

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

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

Now we provide a bit more detailed information about Assumption 5. Assumption 5(i) is a technical assumption that facilitates the proof of Lemma A.1 in Appendix for the finite sample results, and note that this is a strengthened version of Assumption 3 when we have small p, K. 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.

4. 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:

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

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 (2)(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 $\bar{A}_{-j} := A_{-j}\Sigma_f^{1/2}$, and $\bar{a}_j := \Sigma_f^{1/2}a_j$, 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 Callot et al. (2021), Caner et al. (2023), Caner and Kock (2018), 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) propose linear regression coefficient estimation with hidden factor models, and analyzed the risk of an asset in a hidden factor model. But precision matrix estimation or formulae is not discussed in Bing et al. (2021), Bunea et al. (2022). The difficulty is that in 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),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

$$\tilde{\alpha}_j := \hat{B}(Y_{-j}\hat{B})^+ y_j = \hat{B}(\hat{B}' Y_{-j}' Y_{-j} \hat{B})^+ (\hat{B}' Y_{-j}' y_j), \tag{11}$$

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. We let $\hat{B}$ be a deterministic matrix or a matrix that depends on Y_{-j}. Our estimator is based on analysis of precision matrix in Lemma 1 and the equality in (11) 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 based on (B.10)[1]

$$\tilde{\tau}_j^2 := \frac{y_j'(y_j - Y_{-j}\tilde{\alpha}_j)}{n}. \tag{12}$$

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

$$\tilde{\Theta}_{j,j} = 1/\tilde{\tau}_j^2, \tag{13}$$

and the j th row of the precision matrix estimate without j th element (off-diagonal terms in the row) is, via (B.12)

$$\tilde{\Theta}_{j,-j} = -\tilde{\alpha}_j'/\tilde{\tau}_j^2. \tag{14}$$

The j th row of the estimate, to give an example, at $j = 1$ will be $\tilde{\Theta}_1' := (\frac{1}{\tilde{\tau}_1^2}, \frac{-\tilde{\alpha_1}'}{\tilde{\tau}_1^2})$.

5. UPPER BOUND ON l_2 NORM ERROR

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

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

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 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

$$\hat{r}_j := rank(Y_{-j}P_{\hat{B}}). \tag{15}$$

[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.

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 Theorem 1. Define

$$\hat{\eta}_j := \sigma^2_{\hat{r}_j}(Y_{-j}P_{\hat{B}})/n, \tag{16}$$

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 Theorem 1 below. Next, we define

$$\hat{\Psi}_j := \sigma^2_1(Y_{-j}M_{\hat{B}})/n, \tag{17}$$

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, $Klogn \le cn$, with $c > 0$ a positive constant, with probability at least $1 - 2/n$*

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

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

$$\max_{1\le j\le p}\|\tilde{\alpha}_j - \alpha^*_j\|_2^2 = O_p\left(\max(\frac{K}{\bar{\xi}}, \frac{K\bar{\Psi}}{\bar{\xi}\bar{\eta}\delta_n}, \frac{K}{\bar{\xi}\bar{\eta}}, \frac{logn + \bar{r}}{n\bar{\eta}})\right) = O_p(r_{w2}),$$

where r_{w2} is defined in (9)-(10), by using δ_n is nondecreasing in n.

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

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

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

4. 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.

5. 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).

6. 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.

5.1. *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 Theorem 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 $\sigma^2_1(Y_{-j})/n$ (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,

$$\Delta_{-j} := c_\Delta[\|\Sigma_{U,-j}\|_2 + tr(\Sigma_{U,-j})/n], \tag{18}$$

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. This type of choice is done to minimize the risk of adaptive PCR in Bing et al. (2021). 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. Set $k_j = \hat{s}_j$. By definition of $\hat{s}_j$ above

$$\hat{\lambda}_{\hat{s}_{j+1}} < C_0\Delta_{-j} \leq \hat{\lambda}_{\hat{s}_j}. \tag{19}$$

We show in the proof of Corollary 1 in Appendix

$$P(\max_{1\leq j\leq p} \hat{s}_j \leq K) \to 1, \tag{20}$$

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: *Under Assumptions 1-4 and $logp < c'n$, with $0 < c' < 1$ a positive constant, $Klogn \leq cn$, with c a positive constant, with probability at least $1 - 2/n - exp((c' - 1)n)$, and*

(i).

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

(ii).

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

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 $(logn + K)/nr_n \to 0$ and this is easy to satisfy given $Klogn < cn$ 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 (8) we may need $Kr_n/d_{1n}^2 \to 0$.

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 $logp < c'n$, with $0 < c' < 1$.

4. If we fix $p/n = \gamma \in [0, \infty)$ and then we increase n, p together, and we keep γ fixed to check consistency. In $(logn + K)/(nr_n)$ condition, r_n can grow with n, p both increasing and have a positive effect on consistency. The price here is $logn$ rate, but that is not large compared to n rate in the denominator. In the second condition $K/\bar{\xi}$, by using the example (8), it is tied to Kr_n/d_{1n}^2 rate as discussed in Remark 2 here. Increasing n, p together can increase r_n, but also d_{1n}. It is clear that r_n has a tradeoff to get consistency. The noise r_n can reduce variance but increase bias. Note that d_{1n} shows the strength of factor loadings A_{-j}. With $Kr_n/d_{1n}^2 \to 0$ we have consistency.

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

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

with $\lfloor x \rfloor$ represents the floor function, which is the greatest integer strictly less than real number x, $\kappa > 1$ is a positive constant number.

5.2. *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, $K log n < cn$, with c a positive constant, with probability at least $1 - 2/n - 1/n^{c_1/2}$, and

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

(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 asymptotically. GLS will be consistent when $p/n \to 0$.

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: $p - 1 > n$)*

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

(i).

$$\max_{1 \le j \le p} \|\hat{\alpha}_{GLS,j} - \alpha_j^*\|_2^2 \le 2\gamma_e^2 \sigma^2 \left[\frac{2logn + n}{Cnr_n}\right] + \frac{3CK}{\bar{\xi}}\left[1 + \frac{2}{\delta_n}\right].$$

with probability at least $1 - 2/n - exp((-c_2/2)n)$, and

(ii).

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

Remarks.

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

$$\frac{logn + n}{nr_n} \to 0,$$

since r_n may grow larger. Then $K/\bar{\xi}$ can be smaller if we use the example (8), and $Kr_n/d_{1n}^2 \to 0$. So signal-to-noise ratio should grow and also dominate the number of factors as well to have consistency.

2. Note that we may see double descent in GLS for off-diagonal term estimation of the parameter vector α_j^* in the precision matrix. We fix n and K and 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

$$\frac{logn + n}{nr_n} \to 0,$$

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 can be seen. We also see this in our simulations.

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 have 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.

6. 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. Specific results regarding adaptive PCR and GLS will be discussed in subsections 6.1-6.2. 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. This can be seen through comparing the rates in Theorem 1 and Theorem 2 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 2: *Under Assumptions 1-5, with* $r_n/\delta_n \to 1$

$$\max_{1\le j\le p} \|\tilde{\Theta}_j' - \Theta_j'\|_2 = O_p(K\frac{\sqrt{plogp}}{\sqrt{n}} r_{w1}) + O_p(d_{2n} r_{w2}) + O_p\left(\max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\sqrt{\max(logp, logn)}}{n^{1/2}})\right) = o_p(1),$$

where r_{w1}, r_{w2} *are defined in (9)(10).*

Remarks. 1. First, we see that if $p/n \to [0, \infty)$, we obtain consistency and the price that we pay in terms of number of assets, p is just $\sqrt{logp}$ 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}, r_{w2} one of the keys to consistent estimation is the ratio $K/\bar{\xi}$, and this ratio needs to go to zero. So we cannot have a large hidden factor system and expect consistency. Key is the 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 2 is obtained 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 and Appendix Corollary A.3.

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 A.1(i) and Lemma 1(ii) with r_{w2} definition, under Assumptions 1-5, we can see in the proof of Theorem 2

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

Also we may get a better rate than the one in Theorem 2.

6.1. *Adaptive PCR Estimator*

In case of adaptive PCR, the random terms in the upper bound in Theorem 2, can be bounded as: $\bar{r} \le K$ wpa1, $\bar{\Psi}/\bar{\eta} \le 1$ wpa1, $\bar{\eta} \ge r_n$ with wpa1. So Theorem 2 rate can be simplified in adaptive PCR, with $\hat{\Theta}'_{PCR,j}$ representing the estimator of j th row of Θ. The following Corollary follows from Theorem 2 by replacing r_{w1}, r_{w2} with $r_{w1,PCR}, r_{w2,PCR}$ rates in (A.84)(A.85).

COROLLARY 4: *Let $p/n \to [0,\infty)$ and with Assumptions 1-5, and with $r_n/\delta_n \to 1$*

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

Remarks. 1. The price of consistency is $\sqrt{logp}$ 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.

2. We analyze how consistency can be achieved in this adaptive PCR. We start with Assumption 5(iii) which may be the most stringent term for consistency.

We use example (8) in Assumption 5(iii) explanation to illustrate our point. Take the case of $p/n \to \gamma$, with $\gamma \in (1,\infty)$, γ is a positive constant. Case of $p/n \to \gamma$, with $\gamma \in [0,1]$ is similar hence skipped. What if we increase n, p together but keep γ fixed. Consistency will require,

$$\frac{K\sqrt{log(\gamma n)}\sqrt{logn+K}}{\sqrt{nr_n}} \to 0, \tag{21}$$

which is possible with Assumption 5(ii) imposing $K = o(n^{1/3})$, and since r_n is nondecreasing in n . Then we need

$$\frac{K^{3/2}\sqrt{log(\gamma n)}}{\sqrt{\delta_n \bar{\xi}}} \to 0 \tag{22}$$

which can be possible with (8), $r_n/\delta_n \to 1$ and since

$$\frac{K^{3/2}\sqrt{log(\gamma n)}}{d_{1n}} \to 0,$$

with increasing d_{1n} with n, and can grow larger than $K^{3/2}\sqrt{logn}$. Next, we consider in Assumption 5(iii)

$$d_{2n}\sqrt{\frac{logn+K}{nr_n}} \to 0,$$

since $K = o(n^{1/3})$ and d_{2n} can diverge with n but not too fast (i.e. such as a slowly varying. function in n). Then for the second part of that consistency condition we also need with example (8)

$$d_{2n}\sqrt{\frac{K}{\bar{\xi}\delta_n}} \leq \frac{d_{2n}K^{1/2}}{d_{1n}} \to 0.$$

So the signal d_{1n} can be large and bring consistency.

3. Again a key component is signal-to-noise ratio with $\bar{\xi}$ being large estimation errors get smaller. Growing number of factors will have a negative effect on estimation errors.

4. Note that, a hidden factor model with principal components analysis is by Fan et al. (2013). Hidden factors are estimated through principal components and this is fed into covariance matrix of returns and then an inverse is taken through Sherman-Woodbury-Morrison matrix formula. Hidden factor rotation and estimation results in a rate of convergence which heavily depends on p growing. In problems, with few outcomes, p fixed, with larger sample size n, their precision matrix estimate is not consistent, and this can be seen in Theorem 3, and also equation (3.2) rate, and the paragraph below that equation explains this problem is due to effect of estimation of unknown factors. But note that their

error-covariance estimator is consistent only when p, n grow in Remarks 1-2-Theorem 1 on p.614 of Fan et al. (2013), the same is true with Theorem 3 in their paper. So their consistency analysis is limited to the growing p, n case. Also their paper does not take into account the signal-to-noise ratio $\bar{\xi}$ directly in rate of convergence results. Our estimator is consistent, with fixed p, and growing n case, and also tied explicitly to the signal-to-noise ratio.

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

6.2. *GLS Estimators*

When we consider the sub case of GLS with $p < n - 1$, we set wpa1 $\bar{r} < p$, $\bar{\Psi} = 0$, and $\bar{\eta} \geq c$, where c is a positive constant in Theorem 2 result above. So, denote the j th row of estimator of the precision matrix in GLS, $p - 1 < n$ case, as $\hat{\Theta}'_{GLS,j}$. To obtain Corollary 5, we use Corollary A.2 and Corollary 2 with r_{w1}, r_{w2} in Theorem 2 is replaced with r_{GLS-l} rate in (A.86).

COROLLARY 5: *Set $p/n = a_{2n} \to 0$, and under Assumptions 1-5, and $r_n/\delta_n \to 1$*

$$\max_{1 \leq j \leq p} \|\hat{\Theta}'_{GLS,j} - \Theta'_j\|_2 = O_p\left(\max(K\sqrt{a_{2n} logp}, 1) \max(\frac{\sqrt{logn + p}}{\sqrt{n}}, \frac{\sqrt{K}}{\sqrt{\bar{\xi}}})\right)$$
$$+ O_p\left(d_{2n} \max(\frac{\sqrt{logn + p}}{\sqrt{n}}, \frac{\sqrt{K}}{\sqrt{\bar{\xi}}})\right)$$
$$+ O_p\left(\max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\sqrt{logn}}{n^{1/2}})\right) = o_p(1).$$

Remarks.

1. 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 below. 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.

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

3. One remaining question is what if $p/n=a_{2n}\to 0$ then which method will have a better rate. Consider the first part of Assumption 5(iii), then in GLS

$$a_{2n}^{1/2}K(\sqrt{logp})r_{GLS-l}=a_{2n}^{1/2}K\sqrt{logp}\max(\sqrt{\frac{logn+p}{n}},\sqrt{\frac{K}{\bar{\xi}}}),$$

and in adaptive PCR the same rate is

$$a_{2n}^{1/2}K(\sqrt{logp})r_{w1,PCR}=a_{2n}^{1/2}K\sqrt{logp}\max(\sqrt{\frac{logn+K}{nr_n}},\sqrt{\frac{K}{\bar{\xi}\delta_n}}),$$

which will be smaller since $K<p$ and r_n is nondecreasing in n. The second part of Assumption 5(iii) follows a similar pattern and hence adaptive PCR is a better estimator asymptotically than GLS when $p-1<n$.

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 2, 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 $\hat{\Theta}_{INT,j}$. Note that we prove Corollary 6 by combining Corollary 3 and Corollary A.3.

COROLLARY 6: *Under Assumptions 1-4 5(ii), with signal-to-noise condition* $\sqrt{K}d_{2n}/\sqrt{\bar{\xi}}\to 0$, *and* $\min_{1\le j\le p}re(\Sigma_{U_{-j}})>Cn$ *and* $\tilde{U}_{-j}$ *matrix has independent entries,*

$$\max_{1\le j\le p}\|\hat{\Theta}'_{INT,j}-\Theta'_{INT,j}\|_2=O_p\left(\max(\frac{\sqrt{logn+n}}{\sqrt{nr_n}},\frac{\sqrt{K}d_{2n}}{\sqrt{\bar{\xi}}})\right)$$
$$+O_p\left(\max(\frac{K^{3/2}}{n^{1/2}},K\frac{\sqrt{logp)}}{n^{1/2}})\right)=o_p(1).$$

Remarks. 1. First of all we still get consistency with $p-1>n$, as long as signal-to-noise ratio dominates number of factors multiplied by d_{2n}. This can be seen easily by $K\sqrt{logp/n}\to 0$, and also we need signal-to-noise condition: $\sqrt{K}d_{2n}/\sqrt{\bar{\xi}}\to 0$.

2. It is not clear from the rates of convergence that GLS or adaptive PCR may do better. But it is possible that with low number of factors, adaptive PCR can perform better. But there is no tuning parameter estimation in GLS.

3. As described in Remark 2 of Corollary 3 and Remark 1 above, double descent is possible in GLS in precision matrix estimation. This extends the concept of double descent from linear regression to precision matrix estimation. This may have important consequences in empirics. We show one of these findings in Sharpe Ratio in US Stock market.

6.3. *Uses of Precision Matrix Estimate*

We have three examples of our estimators, and the first one is Sharpe-Ratio analysis, and the second one is the de-biased high dimensional estimator formation for the confidence intervals for structural parameters in cross section analysis, and the third one is network analysis. In this part, first, we provide a Sharpe Ratio metric that use precision matrix of asset returns. The metric is the Sharpe ratio (squared) from the maximum Sharpe Ratio portfolio. There are other metrics as discussed heavily in Caner et al. (2023) for example, but the emphasis of our paper is introduction of a new class of estimators for the precision matrix not the detailed analysis of financial metrics. We also think that the analysis of financial metrics can be done in another paper with this approach, since the proofs will be non-standard and will take a larger space with very detailed analysis through several matrix norm inequalities. The metric below is chosen since the contribution of the precision matrix is very clear to the financial metric. The metric is the Sharpe Ratio-squared from the maximum Sharpe Ratio portfolio. To obtain that formula we maximize

$$w_{MS}^{*}:=argmax_{w\in R^{p}}w'\mu \quad subject \quad \text{to} \quad w'\Sigma w\leq\sigma^{2},$$

where σ^2 is the variance of the portfolio. The optimal solution is

$$w_{MS}^{*}=\frac{\sigma\Sigma^{-1}\mu}{\sqrt{\mu'\Sigma^{-1}\mu}}.$$

Substituting this into Sharpe Ratio formula we have

$$SR_{MS} = \frac{w_{MS}^{*'}\mu}{w_{MS}^{*'}\Sigma w_{MS}^{*}} = \sqrt{\mu'\Sigma^{-1}\mu}.$$

The estimator is:

$$\widehat{SR}_{MS}^{2} := \hat{\mu}'\tilde{\Theta}'\hat{\mu}.$$

This can be found in Caner et al. (2023) as well as several other papers in the literature, and used in practice. What has been done in the literature so far is that, either impose sparsity in the covariance matrix of errors or sparsity on the precision matrix of errors, and prove consistency results on those assumptions. Without exact sparsity, the proofs will not be valid for $p > n$ case. The second issue in the current literature is that rates of convergence are tied to number of assets, p, sample size, n, and number of factors K. Here we provide explicit signal-to-noise ratio in the rate of convergence for our metrics.

In our second example, we show how our precision matrix can be modified-used to form a debiased lasso estimator for estimating structural parameters, and this type of estimator can then be used to have confidence intervals for these structural parameters. Our setup is a simple linear high dimensional linear regression as in Caner and Kock (2018) and van de Geer et al. (2014).

$$Z = Y\beta + \epsilon,$$

where Z represents the outcome variables here, and it is $n \times 1$ vector, where $i = 1, \cdots, n$ cross-section units (instead of t index in the other parts of the paper, we use i for cross section data), and $Y : n \times p$ matrix, where each row represents cross section units, and $j = 1, \cdots, p$ represent regressors. Let $\beta : p \times 1$ be the structural parameters of interest, and let s_0 of them to be nonzero, allow s_0 n to be nondecreasing in n, $\epsilon : n \times 1$ represents the errors. Also let $EY_i\epsilon_i = 0$, and ϵ_i is iid as well as Y_i. For example, Z can represent average (over time) growth rate of GDP of a n group of countries, and each column of Y represents each regressor, such as public debt to GDP ratio, current account deficit, inflation, that contributes to this average GDP growth rate. We also assume existence of the following relation

$$Y = F\bar{D}' + u,$$

as can be formed by (4)(5). We assume that regressors are governed by latent factors. Note that using lasso estimation of β results in biased estimates and is not uniformly consistent, hence any confidence intervals will not be valid and not asymptotically honest as discussed in Caner and Kock (2018). To solve this issue, a method is debiasing the lasso estimators. However, debiased lasso estimators depend on the estimation of the inverse of the population inner-product EY_iY_i', where $Y_i : p \times 1$ vector, (row of Y represented in column form). It is not possible to use sample $Y'Y/n$ and invert it to get an estimate since if $p > n$, $Y'Y/n$ is singular. First we show how to form $[EY_iY_i']^{-1}$ which is a $p \times p$ matrix. This can be formed exactly with the analysis in Appendix B.1, with only imposing $\mu_j = 0, \mu_{-j} = 0_{p-1}$, where 0_{p-1} vector is a p-1 vector of zeroes in each element. Also see p.157 of Caner and Kock (2018) for this as well.Our estimator for $\tilde{\Theta}$ is exactly the same as in Section 4 with no change. Then the debiased estimator, $\hat{b}$, can be formed through, using the analysis in Section 3.1, equations (7)-(9) of Caner and Kock (2018) but with our new estimator of inverse of EY_iY_i' rather than the nodewise regression in Caner and Kock (2018),

$$\hat{b} = \hat{\beta} + \tilde{\Theta}Y'(Z - Y\hat{\beta})/n,$$

where $\hat{\beta}$ is the lasso estimator for β by regressing a l_1 penalized regression of Z on Y. We want to focus on this estimator and how to form confidence intervals for our next project.

7. SIMULATION

7.1. *Setup*

By (2) and (3) and Assumption 4, for each $t = 1, ..., n$, we have:

$$y_t = \bar{D} f_t + u_t \tag{23}$$

Where $y_t = (y_{jt}, y_{-jt}')' : p \times 1$ is the vector of excess return, $\bar{A} = (a_j, A_{-j}') : p \times K$ is the matrix of factor loadings, and $u_t = (u_{jt}, u_{-jt})' : p \times 1$, hidden-factors $f_t : K \times 1$ are drawn from $\mathcal{N}(\mu_f, \Sigma_f)$, where μ_f is $p \times 1$ vector with each element equal to 0.5, Σ_f is the hidden-factor's covariance. Following Bing et al. (2021), we set the diagonal elements of Σ_f is the $K-$ length sequence from 2.5 to 3 with equal increments. For example, if $K = 3$, then $\Sigma_f = diag(2.5, 2.75, 3)$. The the off diagonal, i-th row, j-th column of factor covariance

matrix defined as:

$$[\Sigma_f]_{i,j} = (-1)^{i+j}(0.3)^{|i-j|}\min([\Sigma_f]_{i,i}, [\Sigma_f]_{j,j}) \tag{24}$$

Each element of $\bar{D}$ is drawn independently from $\mathcal{N}(0.5, 1/\sqrt{K})$. The idiosyncratic component, u_t, is drawn from $\mathcal{N}(0, \Sigma_u)$. In the first experiment, we define $\Sigma_u : p \times p$ as a diagonal matrix with diagonal elements sampled from Unif$(1, 3) \times \log(p)$. The results are presented here Table I-III, and represent $K = 3, 5, 20$ respectively. In addition, to further demonstrate how our methods behave as the signal-to-noise ratio (SNR) changes, we rewrite (23) in compact form:

$$Y = F\bar{D}' + U \tag{25}$$

Similar to Bing et al. (2021), we multiply $\bar{D}$ by a scalar α chosen from the range $0.1, ..., 1.5$. Fixing $K = 20, n = 400$, we experiment in both low- and high- dimensional case, with $p = 50$ and $p = 450$ respectively. Then for each the α, we calculate the SNR and plot the estimation of errors precision matrix for each methods. The results are presented in Figures 1-2.

We aim to address several questions through simulations. First, as described in the theorems, we investigate whether the estimation errors from our methods decrease when $\gamma = p/n$ [2] is fixed and both p and n jointly increase. Secondly, we compare the performance of our new methods with existing methods. Thirdly, we examine the impact of increasing the number of hidden factors. Lastly, we study the impact of the signal-to-noise ratio (SNR) on estimation when the number of hidden factors K is fixed.

7.2. *Metrics*

We consider metric of estimation error for the precision matrix as stated in Theorem 1. The model comparisons include several approaches from the literature. We consider the nonlinear shrinkage methods of Ledoit and Wolf (2017)(NL-LW), POET method from Fan et al. (2013), and the nodewise (NW) method from Callot et al. (2021) Our new method introduces several variations: Inter-GLS, PCR-3F, PCR-5F, PCR-20F and PCR-Adaptive.

[2] $\gamma = p/n = 0.5$ in low-dimensional set up and $\gamma = p/n = 1.5$ in high-dimensional set up

These denote the use of a GLS estimator, Principal component with fixed 3, 5, or 20 hidden factors, and an adaptive method to select the number of hidden factors, respectively.

7.3. *Results*

Table I- III shows the simulation results. The values in each cell show the average absolute estimation error for estimating the precision matrix. Each two-column block in the tables shows the result for a different sample size. The first column present results for $p = n/2$ (low-dimensional case), while the second corresponds to $p = 3n/2$ (high-dimensional case). We ran 100 iterations in each simulation set up.

First, the tables show that our method achieves consistency as stated in Corollary 4 for adaptive PCR. Analyzing Table I with $K = 3$, we observe that for $p = n/2$, with $n = 100$, our Adaptive-PCR has an estimation error at 0.0583. This error declines to 0.0251 when $p = n/2$ and $n = 400$. This indicates that as both n and p increase jointly, the estimation error decreases. Note that our Inter-GLS estimator in low dimensions (OLS) is consistent only when $p/n \to 0$ as shown in Corollary 5. However, it sill does well with $p = n/2$ in our Tables I- III. To give an example, from Table I, the error is 0.4583 at $n = 100$, and declines to 0.3369 at $n = 400$. We also experimented with $p/n \to 0$ in our setup as expected Inter-GLS error terms decline to zero fast, and these results are available from authors on demand.

To give another example, but from a high dimensional case, from Table III with $K = 20$, the estimation error of our Adaptive-PCR is 0.0674 and Inter-GLS is 0.1554 , with $p = 3n/2, n = 100$, then these errors decline to 0.0359 and 0.1313 with $p = 3n/2, n = 400$. Note that Inter-GLS in high dimensions do not need $p/n \to 0$ as this is clear from Corollary 6.

Next, we consider which method achieves the smallest estimation errors. Our Adaptive-PCR dominates all other techniques except fixed PCR-K. In additional, our Adaptive-PCR effectively identifies the correct number of hidden factors almost all of the scenarios. For instance, at $p = 3n/2, n = 400$ number of factor $K = 5$ in Table II, the estimation of error of both PCR-5F and Adaptive-PCR are the same at 0.0302, which also is the smallest estimation error among all methods.

Lastly, the number of hidden factors used in our PCR methods—whether underfitting or overfitting—significantly impacts the estimation error. When the correct number of hidden factors is $K = 3$, with $n = 400$ and $p = 3n/2$, the estimation error of PCR-3F is 0.0302 in Table I. However, overfitting the model by using $K = 5,$ or 20 hidden factors increases the estimation error; for example, using PCR-20, the estimation error increases to 0.0401.

TABLE I

ESTIMATION ERROR, $K = 3$

Estimator	**n=100**		**n=200**		**n=400**	
	p=n/2	p=3n/2	p=n/2	p=3n/2	p=n/2	p=3n/2
NW	0.0752	0.1016	0.0617	0.0843	0.0521	0.0669
POET	47.0373	76.4717	59.6223	101.4291	84.7820	150.7152
Ledoit-Wolf	40.4218	69.1745	55.2778	97.6674	82.5374	150.9011
Inter-GLS	0.4583	0.2041	0.3858	0.1763	0.3369	0.1562
PCR-3F	0.0583	0.0738	0.0385	0.0464	0.0251	0.0302
PCR-5F	0.0641	0.0803	0.0408	0.0489	0.0263	0.0313
PCR-20F	0.1276	0.1463	0.0637	0.0712	0.0361	0.0401
PCR-Adaptive	0.0583	0.0738	0.0385	0.0464	0.0251	0.0302

TABLE II

ESTIMATION ERROR, $K = 5$

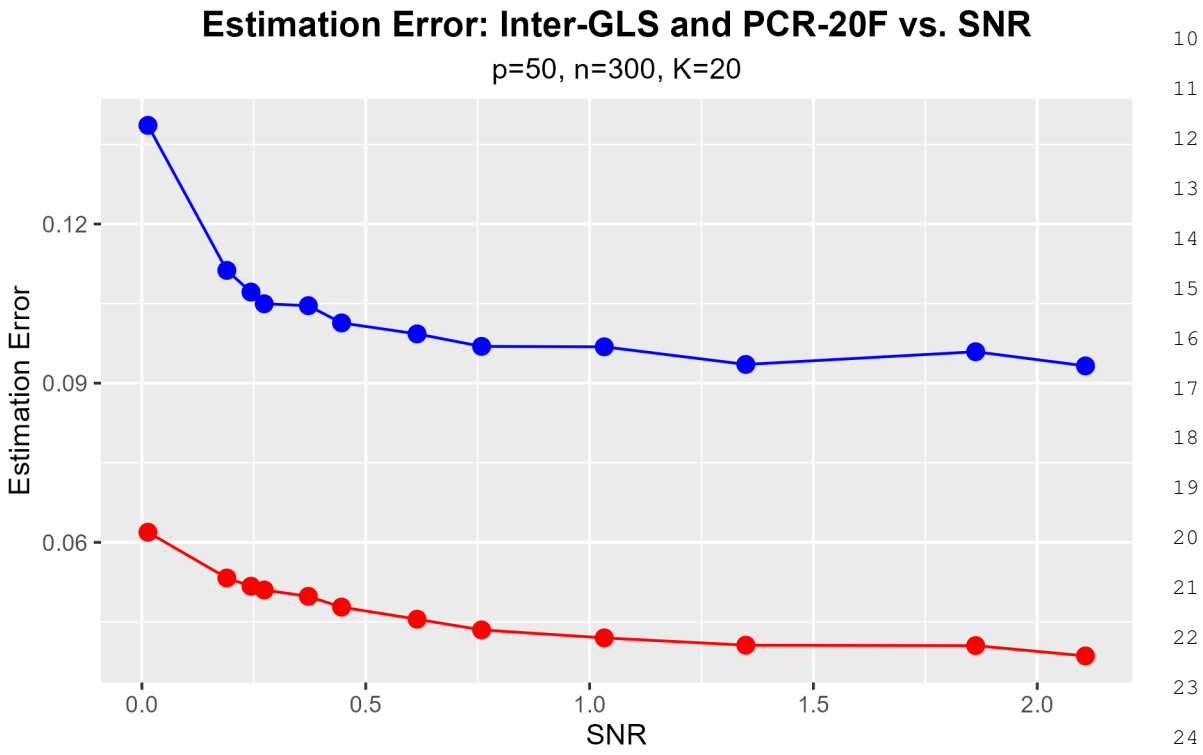
Estimator	**n=100**		**n=200**		**n=400**	
	p=n/2	p=3n/2	p=n/2	p=3n/2	p=n/2	p=3n/2
NW	0.0778	0.1042	0.0683	0.0888	0.0578	0.0728
POET	44.6537	72.7433	58.4469	98.6137	78.9982	135.5738
Ledoit-Wolf	38.8199	66.0089	54.5050	94.8428	76.9331	135.0054
Inter-GLS	0.4307	0.1752	0.3840	0.1577	0.3360	0.1417
PCR-3F	0.0564	0.0686	0.0488	0.0593	0.0452	0.0524
PCR-5F	0.0580	0.0732	0.0404	0.0503	0.0259	0.0302
PCR-20F	0.1193	0.1397	0.0631	0.0725	0.0355	0.0395
PCR-Adaptive	0.0559	0.0732	0.0403	0.0503	0.0259	0.0302

TABLE III

ESTIMATION ERROR, $K = 20$

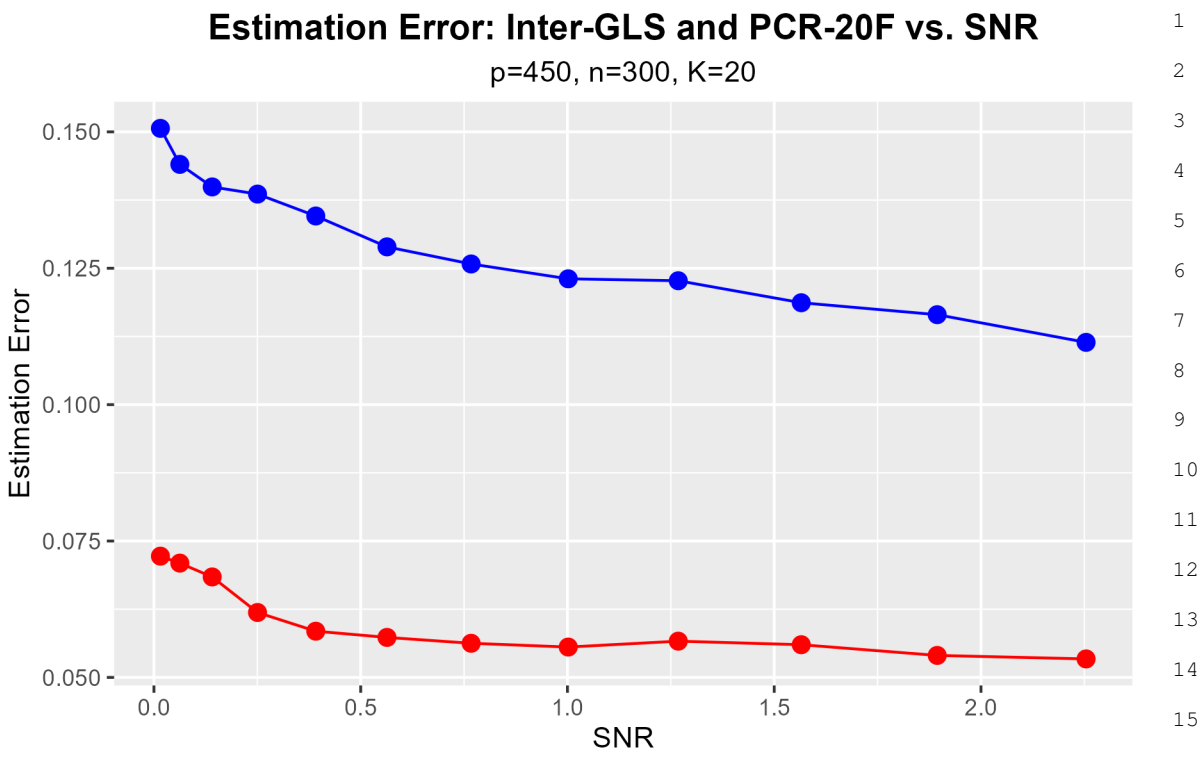
Estimator	**n=100**		**n=200**		**n=400**	
	p=n/2	p=3n/2	p=n/2	p=3n/2	p=n/2	p=3n/2
NW	0.0858	0.1139	0.0703	0.1024	0.0620	0.0870
POET	81.4564	140.6064	110.4248	192.8667	153.6446	270.0705
Ledoit-Wolf	77.0944	135.4704	106.8768	188.9262	150.9729	267.2804
Inter-GLS	0.3987	0.1554	0.3552	0.1431	0.3138	0.1313
PCR-3F	0.0521	0.0694	0.0484	0.0630	0.0471	0.0574
PCR-5F	0.0538	0.0662	0.0454	0.0582	0.0435	0.0522
PCR-20F	0.1047	0.1172	0.0544	0.0596	0.0334	0.0364
PCR-Adaptive	0.0526	0.0674	0.0461	0.0487	0.0359	0.0344

To examine the performance of our methods as the signal-to-noise ratio (SNR) changes, we analyze two figures: Figure 1 shows the estimation error of the precision matrix in the low-dimensional case, and Figure 2 presents the results in high-dimensional cases. We include the performance of Inter-GLS in blue and PCR-20F in red. From the figures, we observe that the performance of GLS and PCR improves as the SNR increases from 0 to 1. As the SNR increases beyond 1, the performance of PCR-20F continues to improve, while the performance of Inter-GLS stabilizes. Our PCR-20R consistently outperforms Inter-GLS across the entire SNR range for both low- and high-dimensional cases.



8. EMPIRICAL APPLICATION

The objective of the empirical application is to conduct an out-of-sample comparison among estimations. We estimate Maximum Sharpe Ratio (MSR) portfolio 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 %.

8.1. *Metrics*

We use a rolling window set up for the out-of-sample estimation of the Sharpe Ratio following Callot et al. (2021) and Caner et al. (2023). The sample size n is divided into in-sample $(1 : n_I)$ and out-of-sample $(n_I + 1 : n)$. Initially, we estimate the portfolio $\hat{w}_{n_I}$ in the in-sample period and the out-of-sample portfolio returns $\hat{w}'_{n_I} y_{n_I+1}$. Then, we roll the window forward by one element $(2 : n_I + 1)$, forming a new in-sample portfolio $\hat{w}_{n_I+1}$

and caculating out-of-sample portfolio returns $\hat{w}'_{n_{I+1}} y_{n_I+2}$. This process is repeated until the end of the sample.

The out-of-sample average return and variance without transaction costs are computed as follows:

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

We estimate the Sharpe Ratios without transaction costs.

$$SR = \frac{\hat{\mu}_{os}}{\hat{\Sigma}_{y,os}}$$

When transaction costs are present, the definition of the metrics is modified. Let c be the transaction cost, which we set 10 basis points as. 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

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

where $\hat{w}^+_{j,t} = \dfrac{\hat{w}_{j,t}(1 + y_{j,t+1})}{1 + y_{P,t+1}}$, with $y_{j,t+1}$ representing the excess return added to risk-free-rate for jth, 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 portfolio returns with transaction costs as:

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

The portfolio returns are replaced by the returns with transaction costs when we calculated the Sharpe Ratio with transaction costs. Finally, portfolio turnover is defined as:

$$\text{turnover} = \sum_{j=1}^{p} |\hat{w}_{j,t+1} - \hat{w}^+_{j,t}|$$

8.2. *Methods*

We conduct a comprehensive empirical analysis by fixing in-sample $n_I = 240$ and considering both low-dimensional and high-dimensional cases across various portfolio sizes. Specifically, we selected 10, 100, and 200 stocks from the pool to estimate the models with the same stocks for all windows in the low-dimensional experiments. For the high-dimensional experiments, we randomly choose 325 and 357 (including all available stocks). We also analyzed portfolios with 25, 50, 100, and 300 stocks, however, since theses results are similar, they are not included here. In total, we examine 9 different scenarios- 5 in low-dimensional and 4 in high-dimensional cases but we report 5 of them to save space.

The model comparisons include several approaches from the literature. We consider the nonlinear shrinkage methods of Ledoit and Wolf (2017)(NL-LW) and the single-factor nonlinear shrinkage (SF-NL-LW), POET method from Fan et al. (2013), and the nodewise (NW) method from Callot et al. (2021), along with the feasible residual-based nodewise approach from Caner et al. (2023). In particular, RNW-SF, RNW-3F, and RNW-5F refer to residual-based nodewise regressions using a single-factor model (market as the only factor), three-factor model, and five-factor Fama and French model, respectively. We also include equally weighted portfolio for comparison. Our new method introduces several variations: Inter-GLS, PCR-1F, PCR-3F, PCR-5F, PCR-7F and PCR-Adaptive. These denote the use of a GLS estimator, Principal component with fixed 1, 3, 5, 7 hidden factors, and an adaptive method to select the number of hidden factors, respectively. We use these number of factors since they are used in the empirics in the literature and seven factor is used to show large number of factors. We did not try 20 factors since it is not working well in simulations.

8.3. *Results*

We report the results for Maximum Sharpe Ratio portfolio, with and without transaction costs. Results for the low-dimensional cases are presented in Table IV - VI, while high-dimensional cases are shown in Tables VII - VIII. 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 IV, for a portfolio of 10 stocks ($p = 10, n_I = 240$), PCR-1F has 0.2260 as the best Sharpe Ratio. In Table VI 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 VIII, 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,

$$H_o : SR_{winner} \leq SR_0 \text{ vs } H_a : SR_{winner} > SR_0, \tag{26}$$

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 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.

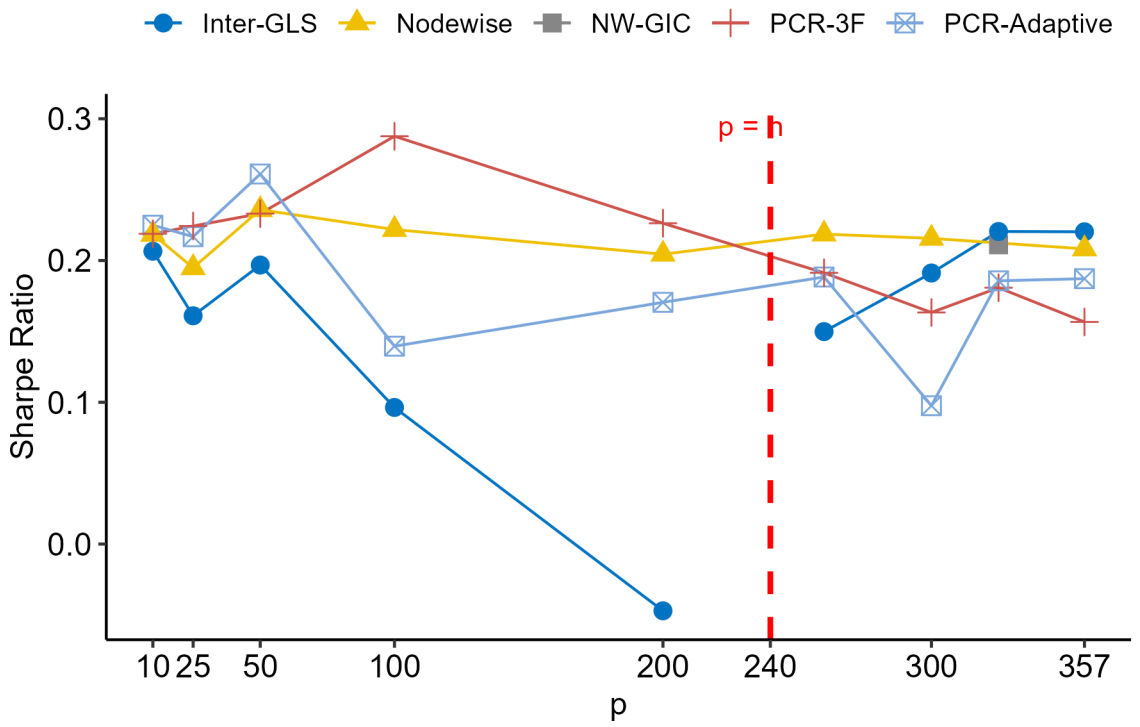


FIGURE 1.—MSR Sharpe Ratio vs. p

TABLE IV

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	NA	NA	NA	NA
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	NA	NA	NA	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 V

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	NA	NA	NA	NA
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	NA	NA	NA	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 VI

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	NA	NA	NA	NA
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	NA	NA	NA	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 VII

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	NA	NA	NA	NA
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	NA	NA	NA	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 VIII

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	NA	NA	NA	NA
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	NA	NA	NA	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

APPENDIX

APPENDIX A: PROOFS

Proof of Lemma 1

Using (4) and (11) we get

$$\tilde{\alpha}_j - \alpha_j^* = \hat{B}(Y_{-j}\hat{B})^+ F a_j + \hat{B}(Y_{-j}\hat{B})^+ u_j - \alpha_j^*.$$

Then we have the following inequality

$$\|\tilde{\alpha}_j - \alpha_j^*\|_2^2 \leq 3\|\hat{B}(Y_{-j}\hat{B})^+ u_j\|_2^2 + 3\|\hat{B}(Y_{-j}\hat{B})^+ F a_j\|_2^2 + 3\|\alpha_j^*\|_2^2. \tag{A.1}$$

Our proof is a detailed extension of Lemma 17 of Bing et al. (2021) to precision matrix estimation from linear regression framework and we provide all details in the proof, and also to use in high dimensional portfolio formation we need uniform results over $j = 1, \cdots, p$, unlike pointwise result for one possible asset in Bing et al. (2021).

Each of the right side terms in (A.1) will be analyzed in each of Steps 1-3.

Step 1.

Start with the first right side term in (A.1)

$$\|\hat{B}(Y_{-j}\hat{B})^+ u_j\|_2^2 = u_j' M_j u_j, \tag{A.2}$$

with $M_j : n \times n$ matrix defined as

$$M_j := (Y_{-j}\hat{B})^{+'} \hat{B}' \hat{B}(Y_{-j}\hat{B})^+. \tag{A.3}$$

Note that u_j is a subgaussian vector, and independent of Y_{-j}, and $\hat{B}$ by Assumption 1 and the definition of $\hat{B}$, with M_j being a symmetric positive semi-definite matrix. Then we use Lemma 21 of Bing et al. (2021) or proof of Lemma 16 in Bing et al. (2021) to have, with $H = M_j, \epsilon = u_j$ there,

$$u_j' M u_j \leq 2\gamma_e^2 \sigma_j^2 [2\|M_j\|_2 logn + tr(M_j)], \tag{A.4}$$

with probability at least 1-1/n. So we can have with probability at least 1-1/n

$$\|\hat{B}(Y_{-j}\hat{B})^{+}u_j\|_2^2 \leq \gamma_e^2\sigma_j^2[2\|M_j\|_2 logn + tr(M_j)]. \tag{A.5}$$

We consider the trace term above in (A.5)

$$tr(M_j) = tr[(Y_{-j}\hat{B})^{+'}\hat{B}'\hat{B}(Y_{-j}\hat{B})^{+}].$$

First of all

$$\begin{aligned} tr[M_j] &\leq rank(M_j)\|M_j\|_2 = rank(Y_{-j}\hat{B})^{+}\|M_j\|_2 \\ &= rank(Y_{-j}\hat{B})\|M_j\|_2 = rank(Y_{-j}P_{\hat{B}})\|M_j\|_2, = \hat{r}_j\|M_j\|_2 \end{aligned} \tag{A.6}$$

where the first inequality is due to M_j being positive semidefinite, and $n \times n$ matrix and Exercise 7.49 of Abadir and Magnus (2005), and the first equality is by Exercise 8.27a and 4.24a of Abadir and Magnus (2005), and the second equality is by Exercise 10.28 of Abadir and Magnus (2005), and the third equality is again by Exercise 10.28 of Abadir and Magnus (2005) with $P_{\hat{B}} = \hat{B}\hat{B}^{+}$ in p.8 of Bing et al. (2021), and the last equality is just the definition of the complexity $\hat{r}_j$. Now we consider, as proved in (44) of Bing et al. (2021)

$$\|M_j\|_2 \leq (n\hat{\eta}_j)^{-1}, \tag{A.7}$$

where $\hat{\eta}_j := \hat{\sigma}^2_{\hat{r}_j}(Y_{-j}P_{\hat{B}})/n$ which is the $\hat{r}_j$ th largest singular value, squared, of the matrix $Y_{-j}P_{\hat{B}}$, and divided by n.

So with probability at least $1 - 1/n$ by (A.5)-(A.7) in (A.2)

$$\max_{1\leq j\leq p}\|\hat{B}(Y_{-j}\hat{B})^{+}u_j\|_2 \leq 2\gamma_e^2\sigma^2[\frac{2logn}{n\bar{\eta}} + \frac{\bar{r}}{n\bar{\eta}}]. \tag{A.8}$$

Step 2.

Now consider the second term on the right side of (A.1), we use $A'_{-j}A^{+'}_{-j} = I_K$ and (5), with $I_{p-1} = P_{\hat{B}} + M_{\hat{B}}$

$$\begin{aligned} \hat{B}(Y_{-j}\hat{B})^{+}F &= \hat{B}(Y_{-j}\hat{B})^{+}FA'_{-j}A^{+'}_{-j} = \hat{B}(Y_{-j}\hat{B})^{+}(Y_{-j} - U_{-j})A^{+'}_{-j} \\ &= \hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}P_{\hat{B}}A^{+'}_{-j} + \hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}M_{\hat{B}}A^{+'}_{-j} \\ &- \hat{B}(Y_{-j}\hat{B})^{+}U_{-j}A^{+'}_{j} \end{aligned}$$

Use simple inequality $(a+b+c)^2 \leq 3a^2+3b^2+3c^2$ inequality, with $M_{\hat{B}} := I_{p-1} - P_{\hat{B}}$,

$$\begin{aligned}\|\hat{B}(Y_{-j}\hat{B})^{+}Fa_j\|_2^2 \leq & 3\|\hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}P_{\hat{B}}A_{-j}^{+'}a_j\|_2^2 \\ & + 3\|\hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}(I_{p-1}-P_{\hat{B}})A_{-j}^{+'}a_j\|_2^2 \\ & + 3\|\hat{B}(Y_{-j}\hat{B})^{+}U_{-j}A_{j}^{+'}a_j\|_2^2. \end{aligned} \tag{A.9}$$

Step 2a.

We analyze each term above. So

$$\|\hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}P_{\hat{B}}A_{-j}^{+'}a_j\|_2^2 \leq \|\hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}P_{\hat{B}}\|_2^2\|A_{-j}^{+'}a_j\|_2^2. \tag{A.10}$$

Note that we can write the first term on the right side of (A.10) as

$$\|\hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}P_{\hat{B}}\|_2 = \|\hat{B}(\hat{B}'Y_{-j}'Y_{-j}\hat{B})^{+}\hat{B}'Y_{-j}'Y_{-j}P_{\hat{B}}\|_2, \tag{A.11}$$

where we use Moore-Penrose inverse rule (134) in Bunea et al. (2022) which is for a rectangular matrix S, we have $(S'S)^{+}S' = S^{+}$, which we input $S = Y_{-j}\hat{B}$ in our case above. To simplify the expressions see that by the proof of Lemma 15 of Bing et al. (2021) $P_{\hat{B}} = V_1V_1'$, where $\hat{B} = V_1DV_2'$ singular value decomposition with V_1, V_2 are orthonormal matrices and with dimension of $p-1 \times r_0, q \times r_0$ respectively with D as r_0 dimensional diagonal matrix, and rank $(\hat{B}) = r_0$. Rewrite right side of (A.11) as

$$\|\hat{B}(Y_{-j}\hat{B})^{+}Y_{-j}P_{\hat{B}}\|_2 = \|V_1DV_2'(V_2DV_1'Y_{-j}'Y_{-j}V_1DV_2')^{+}V_2DV_1'Y_{-j}'Y_{-j}V_1V_1'\|_2. \tag{A.12}$$

Then in (A.12) right side see that

$$(V_2DV_1'Y_{-j}'Y_{-j}V_1DV_2')^{+} = V_2D^{-1}(V_1'Y_{-j}'Y_{-j}V_1)^{+}D^{-1}V_2', \tag{A.13}$$

which is shown in the proof of Lemma 15 of Bing et al. (2021). Then use (A.13) in (A.12), $V_2'V_2 = I_{r_0}$

$$\begin{aligned} V_1DV_2'(V_2DV_1'Y_{-j}'Y_{-j}V_1DV_2')^{+}V_2DV_1' &= V_1DV_2'V_2D^{-1}[V_1'Y_{-j}'Y_{-j}V_1]^{+}D^{-1}V_2'V_2DV_1' \\ &= V_1[V_1'Y_{-j}'Y_{-j}V_1]^{+}V_1' \end{aligned}$$

Use the last expression in (A.12) right side

$$\|V_1 DV_2'(V_2 DV_1' Y_{-j}' Y_{-j} V_1 DV_2')^{+} V_2 DV_1' Y_{-j}' Y_{-j} V_1 V_1'\|_2 = \|V_1[V_1' Y_{-j}' Y_{-j} V_1]^{+} V_1' Y_{-j} Y_{-j}' V_1 V_1'\|_2.$$

Then clearly since $V_1 : p-1 \times r_0$ matrix, and Exercise 10.26 of Abadir and Magnus (2005) shows $[V_1' Y_{-j}' Y_{-j} V_1]^{+} V_1' Y_{-j}' Y_{-j} V_1$ is a symmetric idempotent matrix and all idempotent matrices eigenvalues are either 0 or 1, and since $V_1' V_1 = I_{r_0}$

$$\|V_1[V_1' Y_{-j}' Y_{-j} V_1]^{+} V_1' Y_{-j}' Y_{-j} V_1 V_1'\|_2 \leq \|V_1\|_2 \|[V_1' Y_{-j}' Y_{-j} V_1]^{+} V_1' Y_{-j}' Y_{-j} V_1\|_2 \|V_1'\|_2 \leq 1.$$

So use this last inequality in (A.12)

$$\|\hat{B}(Y_{-j}\hat{B})^{+} Y_{-j} P_{\hat{B}}\|_2 \leq 1, \tag{A.14}$$

Next part of the proof simplifies certain expressions in Lemma 17 of Bing et al. (2021) and ties that to signal-to-noise ratio. Now consider the following term which is common in almost all terms on upper bounds of our estimator, and will link to signal to noise ratio, which is the second term on the right side of (A.10)

$$\|A_{-j}^{+'} a_j\|_2^2 = a_j' A_{-j}^{+} A_{-j}^{+'} a_j = a_j'(A_{-j}' A_{-j})^{-1} a_j,$$

since $A_{-j}^{+} = (A_{-j}' A_{-j})^{-1} A_{-j}^{'}$ with $p-1 > K$. Next

$$\begin{aligned} a_j' A_{-j}^{+}(A_{-j}^{+})' a_j &= a_j' \Sigma_f^{1/2} (\Sigma_f^{1/2} A_{-j}' A_{-j} \Sigma_f^{1/2})^{-1} \Sigma_f^{1/2} a_j \\ &\leq \frac{a_j' \Sigma_f a_j}{\lambda_K(A_{-j} \Sigma_f A_{-j}')}, \end{aligned} \tag{A.15}$$

where we use first Rayleigh inequality on p.182 of Abadir and Magnus (2005) and then Theorem 2.6.3b of Horn and Johnson (2013), where $\lambda_K(.)$ is the K th largest eigenvalue for the matrix $A_{-j}\Sigma_f A_{-j}' : p-1 \times p-1$.

By (6)(7), we try to rewrite (A.15) right-side.

$$\begin{aligned} \lambda_K(A_{-j}' \Sigma_f A_{-j}) &= \left[\frac{\lambda_K(A_{-j}' \Sigma_f A_{-j})}{\|\Sigma_{U,-j}\|_2}\right] \|\Sigma_{U,-j}\|_2 \\ &= \xi_j \|\Sigma_{U,-j}\|_2 \geq \bar{\xi} \|\Sigma_{U,-j}\|_2 \\ &\geq \bar{\xi} \delta_n, \end{aligned} \tag{A.16}$$

where we use $\delta_n := \min_{1\le j\le p}\|\Sigma_{U,-j}\|_2 \ge c > 0$ by Assumption 2(ii), where δ_n is non-decreasing in n. Then by $\max_{1\le j\le p}\max_{1\le k\le K}|a_{jk}| \le C < \infty$, with a_{jk} as the k th element of row a_j in A, which is Assumption 4(iii)

$$a_j'\Sigma_f a_j \le \|a_j\|_2^2 Eigmax(\Sigma_f) \le CK, \tag{A.17}$$

by $Eigmax(\Sigma_f) \le C < \infty$ Assumption 3. So the first term on the right side of (A.15) can be bounded as as

$$\max_{1\le j\le p}\|A_{-j}^{+'}a_j\|_2^2 = \max_{1\le j\le p}\left[\frac{(a_j'\Sigma_f a_j)}{\lambda_K(A_{-j}\Sigma_f A_{-j}')}\right] \le \frac{CK}{\bar{\xi}\delta_n}. \tag{A.18}$$

Clearly by (A.14)(A.18) in (A.10)

$$\max_{1\le j\le p}\|\hat{B}(Y_{-j}\hat{B})^+ Y_{-j}P_{\hat{B}}A_{-j}^{+'}a_j\|_2^2 \le \frac{CK}{\bar{\xi}\delta_n}. \tag{A.19}$$

Step 2b.

Take the second term on the right side of (A.9)

$$\|\hat{B}(Y_{-j}\hat{B})^+ Y_{-j}(I_{p-1} - P_{\hat{B}})A_{-j}^{+'}a_j\|_2^2 \le \|\hat{B}(Y_{-j}\hat{B})^+\|_2^2 \|Y_{-j}(I_{p-1} - P_{\hat{B}})A_{-j}^{+'}a_j\|_2^2. \tag{A.20}$$

See that the first term on the right side of by $\|S_b\|_2^2 = \|S_b'S_b\|_2 = \|S_bS_b'\|_2$ for a rectangular matrix $S_b := \hat{B}(Y_{-j}\hat{B})^+$ and M_j definition in (A.3)

$$\|\hat{B}(Y_{-j}\hat{B})^+\|_2^2 = \|M_j\|_2 \le \frac{1}{n\hat{\eta}_j}, \tag{A.21}$$

where the last inequality is by (A.7). Then by definition of $\hat{\Psi}_j$ in (17)

$$\|Y_{-j}(I_{p-1} - P_{\hat{B}})\|_2^2 = \sigma_1^2[(Y_{-j}(I_{p-1} - P_{\hat{B}})] = n\hat{\Psi}_j. \tag{A.22}$$

Now consider (A.18)(A.21)(A.22) on the second term-right side of (A.9), with $\bar{\Psi} := \max_{1\le j\le p}\hat{\Psi}_j, \bar{\eta} := \min_{1\le j\le p}\hat{\eta}_j$, with $\hat{\eta}_j$ defined in (16)

$$\max_{1\le j\le p}\|\hat{B}(Y_{-j}\hat{B})^+ Y_{-j}(I_{p-1} - P_{\hat{B}})A_{-j}^{+'}a_j\|_2^2 \le C\left(\frac{\bar{\Psi}}{\bar{\eta}}\right)\left(\frac{K}{\bar{\xi}\delta_n}\right). \tag{A.23}$$

Step 2c.

Consider the third term on the right side of (A.9)

$$\|\hat{B}(Y_{-j}\hat{B})^{+}U_{-j}A_j^{+'}a_j\|_2^2 \leq \|\hat{B}(Y_{-j}\hat{B})^{+}\|_2^2\|U_{-j}A_j^{+'}a_j\|_2^2, \tag{A.24}$$

where we analyze the following on the right side of (A.24)

$$\|U_{-j}A_j^{+'}a_j\|_2^2 = a_j'A_{-j}^{+}U_{-j}U_{-j}'A_{-j}^{+'}a_j.$$

Since $U_{-j}A_{-j}^{+'}a_j$ is $\gamma_w\sqrt{a_j'A_{-j}^{+}\Sigma_{U,-j}A_{-j}^{+'}a_j}$ subgaussian vector for each $j = 1, \cdots, p$ via Assumption 4, apply Lemma 21 of Bing et al. (2021) with $\epsilon_j := U_{-j}A_j^{+'}a_j$ and $H := I_n$ and $\gamma_{e,j}^2 = \gamma_w^2 a_j'A_j^{+}\Sigma_{U,-j}A_j^{+'}a_j$ with $t = logn$ as in (41) of Bing et al. (2021)

$$P[\|U_{-j}A_{-j}^{+'}a_j\|_2^2 > (n + 2logn + 2\sqrt{nlogn})\gamma_{e,j}^2] \leq 1/n. \tag{A.25}$$

Now we simplify $\gamma_{e,j}^2$ term. See that, uniformly over j, by Assumption 2 and (A.18)

$$\begin{aligned}\max_{1\leq j\leq p}\gamma_w^2 a_j'A_j^{+}\Sigma_{U,-j}A_j^{+'}a_j &\leq [max_{1\leq j\leq p}Eigmax(\Sigma_{U,-j})][\max_{1\leq j\leq p}\gamma_w^2 a_j'A_j^{+}A_j^{+'}a_j] \\ &\leq \frac{\gamma_w^2 C r_n K}{\bar{\xi}\delta_n}.\end{aligned} \tag{A.26}$$

Use (A.21) and (A.25)(A.26) in (A.24) with probability at least $1 - 1/n$ (note that we ignore the smaller order terms than the right side term in (A.26) since they play no role in the subsequent analysis), by combining γ_w^2 into C,

$$\max_{1\leq j\leq p}\|\hat{B}(Y_{-j}\hat{B})^{+}U_{-j}A_j^{+'}a_j\|_2^2 \leq \frac{CKr_n}{\bar{\eta}\bar{\xi}\delta_n}. \tag{A.27}$$

Combine all Steps 2a-2c results (A.19)(A.23)(A.27) in (A.9) to have, with probability at least $1 - 1/n$

$$\max_{1\leq j\leq p}\|\hat{B}(Y_{-j}\hat{B})^{+}Fa_j\|_2^2 \leq \frac{3CK}{\bar{\xi}\delta_n} + \frac{3\bar{\Psi}K}{\bar{\eta}\bar{\xi}\delta_n} + \frac{3CKr_n}{\bar{\eta}\bar{\xi}\delta_n}. \tag{A.28}$$

Step 2d.

First combine (A.8) (A.28) on the first two right side terms on (A.1), with probability at least $1-2/n$

$$\max_{1\le j\le p}\|\hat{B}(Y_{-j}\hat{B})^{+}u_j\|_2^2+\|\hat{B}(Y_{-j}\hat{B})^{+}Fa_j\|_2^2\le 2\gamma_e^2\sigma^2[\frac{2logn}{n\bar{\eta}}+\frac{\bar{r}}{n\bar{\eta}}]+\frac{3CK}{\bar{\xi}\delta_n}+\frac{3\bar{\Psi}K}{\bar{\eta}\bar{\xi}\delta_n}+\frac{3CK\bar{r}_n}{\bar{\eta}\bar{\xi}\delta_n^2}. \tag{A.29}$$

Step 3.

Note that by Lemma 1 the last term on the (A.1) is, via following (60) of Bunea et al. (2022), by $\bar{G}_j:=I_K+\bar{A}'_{-j}\Sigma_{U,-j}^{-1}\bar{A}_{-j}$ definition, where we also use $\bar{a}_j:=\Sigma_f^{1/2}a_j, \bar{A}_{-j}:=A_{-j}\Sigma_f^{1/2}$ definitions

$$\begin{aligned}\|\Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{a}_j\|_2^2&=\bar{a}'_j\bar{G}_j^{-1}\bar{A}'_{-j}\Sigma_{U,-j}^{-1}\Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{a}_j\\&\le\frac{\bar{a}'_j\bar{G}_j^{-1}\bar{A}'_{-j}\Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{a}_j}{Eigmin(\Sigma_{U,-j})}\\&=\frac{1}{Eigmin(\Sigma_{U,-j})}\bar{a}'_j\bar{G}_j^{-1}(\bar{G}_j-I_K)\bar{G}_j^{-1}\bar{a}_j,\end{aligned}$$

Then

$$\begin{aligned}\|\Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{a}_j\|_2^2&\le\frac{1}{Eigmin(\Sigma_{U,-j})}[\bar{a}'_j\bar{G}_j^{-1}\bar{a}_j-\bar{a}'_j\bar{G}_j^{-1}\bar{G}_j^{-1}\bar{a}_j]\\&\le\frac{1}{Eigmin(\Sigma_{U,-j})}\bar{a}'_j\bar{G}_j^{-1}\bar{a}_j.\end{aligned} \tag{A.30}$$

Next,

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

Define $H_j:=\bar{A}'_{-j}\Sigma_{U,-j}^{-1}\bar{A}_{-j}$, and H_j is invertible, with rank K

$$\begin{aligned}\bar{a}'_j(I_K+H_j)^{-1}\bar{a}_j&=\bar{a}'_jH_j^{-1/2}[I_K+H_j^{-1}]^{-1}H_j^{-1/2}\bar{a}_j\\&\le\frac{\bar{a}'_jH_j^{-1}\bar{a}_j}{Eigmin(I_K+H_j^{-1})}\le\frac{\bar{a}'_jH_j^{-1}\bar{a}_j}{1+Eigmin(H_j^{-1})}\end{aligned}$$

$$\leq \bar{a}_j' H_j^{-1} \bar{a}_j = \bar{a}_j' [\bar{A}_{-j}' \Sigma_{U,-j}^{-1} \bar{A}_{-j}]^{-1} \bar{a}_j, \tag{A.31}$$

where we use Rayleigh inequality for the first inequality, and for the last equality we use H_j definition. Next use $\bar{a}_j := \Sigma_f^{1/2} a_j, \bar{A}_{-j} := A_{-j}\Sigma_f^{1/2}$ definitions

$$\bar{a}_j' [\bar{A}_{-j}' \Sigma_{U,-j}^{-1} \bar{A}_{-j}]^{-1} \bar{a}_j = a_j' (A_{-j}' \Sigma_{U,-j}^{-1} A_{-j})^{-1} a_j.$$

Then using the inequality after proof of (63) in Bunea et al. (2022) at the end of Section A.2.3 of Bunea et al. (2022)

$$\bar{a}_j' [\bar{A}_{-j}' \Sigma_{U,-j}^{-1} \bar{A}_{-j}]^{-1} \bar{a}_j \leq \|\Sigma_{U,-j}\|_2 a_j' (A_{-j}' A_{-j})^{-1} a_j. \tag{A.32}$$

Combine (A.31)(A.32) in (A.30) to have, by (A.15)

$$\|\Sigma_{U,-j}^{-1} \bar{A}_{-j} \bar{G}_j^{-1} \bar{a}_j\|_2^2 \leq \left[\frac{\|\Sigma_{U,-j}\|_2}{Eigmin(\Sigma_{U,-j})}\right] \left[\frac{a_j' \Sigma_f a_j}{\lambda_K(A_{-j}\Sigma_f A_{-j}')}\right]. \tag{A.33}$$

Next,

$$\begin{aligned} \left[\frac{\|\Sigma_{U,-j}\|_2}{Eigmin(\Sigma_{U,-j})}\right] \left[\frac{a_j' \Sigma_f a_j}{\lambda_K(A_{-j}\Sigma_f A_{-j}')}\right] &\leq \left[\frac{\|\Sigma_{U,-j}\|_2}{Eigmin(\Sigma_{U,-j})}\right] \left[\frac{CK}{\xi_j \|\Sigma_{U,-j}\|_2}\right] \\ &= \frac{CK}{Eigmin(\Sigma_{U,-j})\xi_j} \leq \frac{CK}{\bar{\xi}}, \end{aligned} \tag{A.34}$$

where we use $\min_{1\leq j\leq p} Eigmin(\Sigma_{U,-j}) \geq c > 0$ Assumption 2(i) for the last inequality as well as $\bar{\xi}$ definition, and second equality in (A.16), with (A.17).

Now we combine Steps 1-3 to have with probability at least $1 - 2/n$, with definitions of $\max_{1\leq j\leq p} \hat{r}_j := \bar{r}$, $\min_{1\leq j\leq p} \hat{\eta}_j := \bar{\eta}$, and $\max_{1\leq j\leq p} \hat{\Psi}_j := \bar{\Psi}$,

$$\max_{1\leq j\leq p} \|\tilde{\alpha}_j - \alpha_j^*\|_2^2 \leq 2\gamma_e^2 \sigma^2 \left(\frac{2logn + \bar{r}}{n\bar{\eta}}\right) + \frac{3CK}{\bar{\xi}} \left(1 + \frac{1}{\delta_n} + \frac{\bar{\Psi}}{\bar{\eta}\delta_n} + \frac{r_n}{\bar{\eta}\delta_n}\right).$$

(ii). Asymptotics is trivial when $r_n/\delta_n \to 1$, and δ_n is nondecreasing in n.

Q.E.D.

Proof of Corollary 1. (i). We need to simplify three terms on the upper bound in Lemma 1 are $\bar{r}, \bar{\eta}, \bar{\Psi}$. Note that $\bar{r} := \max_{1\leq j\leq p} \hat{r}_j$. In that respect,

$$\hat{r}_j = rank(Y_{-j} P_{V_{1,k_j}}) = rank(Y_{-j} V_{1,k_j}) = k_j, \tag{A.35}$$

where the first equality is definition and the second equality is by the proof of Corollary 5 in Bing et al. (2021). Then the term

$$\hat{\eta}_j = \frac{1}{n}\sigma^2_{k_j}(Y_{-j}P_{U_{k_j}}) = \hat{\lambda}_{k_j}, \tag{A.36}$$

where the first equality is definition and the second equality is by the proof of Corollary 5 in Bing et al. (2021), and note that $\hat{\lambda}_{k_j} = \frac{1}{n}\sigma^2_{k_j}(Y_{-j})$ by definition. Then

$$\hat{\Psi}_j = \frac{\sigma^2_1(Y_{-j}(I_{p-1} - P_{U_{k_j}}))}{n} = \hat{\lambda}_{k_j+1}, \tag{A.37}$$

where the first equality is definition, and the second one is by the proof of Corollary 5 in Bing et al. (2021), and with $\hat{\lambda}_{k_j+1} = \frac{1}{n}\sigma^2_{k_j+1}(Y_{-j})$. We need to prove (20) to proceed further. We start by proving

$$P[\frac{1}{n}\sigma^2_{K+1}(Y_{-j}) < C_0\Delta_{-j}] \to 1. \tag{A.38}$$

In that respect, by using Weyl's inequality

$$\sigma_{K+1}(Y_{-j}) \leq \sigma_{K+1}(FA'_{-j}) + \sigma_1(U_{-j}) = \sigma_1(U_{-j}), \tag{A.39}$$

since $\sigma_{K+1}(FA'_{-j}) = 0$ by FA'_{-j} having rank of K. See that $\sigma^2_1(U_{-j})/n = \|U'_{-j}U_{-j}\|_2/n = \|U_{-j}U'_{-j}\|_2/n$ by singular value definition. We provide a definition of $\tilde{U}_{-j}$, which is an $n \times p-1$ matrix, U_{-j} also based on Assumption 4, with $\tilde{U}'_{-jt} : 1 \times p-1$ t th row of $\tilde{U}_{-j}$, and we stack these n rows to get $\tilde{U}_{-j}$ and then form U_{-j}

$$U_{-j} := \tilde{U}_{-j}\Sigma^{1/2}_{U,-j}.$$

Then by Lemma 22 of Bing et al. (2021) via Assumption 4, and definition (18)

$$P[\frac{1}{n}\|\tilde{U}_{-j}\Sigma_{U,-j}\tilde{U}'_{-j}\|_2 \leq C_0\Delta_{-j}] \geq 1 - exp(-n). \tag{A.40}$$

By (A.39)(A.40)

$$P[\frac{1}{n}\sigma^2_{K+1}(Y_{-j}) < C_0\Delta_{-j}] \geq 1 - exp(-n). \tag{A.41}$$

Note that defining the event $E_j := \{\sigma^2_{K+1}(Y_{-j})/n < C_0\Delta_{-j}\}$, and since the E_j^c is the complement of the event E_j

$$P(\cap_{j=1}^p E_j) = 1 - P((\cap_{j=1}^p E_j)^c) = 1 - P(\cup_{j=1}^p E_j^c) \geq 1 - \sum_{j=1}^p P(E_j^c)$$

Next, the union bound argument

$$P(\cup_{j=1}^p E_j^c) \leq \sum_{j=1}^p P(E_j^c) \leq exp(logp - n) \leq exp[(c' - 1)n], \tag{A.42}$$

given Assumption $logp \leq c'n$ with $c' < 1$ a positive constant upper bounded by one.

So this clearly shows

$$P(\cap_{j=1}^p E_j) \geq 1 - exp((c' - 1)n).$$

Then by (19) we have $\max_{1\leq j\leq p} \hat{s}_j < K + 1$ with probability approaching one, hence $\max_{1\leq j\leq p} \hat{s}_j \leq K$, with $\hat{s}_j$ being a positive integer.

So with probability at least $1 - exp((c' - 1)n$, $\max_{1\leq j\leq p} \hat{r}_j \leq K$, by $\hat{r}_j = k_j$ and definition of $\hat{s}_j$. Since $\hat{s}_{j+1} \leq K + 1$ wpa1, and by definition of E_j, and the analysis in (A.42), $k_{j+1} = \hat{s}_{j+1}$, we have with (A.37), and $\bar{\Psi} := \max_{1\leq j\leq p} \hat{\Psi}_j = \max_{1\leq j\leq p} \hat{\lambda}_{\hat{s}_j+1}$,

$$\bar{\Psi} \leq min_{1\leq j\leq p} C_0\Delta_{-j}, \tag{A.43}$$

with probability at least $1 - exp(c' - 1)n$. Next, by (A.36), $\hat{\eta}_j = \hat{\lambda}_{\hat{s}_j}$ with $k_j = \hat{s}_j$. Then

$$\bar{\eta} := \min_{1\leq j\leq p} \hat{\eta}_j = \min_{1\leq j\leq p} \hat{\lambda}_{\hat{s}_j} \geq \min_{1\leq j\leq p} \hat{\lambda}_K = \min_{1\leq j\leq p} \frac{\sigma^2_K(Y_{-j})}{n},$$

by $\hat{s}_j \leq K$ wpa1, and by definition $\hat{\lambda}_K$ represents the K th largest singular value (squared and then scaled by n). By the same analysis in (A.42)

$$\bar{\eta} \geq \min_{1\leq j\leq p} \hat{\lambda}_K := \min_{1\leq j\leq p} \sigma^2_K(Y_{-j})/n \geq \max_{1\leq j\leq p} C_0\Delta_{-j}, \tag{A.44}$$

$$\geq \max_{1\leq j\leq p} \|\Sigma_{U_{-j}}\|_2 = r_n, \tag{A.45}$$

where for the last inequality we use Δ_{-j} definition with $c_\Delta > 1, C_0 > 1$, and r_n definition. Also see that by (A.43)(A.44) $\bar{\Psi}/\bar{\eta} \leq 1$, and by (A.45) we have $\bar{\eta} \geq r_n$.

(ii). Note that asymptotic case is obtained by seeing δ_n is nondecreasing in n. **Q.E.D.**

Proof of Corollary 2. (i). Here we consider the case of $p-1<n$ in Generalized Least Squares Estimator (GLS) from now on. Basically in this case of $p-1<n$, this is OLS. We simplify the upper bound in Theorem 1 three terms, $\bar{r}, \bar{\Psi}, \bar{\eta}$. In that respect $\hat{B}=I_{p-1}$ by definition of GLS. Then clearly $\hat{r}_j = rank(Y_{-j}) = p-1$ since $p-1<n$ in this Corollary. Hence

$$\bar{r} = \max_{1\leq j\leq p} \hat{r}_j < p. \tag{A.46}$$

Also since $I_{p-1} - P_{I_{p-1}} = 0$ we have

$$\bar{\Psi} = 0. \tag{A.47}$$

This last equality shows that GLS has one less bias term in the upper bound compared to adaptive PCR estimator or compared to the general upper bound in Theorem 1. Use Corollary 11.6.6 in Bernstein (2018) (with $i=p-1, j=1$ there) or (11.6.18) of Bernstein (2018) for the first inequality below, $\Sigma_{-j,-j}$ is defined in Appendix B,

$$\begin{aligned}
\hat{\eta}_j &= \frac{1}{n}\sigma^2_{p-1}(Y_{-j}) = \frac{1}{n}\sigma^2_{p-1}(Y_{-j}\Sigma^{-1/2}_{-j,-j}\Sigma^{1/2}_{-j,-j}) \\
&\geq \frac{1}{n}\sigma^2_{p-1}(Y_{-j}\Sigma^{-1/2}_{-j,-j})\sigma^2_1(\Sigma^{1/2}_{-j,-j}) \\
&\geq \frac{1}{n}\sigma^2_{p-1}(Y_{-j}\Sigma^{-1/2}_{-j,-j})\sigma^2_{p-1}(\Sigma^{1/2}_{-j,-j}) \\
&= [\frac{1}{n}\sigma^2_{p-1}(Y_{-j}\Sigma^{-1/2}_{-j,-j})][Eigmin(\Sigma_{-j,-j})],
\end{aligned} \tag{A.48}$$

where the second inequality is obtained by maximum singular value of $\Sigma^{1/2}_{-j,-j}$ is larger than equal to minimum one, and the last equality is by the definition of singular value. Then clearly by $\Sigma_{-j,-j}, \Sigma_{U,-j}$ definitions, by using (5) and by Assumptions 2-3

$$min_{1\leq j\leq p} Eigmin(\Sigma_{-j,-j}) \geq min_{1\leq j\leq p} Eigmin(\Sigma_{U,-j}) \geq c.$$

So by (A.48) and inequality above by taking minimums

$$\bar{\eta} \geq c \min_{1\leq j\leq p}\left[\frac{1}{n}\sigma^2_{p-1}(Y_{-j}\Sigma^{-1/2}_{-j,-j})\right]. \tag{A.49}$$

To analyze (A.49), define $D_j := \{\sigma^2_{p-1}(Y_{-j}\Sigma^{-1/2}_{-j,-j}) \geq Cn\}$. Since $Y_{j-1} : n \times p-1$, and $n > p-1$ here, by Theorem 5.39 of Vershynin (2012) with Corollary 10 proof in Bing et al. (2021) via Assumption 4, and applying a union bound (with $t_1 = \sqrt{logn}$)and by assuming $plogn \leq c'n$, we have, with c_1 a positive constant, and $c' \in (0,1)$ are positive constants, and also assuming $log2p \leq (c_1/2)logn$

$$P(\cup_{j=1}^p D_j^c) \leq 2pexp(-c_1 logn) = exp(log2p - c_1 logn)$$
$$\leq 1/n^{c_1/2}. \tag{A.50}$$

Then since $P(\cap_{j=1}^p D_j) = 1 - P(\cup_{j=1}^p D_j^c)$ we have

$$\bar{\eta} \geq C, \tag{A.51}$$

with at least probability $1 - n^{c_1/2}$. Now substitute (A.46)(A.47)(A.51) in the upper bound in Theorem 1 to have the result.

(ii). Asymptotics is derived by using $r_n/\delta_n \to 1$.

Q.E.D

Proof of Corollary 3. (i). There are again three terms that we consider, but we have $p-1 > n$ in this GLS case. So

$$\bar{r} = \max_{1 \leq j \leq p} rank(Y_{-j}) = n. \tag{A.52}$$

Then we have $\bar{\Psi} = 0$ as in Corollary 2 due to $\hat{B} = I_{p-1}$. The last term is $\hat{\eta}_j = \frac{1}{n}\sigma^2_n(Y_{-j})$, which is the n th largest singular value of $Y_{-j} : n \times p-1$ matrix. Set $E_j := \{\sigma^2_n(Y_{-j}) \geq ctr(\Sigma_{U,-j})\}$. Also we assume $\min_{1 \leq j \leq p}[tr(\Sigma_{U,-j})/\|\Sigma_{U,-j}\|_2] > Cn$, and $\tilde{U}_{-j}$ has independent entries so p.42 of proof of Proposition 14 of Bunea et al. (2022) but replacing Theorem 4.6.1 of Vershynin (2019) in that proof with Theorem 5.3.9 of Vershynin (2012) to take into account nonzero mean of factors, and taking a union bound

$$P(\cap_{j=1}^p E_j) \geq 1 - exp((-c_2/2)n),$$

where we need also $log2p \leq (c_2/2)n$ for c_2 positive constant. Clearly, with probability exceeding $1 - exp((-c_2/2)n)$, uniformly in j

$$\bar{\eta} \geq Cmintr(\Sigma_{U,-j})/n \geq C \max \|\Sigma_{U,-j}\|_2 \geq Cr_n. \tag{A.53}$$

The desired result follows by incorporating these three terms in the upper bound of Theorem 1.

(ii). Asymptotics follows thru $r_n/\delta_n \to 1$.

Q.E.D.

Before the following lemma, we proceed with three inequalities that are used for rectangular matrices. Take a generic rectangular matrix M and a vector x.

$$\|Mx\|_2 \leq \|M\|_2 \|x\|_2, \tag{A.54}$$

by (11.4.4) of Bernstein (2018). Then by (11.3.15) of Bernstein (2018) where for two generic rectangular matrices M, L

$$\|ML\|_2 \leq \|M\|_2 \|L\|_2, \tag{A.55}$$

We form the following matrix norm inequality for a generic rectangular matrix $M : m_1 \times m_2$

$$\begin{aligned} \|M\|_2 \leq \|M\|_F &= [\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} |M_{ij}|^2]^{1/2} \\ &\leq \sqrt{m_1 m_2} \max_{1 \leq i \leq m_1} \max_{1 \leq j \leq m_2} |M_{ij}| = \sqrt{m_1 m_2} \|M\|_\infty. \end{aligned} \tag{A.56}$$

where $\|.\|_F$ is the Frobenius norm and we get the first inequality by (4.4) of Vershynin (2019).

We have the following lemma for the reciprocal of the main j th diagonal term in the precision matrix of asset returns. This is a new proof. Also, technique in Bing et al. (2021) does only consider simple linear regression, and analysis of precision matrix estimation and especially a diagonal term estimation is not handled, so this proof, Lemma A.1 extends the literature to a different issue of diagonal term estimation. The results are new and shows the rate of convergence and what are the contributing terms, as well as consistency of the estimator. Note that r_{w1}, r_{w2} rates are defined just before Assumption 5 in the main text.

LEMMA A.1: *(i). Under Assumptions 1-4 5(i), with* $logp \leq Cn$*, with* C *a positive constant, and* $r_n/\delta_n \to 1$*,*

$$\max_{1\leq j\leq p} |\tilde{\tau}_j^2 - \tau_j^2| = O_p\left(K(\sqrt{\frac{p(logp)}{n}})r_{w1}\right) + O_p\left(d_{2n}r_{w2}\right) + O_p\left(max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\max(\sqrt{logp}, \sqrt{logn})}{n^{1/2}})\right).$$

(ii). Under Assumptions 1-5, with $r_n/\delta_n \to 1$*,*

$$\max_{1\leq j\leq p} |\tilde{\tau}_j^2 - \tau_j^2| = o_p(1).$$

(iii). Under Assumptions 1-5, with $r_n/\delta_n \to 1$*, where* $c > 0$ *is a positive constant*

$$\min_{1\leq j\leq p} \tau_j^2 \geq c - o(1).$$

Remarks. 1. We show that estimation of the reciprocal of the main diagonal terms in precision matrix estimation depends on number of factors, K, sample size-time span n, number of assets p, and signal-to noise ratio $\bar{\xi}$ as well as three terms $\bar{r}, \bar{\Psi}, \bar{\eta}$ which represents complexity, prediction errors, and prediction performance respectively. These last three terms are random and will be upper bounded with non-random terms, wpa1, in subcases of GLS and adaptive PCR below in Corollaries A.1-A.3.

2. When $p > n$ and $p/n \to (1, \infty)$ it is possible to get consistency since p only enters the estimation errors directly via $\sqrt{logp}$ and with a larger n, the estimation errors will go to zero, if $K\sqrt{logp}\, r_{w1} = o_p(1)$.

3. Number of factors K play a key role, to have Assumption 5(ii), we need $K = o(n^{1/3})$, and there are more restrictions on K in Assumptions 5(i), (iii). Larger number of factors provide a real roadblock in getting consistency.

4. Most important is the fact that estimation of the reciprocal of the main diagonal terms in the precision matrix. This is more difficult compared with estimating each row of the precision matrix (without the main diagonal term). Lemma 1(ii) clearly shows that the rate of estimation error for off-diagonal-row elements is r_{w2}, but our Lemma A.1 shows for the reciprocal of the main diagonal term, this is more complicated and slower than r_{w2} since

via Assumption 5(iii), we need $d_{2n}r_{w2} = o_p(1)$ for consistency in addition to other part of Assumption 5(iii) as well as Assumption 5(ii).

Proof of Lemma A.1.

First note that $Ey_{jt} = \mu_j = a_j' Ef_t$, and $EY_{-jt} = \mu_{-j} = A_{-j}Ef_t$. Also see that

$$y_{jt} - \mu_j = a_j'(f_t - Ef_t) + u_{jt}, \tag{A.57}$$

and

$$Y_{-jt} - \mu_{-j} = A_{-j}(f_t - Ef_t) + U_{-jt}, \tag{A.58}$$

(i). Note that we need to estimate the j th main diagonal term for the precision matrix of asset returns. In that sense the key element is the reciprocal of that term which is defined in (12), and the population version in (B.10). We use triangle inequality to consider the estimation error by (12)(B.10)

$$|\tilde{\tau}_j^2 - \tau_j^2| \leq |\frac{1}{n}\sum_{t=1}^{n}[y_{jt}^2 - E(y_{jt}-\mu_j)^2]| + \left|[n^{-1}\sum_{t=1}^{n} y_{jt}Y_{-jt}']\tilde{\alpha}_j - E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'\alpha_j^*\right|. \tag{A.59}$$

Next add and subtract to the second right side term $E[(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\tilde{\alpha}_j$, by iid nature of data,

$$\begin{aligned}|\tilde{\tau}_j^2 - \tau_j^2| \leq |\frac{1}{n}\sum_{t=1}^{n}[y_{jt}^2 - E(y_{jt}-\mu_j)^2]| + \left|n^{-1}\sum_{t=1}^{n}[y_{jt}Y_{-jt}' - E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'\tilde{\alpha}_j\right| \\ + \left|E[(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'](\tilde{\alpha}_j - \alpha_j^*)\right|. \end{aligned} \tag{A.60}$$

Take uniform bounds over $j = 1, \cdots, p$ and use Cauchy-Schwartz inequality for the second and third right side terms on (A.60)

$$\begin{aligned}\max_{1\leq j\leq p}|\tilde{\tau}_j^2 - \tau_j^2| \leq \max_{1\leq j\leq p}|\frac{1}{n}\sum_{t=1}^{n}[y_{jt}^2 - E(y_{jt}-\mu_j)^2]| \\ + \max_{1\leq j\leq p}\|n^{-1}\sum_{t=1}^{n}[y_{jt}Y_{-jt}' - E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\|_2 \max_{1\leq j\leq p}\|\hat{\alpha}_j\|_2 \\ + \max_{1\leq j\leq p}\|E[(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\|_2 \max_{1\leq j\leq p}\|\hat{\alpha}_j - \alpha_j^*\|_2. \end{aligned} \tag{A.61}$$

Each right side term in (A.61) will be analyzed as a step in the proof.

Step 1.

Start with the first right side term on (A.61), and by (2)

$$y_{jt}^2 = a_j' f_t f_t' a_j + 2a_j' f_t u_{jt} + u_{jt}^2,$$

and by Assumption 1(iv), and using (A.57)

$$E(y_{jt} - \mu_j)^2 = a_j' \Sigma_f a_j + E(u_{jt}^2).$$

Use triangle inequality by the last two equalities, and noting $\Sigma_f := E(f_t - Ef_t)(f_t - Ef_t)'$

$$\begin{aligned} \max_{1\le j\le p} |\frac{1}{n}\sum_{t=1}^n [y_{jt}^2 - E(y_{jt}-\mu_j)^2]| \le & \max_{1\le j\le p} |a_j'[\frac{1}{n}\sum_{t=1}^n (f_t f_t' - \Sigma_f)]a_j| \\ & + \max_{1\le j\le p} |\frac{1}{n}\sum_{t=1}^n [u_{jt}^2 - E(u_{jt}^2)]| \\ & + \max_{1\le j\le p} 2|a_j'[n^{-1}\sum_{t=1}^n f_t u_{jt}]|. \end{aligned} \tag{A.62}$$

Take the first right side term in (A.62) above, and use Cauchy-Schwartz inequality first and then (A.54)

$$\max_{1\le j\le p} |a_j'[\frac{1}{n}\sum_{t=1}^n (f_t f_t' - \Sigma_f)]a_j| \le \left[\max_{1\le j\le p} \|a_j\|_2^2\right] \|\frac{1}{n}\sum_{t=1}^n [f_t f_t' - \Sigma_f]\|_2. \tag{A.63}$$

We consider the second term on the right side of (A.63), with $F : n \times K$ matrix, and $\|\Sigma_f\|_2 \le C < \infty$ by Assumption 3

$$\|\frac{F'F}{n} - \Sigma_f\|_2 \le C[\sqrt{\frac{K}{n}} + \sqrt{\frac{logn}{n}}]\|\Sigma_f\|_2, \tag{A.64}$$

by Lemma 23 of Bing et al. (2021) since rows of F are independent subgaussian vectors, with probability at least $1 - 2/n^c$ via Assumption 1(iii), Assumption 4(ii), and note that $K < n$.

By Assumption 4(iii)

$$\max_{1\le j\le p} \|a_j\|_2^2 \le CK. \tag{A.65}$$

Use (A.64)-(A.65) in (A.63) and Assumption 3 to have

$$\max_{1\leq j\leq p}|a_j'(\frac{1}{n}\sum_{t=1}^{n}[f_tf_t'-\Sigma_f]a_j|=O_p\left(\max(\frac{K^{3/2}}{n^{1/2}},K\sqrt{\frac{logn}{n}})\right). \tag{A.66}$$

Then in (A.62) consider the second right side term

$$\max_{1\leq j\leq p}|\frac{1}{n}\sum_{t=1}^{n}[u_{jt}^2-E(u_{jt}^2)]|=O_p(\sqrt{\frac{logp}{n}}). \tag{A.67}$$

We now show how we obtain (A.67). Note that by Corollary 2.8.3 of Vershynin (2019), for $t_1>0$, and the union bound

$$P\left[\max_{1\leq j_1\leq p}\max_{1\leq j_2\leq p}|\frac{1}{n}\sum_{t=1}^{n}[u_{j_1,t}u_{j_2,t}-E(u_{j_1,t}u_{j_2,t})]|>t_1\right]\leq p^2exp\left(-cmin(\frac{t_1^2}{C_\psi^2},\frac{t_1}{C_\psi})n\right),$$

where we use $u_{j_1,t}$ and $u_{j_2,t}$ are both subgaussian random variables, so $u_{j_1,t}u_{j_2,t}$ are subexponential random variable (independent across $t=1,\cdots,n$) by Lemma 2.7.7 of Vershynin (2019) and exercise 2.7.10 of Vershynin (2019) shows centered subexponential are also subexponential random variables, and we have

$$\max_{1\leq j_1\leq p}\max_{1\leq j_2\leq p}\|u_{j_1,t}u_{j_2,t}\|_\psi\leq C_\psi<\infty$$

where C_ψ are the upper bound constant for the subexponential norm in Definition 2.7.5 of Vershynin (2019). Next

$$\begin{aligned}2p^2exp\left(-cmin(\frac{t_1}{C_\psi^2},\frac{t_1}{C_\psi})n\right)&=exp\left(log(2p^2)-cmin(\frac{t_1^2}{C_\psi^2},\frac{t_1}{C_\psi})n\right)\\&=exp\left(log(2p^2)-c\frac{t_1^2n}{C_\psi^2}\right),\end{aligned}$$

with $t_1:=C\frac{\sqrt{logp}}{\sqrt{n}}$ and with sufficiently large n (i.e.$\sqrt{n}\geq\frac{C}{C_\psi}(logp)^{1/2}$). Set

$$c_\psi:=c\frac{C^2}{C_\psi^2}-2>0$$

to have

$$exp[log2 + 2logp - c\frac{C^2}{C_\psi^2}logp] = \frac{2}{p^{c_\psi}}.$$

Then

$$P\left(\max_{1\le j_1\le p}\max_{1\le j_2\le p}|\frac{1}{n}\sum_{t=1}^{n}u_{j_1,t}u_{j_2,t} - Eu_{j_1,t}u_{j_2,t}| > C\frac{\sqrt{logp}}{\sqrt{n}}\right) \le \frac{2}{p^{c_\psi}}.$$

The asymptotics can be obtained since when $n \to \infty, p \to \infty$ and we obtain (A.67).

Consider the third right side term in (A.62). By Lemma 2.7.7 and Corollary 2.8.3 of Vershynin (2019) and Assumptions 1, 4 by Holder's inequality

$$\begin{aligned}\max_{1\le j\le p}|a_j'(\frac{1}{n}\sum_{t=1}^{n}f_tu_{jt})| &\le \max_{1\le j\le p}\|a_j\|_1\max_{1\le j\le p}\|\frac{1}{n}\sum_{t=1}^{n}f_tu_{jt}\|_\infty \\ &= O_p(K\sqrt{\frac{log(Kp)}{n}}) = O_p(\frac{K\sqrt{logp}}{\sqrt{n}}), \qquad \text{(A.68)}\end{aligned}$$

where the last equality uses $p - 1 > K$. By (A.66)-(A.68) we have the first right side term in (A.61)

$$\max_{1\le j\le p}|\frac{1}{n}\sum_{t=1}^{n}[y_{jt}^2 - E(y_{jt}-\mu_j)^2]| = O_p\left(\max(\frac{K^{3/2}}{n^{1/2}}, K\sqrt{\frac{\max(logp, logn)}{n}})\right). \quad \text{(A.69)}$$

Step 2.

We consider the second right side term in (A.61) and specifically by (2)(3) and Assumption 1(iv)

$$y_{jt}Y_{-jt}' = a_j'f_tf_t'A_{-j}' + a_j'f_tU_{-jt}' + u_{jt}f_t'A_{-j}' + u_{jt}U_{-jt}',$$

and by (A.57) (A.58)

$$E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})' = a_j'\Sigma_fA_{-j}'.$$

Use these last two equations

$$\max_{1\le j\le p}\|\frac{1}{n}\sum_{t=1}^{n}[y_{jt}Y_{-jt}' - E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\|_2 \le \max_{1\le j\le p}\|a_j'\left(\frac{1}{n}\sum_{t=1}^{n}[f_tf_t' - \Sigma_f]\right)\hat{A}_{-j}'\|_2$$

$$+ \max_{1\le j\le p} \|(\frac{1}{n}\sum_{t=1}^{n} u_{jt}f_t')A'_{-j}\|_2 + \max_{1\le j\le p} \|a_j'(\frac{1}{n}\sum_{t=1}^{n} f_t U'_{-jt})\|_2 + \max_{1\le j\le p} \|\frac{1}{n}\sum_{t=1}^{n} u_{jt}U'_{-jt}\|_2. \quad \text{(A.70)}$$

Consider the first term on the right side of (A.70) by (A.54)(A.55)

$$\max_{1\le j\le p} \|a_j'(\frac{1}{n}\sum_{t=1}^{n}[f_t f_t' - \Sigma_f]A'_{-j}\|_2 \le \max_{1\le j\le p} \|a_j\|_2 \max_{1\le j\le p} \left\| \left(\frac{1}{n}\sum_{t=1}^{n}[f_t f_t' - \Sigma_f] \right) A'_{-j} \right\|_2$$

$$\le (\max_{1\le j\le p} \|a_j\|_2)\|\frac{1}{n}\sum_{t=1}^{n}[f_t f_t' - \Sigma_f]\|_2 \times (\max_{1\le j\le p} \|A'_{-j}\|_2)$$

$$= O(K^{1/2})O_p(\max(\frac{K^{1/2}}{n^{1/2}}, \frac{\sqrt{logn}}{n^{1/2}}))O(d_{1n}) = O_p(max(\frac{Kd_{1n}}{n^{1/2}}, \frac{d_{1n}\sqrt{Klogn}}{n^{1/2}})), \quad \text{(A.71)}$$

and the rates are by (A.64)(A.65) and Assumption 3.

Then consider the second term on the right side of (A.70)

$$\begin{aligned}\max_{1\le j\le p} \|(\frac{1}{n}\sum_{t=1}^{n} u_{jt}f_t')A'_{-j}\|_2 &\le \max_{1\le j\le p} \|(\frac{1}{n}\sum_{t=1}^{n} u_{jt}f_t')\|_2 \max_{1\le j\le p} \|A'_{-j}\|_2 \\ &\le \sqrt{K} \max_{1\le j\le p} \|(\frac{1}{n}\sum_{t=1}^{n} u_{jt}f_t')\|_\infty \max_{1\le j\le p} \|A'_{-j}\|_2 \\ &= O_p(\sqrt{K}\frac{\sqrt{logp}}{\sqrt{n}})O(d_{1n}), \quad \text{(A.72)}\end{aligned}$$

where we use (A.54) for the first inequality since $u_{jt}f_t' : 1 \times K$ vector and $A'_{-j} : K \times p-1$ matrix, and the second inequality is by simple vector norm inequality tying l_2, l_∞ vector norms, and the rates are by Assumption 3(iii) $\max_{1\le j\le p} \|A'_{-j}\|_2 = O(d_{1n})$ and Lemma 2.7.7 and Corollary 2.8.3. of Vershynin (2019) provides $\|n^{-1}\sum_{t=1}^{n} u_{jt}f_t'\|_\infty = O_p(\sqrt{logp/n})$. Now third term on the right side of (A.70)

$$\begin{aligned}\max_{1\le j\le p} \|a_j'(\frac{1}{n}\sum_{t=1}^{n} f_t U'_{-jt})\|_2 &\le \max_{1\le j\le p} \|a_j\|_2 \max_{1\le j\le p} \|\frac{1}{n}\sum_{t=1}^{n} f_t U'_{-jt}\|_2 \\ &\le \max_{1\le j\le p} \|a_j\|_2 \sqrt{Kp} \max_{1\le j\le p} \|\frac{1}{n}\sum_{t=1}^{n} f_t U'_{-jt}\|_\infty\end{aligned}$$

$$= O(\sqrt{K})O_p(\frac{\sqrt{Kp(logp)}}{\sqrt{n}}) = O_p(K\sqrt{p}\frac{\sqrt{logp}}{\sqrt{n}}), \quad (A.73)$$

where we use for the first inequality (A.54) and then for the second inequality we use (A.56) and for the rates we use Assumption 5(iii), and Lemma 2.7.7 and Corollary 2.8.3 of Vershynin (2019) with Assumption 4 since $f_t U'_{-jt}$ is $K \times p-1$ matrix. Then for the fourth term on the right side of (A.70) via Corollary 2.8.3 of Vershynin (2019) and Assumption 4 and simple vector norm inequality

$$\max_{1\le j\le p} \|\frac{1}{n}\sum_{t=1}^{n} u_{jt}U'_{-jt}\|_2 \le \sqrt{p} \max_{1\le j\le p} \|\frac{1}{n}\sum_{t=1}^{n} u_{jt}U'_{-jt}\|_\infty = O_p(\sqrt{p}\frac{\sqrt{logp}}{\sqrt{n}}). \quad (A.74)$$

By (A.71)(A.72)(A.73)(A.74) and the slowest rate is by (A.73) by Assumption 5(i), and we have the following on the second right side term in (A.61)

$$\max_{1\le j\le p} \|\frac{1}{n}\sum_{t=1}^{n}[y_{jt}Y'_{-jt} - E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\|_2 = O_p(K\sqrt{p}\frac{\sqrt{logp}}{\sqrt{n}}). \quad (A.75)$$

First we remind the rate definition:

$$r_{w1} := \max\left(\sqrt{\frac{logn+\bar{r}}{n\bar{\eta}}}, \sqrt{\frac{K}{\bar{\xi}\delta_n}}, \sqrt{\frac{K\bar{\Psi}}{\bar{\xi}\bar{\eta}\delta_n}}, \sqrt{\frac{K}{\bar{\xi}\bar{\eta}}}\right), \quad (A.76)$$

Next by the proof of Theorem 1, Step 2d we obtain, with $r_n/\delta_n \to 1$

$$\max_{1\le j\le p} \|\tilde{\alpha}_j\|_2 = O_p(r_{w1}). \quad (A.77)$$

Combine (A.75)(A.77) for the second right side term in (A.61)

$$\max_{1\le j\le p} \|\frac{1}{n}\sum_{t=1}^{n}[y_{jt}Y'_{-jt} - E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\|_2 \max_{1\le j\le p} \|\tilde{\alpha}_j\|_2$$
$$= O_p((K\sqrt{plogp/n})r_{w1}) \quad (A.78)$$

Step 3.

Now consider the third term on the right side of (A.61). The key term is

$$E[(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'] = a_j\Sigma_f A'_{-j},$$

where we use Assumption 1 and (2)(3). Since we have for rectangular matrices (A.54)(A.55)

$$\max_{1\le j\le p}\|E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'\|_2=\max_{1\le j\le p}\|a_j'\Sigma_f A_{-j}'\|_2=d_{2n}, \tag{A.79}$$

By (A.79), the third term on the right side of (A.61) is by Theorem 1

$$\max_{1\le j\le p}\|E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'\|_2\max_{1\le j\le p}\|\hat\alpha_j-\alpha_j^*\|_2=O_p(d_{2n}r_{w2}), \tag{A.80}$$

with $r_{w2}:=\max(r_{w1},\sqrt{\frac{K}{\bar\xi}})$.

Step 4.

We combine (A.69)(A.78)(A.80) with Assumption 5(i) on the right side of (A.61) to have the desired result.

Q.E.D.

(ii). By Assumption 5(ii)-(iii) we obtain the consistency. **Q.E.D.**

(iii). By (B.10)

$$\tau_j^2:=E[(y_{jt}-\mu_j)^2-E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\alpha_j^*.$$

Then by Cauchy-Schwartz inequality

$$\begin{aligned}\max_{1\le j\le p}E[(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\alpha_j^*&\le\left[\max_{1\le j\le p}\|E[(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})']\|_2\right]\left[\max_{1\le j\le p}\|\alpha_j^*\|_2\right]\\&\le\frac{CK^{1/2}d_{2n}}{\sqrt{\bar\xi}}\to 0,\end{aligned} \tag{A.81}$$

where we use Step 3 of Lemma A.1 together with Step 3 of Theorem 1 proof (A.34), and (A.79), and Assumption 5(iii). Next by using (2), Assumptions 1-3

$$\begin{aligned}\min_{1\le j\le p}E(y_{jt}-\mu_j)^2&\ge min_{1\le j\le p}a_j'\Sigma_f a_j+\min_{1\le j\le p}E(u_{jt}^2)\\&\ge Eigmin(\Sigma_f)\min_{1\le j\le p}\|a_j\|_2^2+\min_{1\le j\le p}\sigma_j^2\ge c>0.\end{aligned}$$

The last two results provide the desired result. **Q.E.D.**

Proof of Theorem 2. First of all

$$\max_{1\le j\le p}\|\tilde\Theta_j'-\Theta_j'\|_2\le\max_{1\le j\le p}\left|\frac{1}{\tilde\tau_j^2}-\frac{1}{\tau_j^2}\right|+\max_{1\le j\le p}\|\frac{\tilde\alpha_j}{\tilde\tau_j^2}-\frac{\alpha_j^*}{\tilde\tau_j^2}\|_2.$$

Then add and subtract $\alpha_j^*/\tilde{\tau}_j^2$ to the right hand side of the immediately above inequality, and via triangle inequality and arranging

$$\max_{1\leq j\leq p}\|\tilde{\Theta}'_j-\Theta'_j\|_2\leq \max_{1\leq j\leq p}\left|\frac{1}{\tilde{\tau}_j^2}-\frac{1}{\tau_j^2}\right|+\max_{1\leq j\leq p}\|\tilde{\alpha}_j-\alpha_j^*\|_2\max_{1\leq j\leq p}[\frac{1}{\tau_j^2}]$$
$$+\max_{1\leq j\leq p}\|\alpha_j^*\|_2\max_{1\leq j\leq p}\left|\frac{1}{\tilde{\tau}_j^2}-\frac{1}{\tau_j^2}\right|. \tag{A.82}$$

Note that since $\min_{1\leq j\leq p}\tilde{\tau}_j^2\geq\min_{1\leq j\leq p}\tau_j^2-\max_{1\leq j\leq p}|\tilde{\tau}_j^2-\tau_j^2|$

$$\max_{1\leq j\leq p}\left(\frac{1}{\tilde{\tau}_j^2}\right)\leq\frac{1}{\min_{1\leq j\leq p}\tau_j^2-\max_{1\leq j\leq p}|\tilde{\tau}_j^2-\tau_j^2|}$$
$$\leq\frac{1}{c-o_p(1)},$$

by Lemma A.1(ii)(iii) under Assumptions 1,5. Then by (A.34)

$$\max_{1\leq j\leq p}\|\alpha_j^*\|_2=O(\frac{\sqrt{K}}{\sqrt{\bar{\xi}}})=o(1), \tag{A.83}$$

by Assumption 5(iii). So the third term on right side of (A.82) converges to zero faster in probability than the first right side term in (A.82) given Lemma A.1. Then Remark 4 of Lemma A.1 shows that the first term on the right side of (A.82) converges to zero slower in probability than the second term on the same right side. So the main rate comes from the first term, which is the main diagonal term rate as shown in Lemma A.1(i). **Q.E.D**

Note that we provide the specific case of adaptive PCR estimator of τ_j^2, which is defined as $\hat{\tau}_{j,PCR}^2$. In that respect the rates r_{w1}, r_{w2} simplify considerably. In adaptive PCR case these rates are with wpa1

$$r_{w1,PCR}=\max\left(\sqrt{\frac{logn+K}{nr_n}},\sqrt{\frac{K}{\bar{\xi}\delta_n}}\right), \tag{A.84}$$

and

$$r_{w2,PCR} = \max\left(\sqrt{\frac{logn+K}{nr_n}}, \sqrt{\frac{K}{\bar{\xi}}}\right). \tag{A.85}$$

See that rates (A.84)(A.85) can be obtained by seeing $\bar{r} \leq K, \bar{\Psi}/\bar{\eta} \leq 1, \bar{\eta} \geq r_n$ wpa1 as shown in Corollary 1. Corollary A.1 below follows through replacing r_{w1}, r_{w2} in Lemma A.1 with $r_{w1,PCR}, r_{w2,PCR}$ respectively.

Corollary A.1. (i). *Under Assumptions 1-5(i)and using* $r_{w1,PCR}, r_{w2,PCR}$ *instead of* r_{w1}, r_{w2} *respectively, and* $r_n/\delta_n \to 1$,

$$\max_{1\leq j\leq p} |\hat{\tau}^2_{PCR,j} - \tau_j^2| = O_p\left(K(\sqrt{plogp/n})r_{w1,PCR}\right) + O_p\left(d_{2n}r_{w2,PCR}\right) + O_p\left(max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\max(\sqrt{logp}, \sqrt{logn})}{n^{1/2}})\right).$$

(ii). Under Assumptions 1-5, $r_n/\delta_n \to 1$,

$$\max_{1\leq j\leq p} |\hat{\tau}^2_{PCR,j} - \tau_j^2| = o_p(1).$$

(iii). Under Assumptions 1-5, $r_n/\delta_n \to 1$,

$$\min_{1\leq j\leq p} \tau_j^2 \geq c - o_p(1).$$

Here we consider the case of $p-1 < n$ in GLS estimator of τ_j^2. We define them as $\hat{\tau}^2_{GLS,j}$. We define the following rate which will be used instead of r_{w1}, r_{w2}, wpa1, which can be deduced from Corollary 2 proof,

$$r_{GLS-l} := \max(\sqrt{\frac{logn+p}{n}}, \sqrt{\frac{K}{\bar{\xi}}}). \tag{A.86}$$

The proof is a simple application of Lemma A.1.

Corollary A.2. *(i).Under Assumptions 1-5(i) and using* r_{GLS-l} *instead of both* r_{w1}, r_{w2}, *setting* $p/n = a_{2n} \to 0$, *and* $r_n/\delta_n \to 1$

$$\max_{1\leq j\leq p}|\hat{\tau}^2_{GLS,j}-\tau^2_j| = O_p\left(K(\sqrt{a_{2n}logp})r_{GLS-l}\right) + O_p\left(d_{2n}r_{GLS-l}\right) + O_p\left(max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\max(\sqrt{logp},\sqrt{logn})}{n^{1/2}})\right)$$

(ii). Under Assumptions 1-5 with the conditions in (i)

$$\max_{1\leq j\leq p}|\hat{\tau}^2_{GLS,j}-\tau^2_j| = o_p(1).$$

(iii). Under Assumptions 1-5 with the conditions in (i)

$$\min_{1\leq j\leq p}\tau^2_j \geq c - o_p(1).$$

We now consider the case of GLS estimator when $p-1>n$. Note that this subcase will be different than the others, and we use an estimator of τ^2_j which is different due to interpolation property of GLS in this case. First let us show the interpolation and how it affects the estimator $\tilde{\tau}^2_j$ proposed, and the need to modify. In case of $p-1>n$, GLS estimator becomes the l_2 norm interpolator. Define the minimum norm interpolator as in (1) of Bunea et al. (2022), for each $j=1,\cdots,p$

$$\hat{\alpha}_j := argmin\{\|\alpha_j\|_2 : \|Y_{-j}\alpha_j - y_j\|_2 = \min_{\gamma_j}\|Y_{-j}\gamma_j - y_j\|_2.\}$$

Lemma 28 of Bunea et al. (2022) shows

$$\hat{\alpha}_j = Y^+_{-j}y_j = \hat{\alpha}_{GLS,j},$$

where $Y^+_{-j}: p-1\times n$ is the Moore-Penrose inverse of Y_{-j}. Then when $p-1>n$

$$Y_{-j}\hat{\alpha}_j - y_j = Y_{-j}Y^+_{-j}y_j - y_j = y_j - y_j = 0,$$

with $Y_{-j}Y^+_{-j}=I_n$ by (136) of Bunea et al. (2022). This is interpolation, and we cannot use the formula for $\tilde{\tau}^2_j$ since with interpolation it can be shown, $\tilde{\tau}^2_j=0$ in GLS in (12), with $p-1>n$ case. So we propose the following estimator, which can be considered a subcase

of general $\tilde{\tau}_j^2$. Define

$$\hat{\tau}_{INT,j}^2 := \frac{y_j' y_j}{n}. \tag{A.87}$$

Corollary A.3. *Under Assumptions 1-4,*

(i).

$$\max_{1\leq j\leq p} |\hat{\tau}_{INT,j}^2 - \tau_j^2| = O_p\left(\frac{\sqrt{K}d_{2n}}{\sqrt{\bar{\xi}}}\right) + O_p\left(max(\frac{K^{3/2}}{n^{1/2}}, K\frac{\sqrt{logp}}{n^{1/2}})\right).$$

(ii). Under Assumptions 1-4, Assumption 5(ii) with $\sqrt{K}d_{2n}/\sqrt{\bar{\xi}} \to 0$

$$\max_{1\leq j\leq p} |\hat{\tau}_{INT,j}^2 - \tau_j^2| = o_p(1).$$

(iii).

$$\min_{1\leq j\leq p} \tau_j^2 \geq c - o_p(1).$$

Remarks. 1. Compared with our cases of interest, we can use a weaker version of Assumption 5(iii), $\sqrt{K}d_{2n}/\sqrt{\bar{\xi}} \to 0$ to get consistency.

2. A second question concerns what happens to consistency if $p \to \infty$ and $p/n \to (1, \infty)$. Given Assumption $K\sqrt{logp/n} \to 0$ the only point p enters is logarithmically (natural logarithm), increasing p will have a minor negative effect but still we get consistency. p can effect the signal to noise ratio but its not clear how the tradeoff may happen, an example is given in (8).

3. It will not be a good idea to compare this estimator, $\hat{\tau}_{INT,j}^2$ with adaptive PCR based one, since its structure is different. Also as discussed above Corollary A.3, since there is interpolation, the general estimator $\tilde{\tau}_j^2 = 0$ in this case, hence cannot invert the general estimator for the precision matrix in GLS with $p-1 > n$ case, since $1/\tilde{\tau}_j^2$, $j = 1, \cdots, p$, are the main diagonal terms in the precision matrix estimation. We see the differences in our simulation design and empirics more clearly.

Proof of Corollary A.3. (i). By (B.10)(A.87) and triangle inequality

$$\max_{1\leq j\leq p} |\hat{\tau}_{INT,j}^2 - \tau_j^2| \leq \max_{1\leq j\leq p} |\frac{1}{n}\sum_{t=1}^{n}[y_{jt}^2 - E(y_{jt} - \mu_j)^2]|$$

$$+ \max_{1\leq j\leq p} \|E(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'\|_2 \max_{1\leq j\leq p} \|\alpha_j^*\|_2.$$

Next, Step 1 in the proof of Lemma A.1 with (A.69), and (A.81) provides the result.

(ii). Given Assumptions in part (i) with $\sqrt{K}d_{2n}/\sqrt{\bar{\xi}} \to 0$ provides the consistency.

(iii). This is the same as in the proof of Lemma A.1 (iii). **Q.E.D.**

APPENDIX B

In this part, we provide a precision matrix formula. Different versions with time series, and for second moments of the regressor matrix can also be found in Callot et al. (2021), Caner and Kock (2018). Its relation to a hidden factor model context is established here for the first time in the literature and in detail.

We follow matrix algebra and establish a formula for each row of the precision matrix. This relation will be essential to build up our estimator. Define all assets in the portfolio as $y_t = (y_{jt}, Y'_{-jt})' : p \times 1$. In case of $j = 3$, this represents y_{3t} in the third cell in the vector y_t and Y_{-3t} represents elements $1, 2, 4, \cdots, p$ in y_t.

Proof of Lemma 1 (i). We will provide formula for j th row and then stack all rows to get an expression. Before that define $\Sigma_{j,j} := E[(y_{jt}-\mu_j)^2]$ and $\Sigma_{j,-j} := E[(y_{jt}-\mu_j)(Y_{-jt}-\mu_{-j})'] : 1 \times p-1$ and represents j th row of Σ with j th element missing, and by symmetricity of Σ, $\Sigma_{j,-j} = (\Sigma_{-j,j})'$. Specifically, define $\Sigma_{-j,j} := E[(Y_{-jt}-\mu_{-j})(y_{jt}-\mu_j)] : p-1 \times 1$, as the j th column of Σ where j th element is missing. In the same way we define $\Sigma_{-j,-j} := E[(Y_{-jt}-\mu_{-j})(Y_{-jt}-\mu_{-j})'] : p-1 \times p-1$ where this matrix is j the row and column deleted from Σ. By Section 2.1 of Yuan (2010), the main diagonal elements are given by the scalar

$$\Theta_{j,j} := (\Sigma_{j,j} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j})^{-1}. \tag{B.1}$$

and the j th row without the j th element is:

$$\Theta_{j,-j} := -\Theta_{j,j}\Sigma_{j,-j}\Sigma_{-j,-j}^{-1}. \tag{B.2}$$

Now we relate these quantities to a linear regression context. Define the following minimizer

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

with given zero mean $(y_{jt} - \mu_j), (Y_{-jt} - \mu_{-j})$ and finite second moments for y_{jt}, Y_{-jt}, Lemma 27 in Bunea et al. (2022) shows that

$$\alpha_j^* = \Sigma_{-j,-j}^{+} \Sigma_{-j,j}, \tag{B.3}$$

where $\Sigma_{-j,-j}^{+}$ is the Moore-Penrose inverse of $\Sigma_{-j,-j}$. We will simplify that expression. First,

$$\Sigma_{-j,-j} = E[(Y_{-jt} - \mu_{-j})(Y_{-jt} - \mu_{-j})'] = A_{-j}\Sigma_f A'_{-j} + \Sigma_{U,-j},$$

by (3) and Assumption 1(iv). Then

$$\begin{aligned} \min_{1\leq j\leq p} Eigmin(\Sigma_{-j,-j}) &= \min_{1\leq j\leq p} Eigmin(A_{-j}\Sigma_f A'_{-j} + \Sigma_{U,-j}) \\ &\geq min_{1\leq j\leq p} Eigmin(\Sigma_{U,-j}) \geq c > 0, \end{aligned}$$

by Assumption 2-3. So (B.3) simplifies to

$$\alpha_j^* = \Sigma_{-j,-j}^{-1} \Sigma_{-j,j}. \tag{B.4}$$

Define $\eta_{jt} := (y_{jt} - \mu_j) - (Y_{-jt} - \mu_{-j})'\alpha_j^*$ for $j = 1, \cdots, p$. Now consider

$$\begin{aligned} E[(Y_{-jt} - \mu_{-j})\eta_{jt}] &= E[(Y_{-jt} - \mu_{-j})(y_{jt} - \mu_j)] - E[(Y_{-jt} - \mu_{-j})(Y_{-jt} - \mu_{-j})']\alpha_j^* \\ &= \Sigma_{-j,j} - \Sigma_{-j,-j}\alpha_j^* \\ &= \Sigma_{-j,j} - \Sigma_{-j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j} = 0, \end{aligned} \tag{B.5}$$

where the first equality uses η_{jt} definition, and the second equality uses definitions of moment matrices, and the third one uses (B.4). Given (B.5) we can write the following infeasible linear regression model,

$$y_{jt} - \mu_j = (Y_{-jt} - \mu_{-j})'\alpha_j^* + \eta_{jt}. \tag{B.6}$$

Now we relate (B.2) using (B.4). Then

$$\Theta_{j,-j} = -\Theta_{j,j}(\alpha_j^*)', \tag{B.7}$$

and then define $\tau_j^2 := E\eta_{jt}^2$. See that by (B.5)(B.6)

$$E[y_{jt} - \mu_j]^2 = (\alpha_j^*)'E[(Y_{-jt} - \mu_{-j})(Y_{-jt} - \mu_{-j})']\alpha_j^* + E[\eta_{jt}^2],$$

which implies by definitions $\Sigma_{j,j} := E[y_{jt} - \mu_j]^2, \Sigma_{-j,-j} := E[(Y_{-jt} - \mu_{-j})(Y_{-jt} - \mu_{-j})']$

$$\begin{aligned}\tau_j^2 &= \Sigma_{j,j} - (\alpha_j^*)'\Sigma_{-j,-j}\alpha_j^* \\ &= \Sigma_{j,j} - \Sigma_{j,-j}\Sigma_{-j,-j}^{-1}\Sigma_{-j,j} \qquad \text{(B.8)} \\ &= 1/\Theta_{j,j}, \qquad \text{(B.9)}\end{aligned}$$

where we use (B.4) for the second equality and (B.1) for the third one. Clearly, also by (B.4)(B.8)

$$\tau_j^2 = \Sigma_{j,j} - \Sigma_{j,-j}\alpha_j^*. \qquad \text{(B.10)}$$

We can form the j th row of precision matrix by using j the main diagonal term, via (B.9)(B.10)

$$\Theta_{j,j} = \frac{1}{\Sigma_{j,j} - \Sigma_{j,-j}\alpha_j^*}, \qquad \text{(B.11)}$$

and the rest of that row, via (B.7)(B.9)

$$\Theta_{j,-j} = \frac{-(\alpha_j^*)'}{\tau_j^2}. \qquad \text{(B.12)}$$

We start simplifying α_j^* formula in case of hidden factor models as in (2)(3). The proof for (i) here is also a detailed version of the proof of Lemma 7 and (57) of Bunea et al. (2022). This is repeated to show that our scenario fits into their framework. Note that $\Sigma_f := E(f_t - Ef_t)(f_t - Ef_t)'$. In that respect in (B.4)

$$\begin{aligned}\Sigma_{-j,j} &= E[(Y_{-jt} - \mu_{-j})(y_{jt} - \mu_j)] = E\{[A_{-j}(f_t - Ef_t) + U_{-jt}][(f_t - Ef_t)'a_j + u_{jt}]\} \\ &= A_{-j}\Sigma_f a_j + A_{-j}E[(f_t - Ef_t)u_{jt}] + E[U_{-jt}(f_t - Ef_t)']a_j + E[U_{-jt}u_{jt}] \\ &= A_{-j}\Sigma_f a_j = \bar{A}_{-j}\bar{a}_j, \qquad \text{(B.13)}\end{aligned}$$

by Assumption 1(iv). Then the same analysis in (B.13) provides with $\bar{A}_{-j} := A_{-j}\Sigma_f^{1/2}$,

$$\Sigma_{-j,-j} = E[(Y_{-jt} - \mu_{-j})(Y_{-jt} - \mu_{-j})'] = A_{-j}\Sigma_f A'_{-j} + \Sigma_{U,-j} = \bar{A}_{-j}\bar{A}'_{-j} + \Sigma_{U,-j}. \qquad \text{(B.14)}$$

Using Sherman-Morrison-Woodbury formula in p.19 of Horn and Johnson (2013)

$$\begin{aligned}\Sigma_{-j,-j}^{-1} &= (A_{-j}\Sigma_f A_{-j}' + \Sigma_{U,-j})^{-1} \\ &= \Sigma_{U,-j}^{-1} - \Sigma_{U,-j}^{-1}A_{-j}(\Sigma_f^{-1} + A_{-j}'\Sigma_{U,-j}^{-1}A_{-j})^{-1}A_{-j}'\Sigma_{U,-j}^{-1} \\ &= \Sigma_{U,-j}^{-1} - \Sigma_{U,-j}^{-1}\bar{A}_{-j}(I_K + \bar{A}_{-j}'\Sigma_{U,-j}^{-1}\bar{A}_{-j})^{-1}\bar{A}_{-j}'\Sigma_{U,-j}^{-1}, \qquad \text{(B.15)}\end{aligned}$$

where we use Assumption 3, Σ_f is positive definite. Note also that

$$\bar{A}_{-j}^{+}\bar{A}_{-j} = \bar{A}_{-j}'(\bar{A}_{-j}^{+})' = I_K, \qquad \text{(B.16)}$$

since A_{-j}, Σ_f has full rank K by Assumption 3, so (135) of Bunea et al. (2022) provides the last result (B.16). Define $\bar{a}_j := \Sigma_f^{1/2}a_j$ and (B.4)(B.13)

$$\begin{aligned}\alpha_j^* &= \Sigma_{-j,-j}^{-1}\bar{A}_{-j}\bar{a}_j \\ &= \Sigma_{-j,-j}^{-1}[\bar{A}_{-j}\{\bar{A}_{-j}'(\bar{A}_{-j}^{+})'\}\bar{a}_j] \\ &= \Sigma_{-j,-j}^{-1}[\Sigma_{-j,-j} - \Sigma_{U,-j}](\bar{A}_{-j}^{+})'\bar{a}_j \\ &= [I_{p-1} - \Sigma_{-j,-j}^{-1}\Sigma_{U,-j}](\bar{A}_{-j}^{+})'\bar{a}_j, \qquad \text{(B.17)}\end{aligned}$$

where we use (B.16) for the second equality and for the third equality (B.14) and $\bar{A}_{-j}$ definition. By using $\bar{G}_j$ definition and using (B.15) on right side of (B.17) for $\Sigma_{-j,-j}^{-1}$

$$\begin{aligned}I_{p-1} - \Sigma_{-j,-j}^{-1}\Sigma_{U,-j} &= I_{p-1} - [\Sigma_{U,-j}^{-1} - \Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{A}_{-j}'\Sigma_{U,-j}^{-1}]\Sigma_{U,-j} \\ &= \Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{A}_{-j}' \qquad \text{(B.18)}\end{aligned}$$

Use (B.18) in (B.17)

$$\begin{aligned}\alpha_j^* &= \Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{A}_{-j}'(\bar{A}_{-j}^{+})'\bar{a}_j \\ &= \Sigma_{U,-j}^{-1}\bar{A}_{-j}\bar{G}_j^{-1}\bar{a}_j, \qquad \text{(B.19)}\end{aligned}$$

by (B.16) for the second equality. **Q.E.D.**

(ii). One of the other elements in the precision matrix is the reciprocal of the main diagonal term, τ_j^2 in (B.10). We provide a formula for that in terms of hidden factor regression model, which will be new. In that respect, by (2)(3), Assumption 1

$$\Sigma_{j,j} := E(y_{jt} - \mu_j)^2 = a_j'\Sigma_f a_j + \sigma_j = \bar{a}_j'\bar{a}_j + \sigma_j^2. \qquad \text{(B.20)}$$

Similarly, using (2)(3), Assumption 1

$$\Sigma_{j,-j} := E(y_{jt} - \mu_j)(Y_{-jt} - \mu_{-j})' = \bar{a}_j' \bar{A}_{-j}'. \tag{B.21}$$

Then combine (B.20)(B.21) and Lemma 1(i) to have the desired result. **Q.E.D.**

REFERENCES

ABADIR, K. AND J. MAGNUS (2005): *Matrix Algebra*, Cambridge University Press. [47, 49]

AGARWAL, A., M. DAHLEH, D. SHAH, AND D. SHEN (2023): "Causal Matrix Completion," in *Proceedings of Thirty Sixth Conference on Learning Theory*, PMLR, vol. 195 of *Proceedings of Machine Learning Research*, 3821–3826. [3]

ANGRIST, J. AND J-S PISCHKE (2009): *Mostly Harmless Econometrics: An empiricist's companion*, Princeton University Press. [2]

ATHEY, S., M. BAYATI, N. DOUDCHENKO, G. IMBENS, AND K. KHOSRAVI (2021): "Matrix Completion Methods for Causal Panel Data Models," *Journal of the American Statistical Association*, 116 (536), 1716–1730. [3]

ATHEY, S., M. BAYATI, G. IMBENS, AND Z. QU (2019): "Ensemble Methods for Causal Effects in Panel Data Settings," *AEA Papers and Proceedings*, 109, 65–70. [3]

BAI, J. AND S. NG (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70 (1), 191–221. [6]

——— (2006): "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions," *Econometrica*, 76 (4), 1133–1150. [6]

——— (2021): "Matrix Completion, Counterfactuals, And Factor Analysis of Missing Data," *Journal of the American Statistical Association*, 116 (536), 1746–1763. [3]

BELKIN, M., D. HSU, S. MA, AND S. MANDAL (2019): "Reconciling modern machine-learning practice and the classical bias-variance trade-off," *Proc. Natl. Acad. Sci. USA*, 116, 15849–15854. [6]

BELKIN, M., D. HSU, AND J. XU (2020): "Two Models of Double Descent for Weak Features," *SIAM Journal on Mathematics of Data Science*, 2 (4), 1167–1180. [6]

BERNSTEIN, D.S. (2018): *Scalar vector, and matrix mathematics, theory facts and formulas*, Princeton University Press. [56, 58]

BING, X., F. BUNEA, S. STRIMAS-MACKEY, AND M. WEGKAMP (2021): "Prediction under latent factor regression: adaptive pcr, interpolating predictors and beyond," *Journal of Machine Learning Research*, 22, 1–50. [4, 5, 8, 9, 10, 11, 14, 17, 18, 19, 30, 31, 46, 47, 48, 49, 51, 54, 57, 58, 61]

BING, X. AND M. WEGKAMP (2019): "Adaptive estimation of the rank of the coefficient matrix in high dimensional multivariate response regression models," *Annals of Statistics*, 47, 3157–3184. [19]

BROOKS, R. AND M. DEL NEGRO (2004): "A Latent Factor Model with Global Country, Industry Shocks for International Stock Returns," *Working Paper, Federal Reserve Bank of Atlanta*. [3]

BRYZGALOVA, S., S. LERNER, M. LETTAU, AND M. PELGER (2023): "Missing Financial Data," *Working Paper*. [3]

BUNEA, F., S. STRIMAS-MACKEY, AND M. WEGKAMP (2022): "Interpolating Predictors in High Dimensional Factor Estimation," *Journal of Machine Learning Research*, 23, 1–60. [4, 5, 8, 9, 10, 11, 14, 15, 17, 20, 48, 52, 53, 57, 69, 72, 73, 74]

CAHAN, E., J. BAI, AND S. NG (2023): "Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions," *Journal of Econometrics*, 233 (1), 113–131. [3]

CALLOT, L., M. CANER, O. ONDER, AND E. ULASAN (2021): "A nodewise regression approach to estimating large portfolios," *Journal of Business and Economic Statistics*, 39, 520–531. [2, 7, 14, 31, 36, 38, 71]

CANER, M. (2023): "Generalized Linear Models with Structured Sparsity Estimators," *Journal of Econometrics*, 236(2), 105478. [3]

CANER, M AND A.B. KOCK (2018): "Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso," *Journal of Econometrics*, 203, 143–168. [3, 7, 14, 29, 30, 71]

CANER, M., M. MEDEIROS, AND G.F.R. VASCONCELOS (2023): "Sharpe Ratio analysis in high dimensions: Residual based nodewise regression in factor models," *Journal of Econometrics*, 235(2), 393–417. [2, 7, 10, 14, 28, 29, 36, 38]

CHEN, M., I. FERNANDEZ, AND M. WEIDNER (2021): "Nonlinear factor models for network and panel data," *Journal of Econometrics*, 220 (2), 296–324. [3]

DUAN, J., M. PELGER, AND R. XIONG (2024): "Target PCA: Transfer Learning Large Dimensional Panel Data," *Journal of Econometrics*, 244 (2), 105521. [3]

FAN, J., Y. FAN, AND J. LV (2008): "High-dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, 147, 186–197. [9]

FAN, J., Y. LIAO, AND M. MINCHEVA (2011): "High-dimensional covariance matrix estimation in approximate factor models," *The Annals of Statistics*, 39, 3320–3356. [6]

FAN, JIANQING, YUAN LIAO, AND MARTINA MINCHEVA (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75 (4), 603–680. [4, 6, 7, 25, 26, 31, 38]

HASTIE, T., A. MONTANARI, S. ROSSET, AND R.J. TIBSHIRANI (2022): "Surprises in high dimensional ridgeless least squares interpolation," *Annals of Statistics*, 50, 949–987. [6]

HORN, R. AND C. JOHNSON (2013): *Matrix Analysis*, Cambridge University Press. [49, 74]

IMBENS, G. AND D. RUBIN (2015): *Causal Inference in Statistics, Social and Biomedical Sciences*, Cambridge University Press. [2]

KALLUS, N., X. MAO, AND M. UDELL (2018): "Causal inference with noisy and missing covariates via matrix factorization," *Proceedings of the 32rd International Conference on Neural Information Processing*, 6921–6932. [3]

KALTENPOTH, D. AND J. VREEKE (2023): "Casual Discovery with Hidden Confounders using the Algorithmic Markov Condition," *Proceedings of Machine Learning Research*, 216, 1016–1026. [2]

LEDOIT, O, M. AND M. WOLF (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15 (5), 850–859. [39]

——— (2017): "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks," *The Review of Financial Studies*, 30, 4349–4388. [7, 31, 38]

LETTAU, M. AND M. PELGER (2020a): "Estimating latent asset pricing," *Journal of Econometrics*, 218, 1–31. [3, 4]

——— (2020b): "Factors that fit time series and cross section of stock returns," *Review of Financial Studies*, 33, 2274–2325. [4]

MEINSHAUSEN, NICOLAI AND PETER BÜHLMANN (2006): "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, 1436–1462. [2, 4, 7]

MUTHUKUMAR, V., K. VODRAHALLI, V. SUBRAMANIAN, AND A. SAHAI (2020): "Harmless Interpolation of Noisy Data in Regression," *IEEE J. Sel. Areas Inf. Theory*, 1 (1), 67–83. [6]

SPIGLER, S., M. GEIGER, S. D'ASCOLI, L. SAGUN, G. BIROLLI, AND M. WYART (2019): "A jamming transition from under- to over-parametrization affects generalization in deep learning," *J. Phys. A*, 52, 474001. [6]

STOCK, J. H AND M. W. WATSON (2002a): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97 (460), 1167–1179. [3]

——— (2002b): "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, 27 (2), 147–162. [3]

——— (2016): "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," in *Handbook of macroeconomics*, Elsevier, vol. 2, 415–525. [3]

VAN DE GEER, SARA, PETER BÜHLMANN, YA'ACOV RITOV, AND RUBEN DEZEURE (2014): "On asymptotically optimal confidence regions and tests for high-dimensional models," *The Annals of Statistics*. [3, 7, 29]

VERSHYNIN, R. (2012): *Introduction to non-asymptotic analysis of random matrices*, Cambridge University Press. [57]

——— (2019): *High Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press. [57, 58, 62, 63, 64, 65]

XIONG, R. AND M. PELGER (2023): "Large Dimensional Latent Factor Modeling with Missing Observations and Applications to Causal Inference," *Journal of Econometrics*, 233 (1), 271–301. [3]

YAN, Y. AND M. J. WAINWRIGHT (2024): "Entrywise Inference for Causal Panel Data: A Simple and Instance-Optimal Approach," *Working paper*. [3]

YUAN, M. (2010): "High dimensional inverse covariance matrix estimation via linear programming," *Journal of Machine Learning Research*, 11, 2261–2286. [71]