Essays on Forecasting with Many Predictors

by

Juhui Bae

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2021 by the
Graduate Supervisory Committee:

Seung Ahn, Chair
Seth Pruitt
Nicolai Kuminoff
Domenico Ferraro

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

This dissertation studies how forecasting performance can be improved in big data. The first chapter with Seung C. Ahn considers Partial Least Squares (PLS) estimation of a time-series forecasting model with data containing a large number of time series observations of many predictors. In the model, a subset or a whole set of the latent common factors in predictors determine a target variable. First, the optimal number of the PLS factors for forecasting could be smaller than the number of the common factors relevant for the target variable. Second, as more than the optimal number of PLS factors is used, the out-of-sample explanatory power of the factors could decrease while their in-sample power may increase. Monte Carlo simulation results also confirm these asymptotic results. In addition, simulation results indicate that the out-of-sample forecasting power of the PLS factors is often higher when a smaller than the asymptotically optimal number of factors are used. Finally, the out-of-sample forecasting power of the PLS factors often decreases as the second, third, and more factors are added, even if the asymptotically optimal number of the factors is greater than one. The second chapter studies the predictive performance of various factor estimations comprehensively. Big data that consist of major U.S. macroeconomic and finance variables, are constructed. 148 target variables are forecasted, using 7 factor estimation methods with 11 information criteria. First, the number of factors used in forecasting is important and Incorporating more factors does not always provide better forecasting performance. Second, using consistently estimated number of factors does not necessarily improve predictive performance. The first PLS factor, which is not theoretically consistent, very often shows strong forecasting performance. Third, there is a large difference in the forecasting performance across different information criteria, even when the same factor estimation method is used. Therefore, the choice of factor estimation method, as well as the information criterion, is crucial in forecasting practice. Finally, the first PLS factor yields forecasting performance very close to the best result from the total combinations of the 7 factor estimation methods and 11 information criteria.

DEDICATION

이 학위 논문을 제가 사랑하는 가족들, 아버지, 어머니, 성훈이, 막내 초롱이, 할머니, 그리고 나즘에게 바칩니다. 특히 부모님의 헌신과 도움으로 박사 학위를 받을 수 있었습니다. 감사합니다.

ACKNOWLEDGMENTS

I would like to thank Seung C. Ahn for his guidance and support. I am also grateful to Domenico Ferraro, Nicolai Kuminoff, Seth Pruitt, Berthold Herrendorf, Bart Hobijn and Gustavo Ventura for their feedback and comments. I also thank Nazim Tamkoç for his support and help throughout the program. This paper was presented in seminars and workshops at Arizona State University, University of Glasgow, Northeastern University and Korea Institute for International Economic Policy (KIEP). I would like to thank the participants in the seminars and workshops.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

FORECASTING WITH PARTIAL LEAST SQUARES WHEN A LARGE NUMBER OF PREDICTORS ARE AVAILABLE (WITH SEUNG C. AHN)

1.1 Introduction

When a large number (N) of predictor variables are available for forecasting a single target variable, the Ordinary Least Squares (OLS) regression produces poor forecasting results because of high multicollinearity among predictors, especially when the number (T) of time series observations is not sufficiently larger than N; see Huber *et al.* (1973), Stein (1956), and Stock and Watson (2006), among many. A treatment to this large-dimensionality problem is the use of shrinkage estimation methods such as Ridge, Bayesian, and Principal Component (PC) regressions; see De Mol *et al.* (2008). Another possible choice is the Partial Least Squares (PLS) regression that was originally introduced and developed by Wold (1966, 1973, 1982). The PLS regression is also a shrinkage estimation method; see, for example, Jong (1993) and Phatak and de Hoog (2002).[1] The PLS regression has been popularly used in chemometrics, bioinformatics, machine learning and marketing research, especially for the cases in which obtaining a sufficiently large number of data observations per each predictor variable is restrictively expensive. Recently, use of the PLS regression has been increasingly popular in the fields of finance and economics, especially for the analysis

---

[1] The PLS regression is a shrinkage method in the sense that the norm of the OLS estimates of the coefficients of the PLS factors is not greater than that of the OLS estimates of the coefficients of all predictors. However, differently from the ridge and the Bayesian regressions, the PLS regression does not shrink all of the regressor coefficients. It could rather expand some coefficients; see Butler and Denham (2000).

of the data with both large N and large T; see, for example, Groen and Kapetanios (2009, 2016), Kelly and Pruitt (2013, 2015), Huang *et al.* (2015), Carrasco and Rossi (2016), Light *et al.* (2017),Tu and Lee (2019), and Rytchkov and Zhong (2020).

The PC and PLS regressions are similar in the sense that both extract a small number of common factors in predictor variables and use the extracted factors to forecast a target variable. However, they use different methods to extract the factors. Specifically, the PC regression estimates and uses for forecasting all of the common factors in predictor variables even if some of the factors are in fact uncorrelated with the target variable. For this reason, the PC method is viewed as an "unsupervised" method because the common factors are estimated independently from the target variable. In contrast, the PLS regression generates relevant factors sequentially by the "Nonlinear Iterative Partial Least Squares" (NIPLS) algorithm of Wold (1966). The PLS regression is a "supervised" method because it isolates and estimates relevant factors from the latent factors that are correlated with the target variable; see Mehmood *et al.* (2012). For this reason, many previous studies have conjectured that the PLS factors may have greater predicting power than the PC factors. The purpose of this paper is to revisit this conjecture investigating the asymptotic and finite sample properties of the PLS factors when they are obtained from the data with both large N and large T.

The large-N and large-T properties of the PLS factors have been studied by Kelly and Pruitt (2015) and Groen and Kapetanios (2016). Kelly and Pruitt (2015) consider the cases in which a subset or a whole set of the common factors in predictors are the determinant of a target variable and individual predictor variables are correlated with the target variable only through the common factors. Groen and Kapetanios (2016) examine the forecasting power of PLS factors for the cases in which the predictor variables are directly correlated with the target variables, not just indirectly through the latent factors. We do not consider the model of Groen and Kapetanios (2016) in this paper. Our asymptotic analysis is conducted for a model in which predictor variables are correlated with the target variable only through the latent factors. However, our model is more general than that of Kelly and Pruitt (2015). For the general model, we investigate the asymptotically optimal number of the PLS factors that have the maximum explanatory power for the target variable. We also conduct some Monte Carlo simulations to examine the finite-sample properties of the forecasting results by the PLS regression. For our simulation exercises, we consider some cases in which the idiosyncratic components of predictor variables, as well as the common latent factors, are correlated with the target variable.

The PLS regression uses a smaller number of factors than the PC regression to reach the maximum prediction power. For the cases in which asymptotic theory applies as T grows infinitely with fixed N, Helland (1988, 1990) has shown that the number of the distinct eigenvalues of the population variance-covariance matrix of the predictor variables is the optimal number of the PLS factors to be used. In this paper we examine how his result can be generalized to the cases in which asymptotic theory applies as both N and T jointly grow infinitely. Most of the previous studies related to large-N and large-T properties of the PC or PLS factors have considered the cases in which predictor variables contain K common latent factors and the first K largest eigenvalues of the sample variance-covariance matrix of the predictor variables are asymptotically distinct (e.g., converges to different limits in probability); see Bai (2003), Stock and Watson (2002a), and Kelly and Pruitt (2015). For such cases, each of the eigenvectors corresponding to the K eigenvalues is asymptotically unique up to sign and scale. A novelty of our model is that it allows some or all of the K largest eigenvalues to have the same probability limits. For this general model, the eigenvectors corresponding to the eigenvalues having the same probability limit are unique only up to orthonormal transformation. We find that this generalization is important to understand the asymptotic and finite-sample properties of the PLS factors.

There are two major findings from our asymptotic analysis. First, we find that the asymptotically optimal number of the PLS factors crucially depends on the asymptotic distribution of the eigenvalues of the sample variance-covariance matrix of the predictors. For example, if all of the K largest eigenvalues converge to the same probability limit, the first PLS factor has the maximum prediction power that the PLS regression can have. In contrast, if the K eigenvalues are all asymptotically distinct as in Kelly and Pruitt (2015), the asymptotically optimal number of the PLS factors equals the number of the common factors in predictor variables that are correlated with the target variable. Second, using overly many PLS factors could substantially decrease the out-of-sample forecasting power of the PLS regression unless the ratio N/T is sufficiently small. While using more PLS factors does not decrease the PLS regression's in-sample prediction power, it can deteriorate the regression's out-of-sample forecasting power.

The three major findings from our simulation experiments and topical empirical study are the following. First, in finite samples, the out-of-sample prediction power of the PLS regression often sharply drops as more than the asymptotically optimal number of factors are used. Second, unless the ratio N/T is sufficiently small, the out-of-sample prediction power of the PLS regression is often peaked when a fewer number of factors are used. The first PLS factor has dominantly strong forecasting power than other PLS factors, even for the cases in which the asymptotically optimal number of PLS factors is greater than one. The gain by using the second or other PLS factors in addition to the first PLS factor is generally small. Third and finally, cross-validation methods are not always successful in the number of factors that maximizes the out-of-sample forecasting power of the PLS regression. Our simulation experiments and actual data analysis show that using only the first PLS factor often produces better forecasting results.

This paper is organized as follows. Section 1.2 introduces the model we consider and states the asymptotic properties of the PLS factors. Our Monte Carlo simulation results are reported in Section 1.3, while some results from a topical empirical study are reported in Section 1.4. Some concluding remarks follow in Section 1.5. Proofs of the theorems and lemmas are all given in Appendix A.

Throughout this paper, we use the following notation. For an $a \times a$ symmetric matrix $\boldsymbol{A}$, $\lambda_h(\boldsymbol{A})$ denotes the hth largest eigenvalue of $\boldsymbol{A}$; $\boldsymbol{\Lambda}(\boldsymbol{A}|h'+1:h'')$ denotes the diagonal matrix of $\lambda_{h'+1}(\boldsymbol{A})$, ... , $\lambda_{h''}(\boldsymbol{A})$, where h', $h'' \leq a$. The notation $\xi_h(\boldsymbol{A})$ stands for the $a \times 1$ eigenvector of $\boldsymbol{A}$ corresponding to $\lambda_h(\boldsymbol{A})$. We also use $\boldsymbol{\Xi}(\boldsymbol{A}|h'+1:h'') = [\xi_{h'+1}(\boldsymbol{A}),\ \dots\ ,\xi_{h''}(\boldsymbol{A})]$. For an $a \times b$ full-column rank matrix $\boldsymbol{B}$, $\mathcal{P}(\boldsymbol{B}) = \boldsymbol{B}(\boldsymbol{B'}\boldsymbol{B})^{-1}\boldsymbol{B'}$ and $\mathcal{Q}(\boldsymbol{B}) = \boldsymbol{I}_a - \mathcal{P}(\boldsymbol{B})$. For an $a \times b$ matrix $\boldsymbol{B}$(not necessarily a full-column rank matrix), the spectral and the Frobenius norms of $\boldsymbol{B}$ are respectively denoted by $\|\boldsymbol{B}\|_2 = [\lambda_1(\boldsymbol{B'}\boldsymbol{B})]^{1/2}$ and $\|\boldsymbol{B}\|_F = [\mathrm{trace}(\boldsymbol{B'}\boldsymbol{B})]^{1/2} = [\Sigma_{h=1}^{b}\lambda_h(\boldsymbol{B'}\boldsymbol{B})]^{1/2}$. Finally, we denote "converges in probability" and "converges in distribution" by "$\rightarrow_p$" and "$\rightarrow_d$", respectively.

1.2 Model and Asymptotic Properties of PLS Factors

1.2.1 Model and Some Preliminary Results

This subsection introduces the model for which we investigate the large-N and large-T asymptotic properties of PLS factors. The model we consider is a forecasting model in which N predictor variables are available for forecasting a single target variable. The model consists of two parts. The first one is a factor model in which N predictor variables are generated by K latent factors, and the second part is a forecasting model for a single target variable. Stated formally:

$$x_{it} = \boldsymbol{f'}_{\cdot t}\boldsymbol{\phi}_{\cdot i} + e_{it} = \Sigma_{j=1}^{J}\boldsymbol{f'}_{(j)t}\boldsymbol{\phi}_{(j)i} + e_{it} \tag{1.1}$$

$$y_{t+1} = \Sigma_{j=1}^{J}\boldsymbol{f'}_{(j)t}\boldsymbol{\beta}_{(j)} + u_{t+1} = \boldsymbol{f'}_{\cdot t}\boldsymbol{\beta} + u_{t+1} \tag{1.2}$$

where $i(=1,...,N)$ indexes different predictor variables, $t(=1,...,T)$ indexes time, $\boldsymbol{f}_{(j)t}$ is a $k(j)\times 1$ random vector of latent factors, $\boldsymbol{\phi}_{(j)i}$ is a $k(j)\times 1$ vector of factor loadings corresponding to $\boldsymbol{f}_{(j)t}$, $\boldsymbol{f}_{\cdot t} = (\boldsymbol{f'}_{(1)t}, ..., \boldsymbol{f'}_{(J)t})'$, $\boldsymbol{\phi}_{\cdot i} = (\boldsymbol{\phi'}_{(1)i}, ..., \boldsymbol{\phi'}_{(J)i})'$, $\boldsymbol{\beta}_{(j)}$ is $k(j)\times 1$ vector of regression coefficients on $\boldsymbol{f}_{(j)t}$, $\boldsymbol{\beta} = (\boldsymbol{\beta'}_{(1)}, ..., \boldsymbol{\beta'}_{(J)})'$, the e_{it} and u_{t+1} are random noises, and $K = \Sigma_{j=1}^{J}k(j)$. We later discuss how the factors in $\boldsymbol{f}_{\cdot t}$ are sorted into the J different groups, $\boldsymbol{f}_{(1)t}$, ... , $\boldsymbol{f}_{(J)t}$. Without loss of generality, we assume that $\mathrm{E}(\boldsymbol{f}_{\cdot t}) = 0_{K\times 1}$ and $\mathrm{E}(e_{it}) = \mathrm{E}(u_{t+1}) = 0$, for all i and t. For the cases in which $\boldsymbol{f}_{\cdot t}$, x_{it} and y_{t+1} have non-zero means, we can replace them in (1.1) and (1.2) respectively by their demeaned versions, $\boldsymbol{f}_{\cdot t} - \bar{\boldsymbol{f}}_{\cdot}$, $\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i$, and $y_{t+1} - \bar{y}$, where $\bar{x}_i = T^{-1}\Sigma_{t=1}^{T}x_{it}$, $\bar{\boldsymbol{f}}_{\cdot} = T^{-1}\Sigma_{t=1}^{T}\boldsymbol{f}_{\cdot t}$, and $\bar{y} = T^{-1}\Sigma_{t=1}^{T}y_{t+1}$.

Stacking the equations for individual predictors in (1.1) vertically, we have

$$\boldsymbol{x}_{\cdot t} = \Sigma_{j=1}^{J}\boldsymbol{\Phi}_{(j)}\boldsymbol{f}_{(j)t} + \boldsymbol{e}_{\cdot t} = \boldsymbol{\Phi}\boldsymbol{f}_{\cdot t} + \boldsymbol{e}_{\cdot t} \tag{1.3}$$

where $\boldsymbol{x}_{\cdot t} = (x_{1t}, ..., x_{Nt})'$ and $\boldsymbol{e}_{\cdot t} = (e_{1t}, ..., e_{Nt})'$, $\boldsymbol{\Phi}_{(j)} = (\boldsymbol{\phi}_{(j)1}, ..., \boldsymbol{\phi}_{(j)N})'$, and $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_{(1)}, ..., \boldsymbol{\Phi}_{(J)})$. The equations in (1.3) and (1.2) can be rewritten by the following two matrix equations:

$$\boldsymbol{X} = \Sigma_{j=1}^{J}\boldsymbol{F}_{(j)}\boldsymbol{\Phi'}_{(j)} + \boldsymbol{E} = \boldsymbol{F}\boldsymbol{\Phi'} + \boldsymbol{E} \tag{1.4}$$

$$\boldsymbol{y} = \Sigma_{j=1}^{J}\boldsymbol{F}_{(j)}\boldsymbol{\beta}_{(j)} + \boldsymbol{u} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{u} \tag{1.5}$$

where $\boldsymbol{X} = (\boldsymbol{x}_{\cdot 1}, ..., \boldsymbol{x}_{\cdot T})'$, $\boldsymbol{F}_{(j)} = (\boldsymbol{f}_{(j)1}, ..., \boldsymbol{f}_{(j)T})'$, $\boldsymbol{F} = (\boldsymbol{F}_{(1)}, ..., \boldsymbol{F}_{(J)})$, $\boldsymbol{E} = (\boldsymbol{e}_{\cdot 1}, ..., \boldsymbol{e}_{\cdot T})'$, $\boldsymbol{y} = (y_2, ..., y_{T+1})'$, and $\boldsymbol{u}$ is similarly defined.

For the model given in (1.4) and (1.5), our interest lies in forecasting y_{T+2} using the data available up to time $T+1$. We can forecast y_{T+2} using the PC or PLS factors. For heuristic discussions, we momentarily consider the model in (1.4) and (1.5) under some preliminary assumptions that are unrealistically restrictive.

Preliminary Assumption (PA): (i) $\boldsymbol{E} = \mathbf{0}_{T\times N}$. (ii) The variable groups, $\{\boldsymbol{f}_{\bullet t}\}$ and $\{u_{t+1}\}$, are mutually independent. (iii) The factor vectors $\boldsymbol{f}_{\bullet t}$ are independently and identically distributed (*iid*) over time with $\mathrm{Var}(\boldsymbol{f}_{(j)t}) = \sigma_j^2 \boldsymbol{I}_{k(j)}$ where $\sigma_1^2 > \sigma_2^2 > ... > \sigma_J^2$. (iv) The errors u_{t+1} are *iid* with $\mathrm{var}(u_{t+1}) = \sigma_u^2$. (v) $\boldsymbol{\Phi}$ is a fixed matrix with $\boldsymbol{\Phi'\Phi}/N = \boldsymbol{I}_K$.

Some remarks follow on PA. First, under (i), the predictors x_{it} do not have any idiosyncratic components. This assumption is made to find more clearly what the PC and PLS factors estimate. Second, the assumptions (iii) and (v) are by no means too restrictive assumptions. Suppose that the true factor vector $\boldsymbol{f}_{\bullet t}^*$ have an unrestrictive variance-covariance matrix $\boldsymbol{\Sigma}^*$and the factor loading matrix $\boldsymbol{\Phi}^*$ does not satisfy the assumption (v). Let $\boldsymbol{f}_{\bullet t} = \boldsymbol{f}_{\bullet t}^*(N^{-1}\boldsymbol{\Phi'\Phi})^{1/2}\boldsymbol{\Xi}^*$ and $\boldsymbol{\Phi} = \boldsymbol{\Phi}^*(N^{-1/2}\boldsymbol{\Phi'\Phi})^{-1/2}\boldsymbol{\Xi}^*$, where $\boldsymbol{\Xi}^* = \boldsymbol{\Xi}((N^{-1}\boldsymbol{\Phi'\Phi})^{1/2}\boldsymbol{\Sigma}^*(N^{-1}\boldsymbol{\Phi'\Phi})^{1/2}|1 : K)$. Then, we can easily see that $\boldsymbol{\Phi f}_{\bullet t} = \boldsymbol{\Phi}^*\boldsymbol{f}_{\bullet t}^*$ and $N^{-1/2}\boldsymbol{\Phi'\Phi} = \boldsymbol{I}_K$. That is, unrestricted factors and factor loadings can be reparameterized so that they can satisfy conditions (iii) and (v). Third and finally, for the factors having the same variances, it is not possible to identify which factors among them are correlated with y_{t+1} and which factors are not. Such factors are identified only up to an orthogonal transformation.[2]

Under condition (v), the explanatory power of a factor in $\boldsymbol{f}_{\bullet t}$ for individual predic-

---

[2] This result is for the same reason that the eigenvectors corresponding to a repetitive eigenvalue of a matrix are unique up to an orthogonal transformation.

tor variables x_{it} are on average proportional to the factor's variance. In the literature, it is often assumed that the individual factors in $\boldsymbol{f}_{\boldsymbol{.}t}$ have distinctly different average explanatory power for response variables (predictor variables in our case); for example, see Stock and Watson (2002a), Bai (2003), and Kelly and Pruitt (2015). A novelty of our analysis is that we allow some factors to have the same explanatory power. This generation is important to understand the asymptotic and finite-sample properties of PLS factors.

The asymptotic properties of the PC and PLS factors depend on two terms:

$$\boldsymbol{S}_{NT} = \frac{\boldsymbol{X'X}}{NT}; \quad \boldsymbol{b}_{NT} = \frac{\boldsymbol{X'y}}{N^{1/2}T} \tag{1.6}$$

We scale down each term by NT and $N^{1/2}T$, respectively, to facilitate our asymptotic analysis. For the forecasting with the PC factors, we define the following. For an integer $q = 1, ...,$

$$\hat{\boldsymbol{A}}^{PC}_{1:q} = (\hat{\boldsymbol{\alpha}}^{PC}_1, ..., \hat{\boldsymbol{\alpha}}^{PC}_q) = \Xi(\mathbf{S}_{NT}|1:q) \tag{1.7}$$

$$\hat{\boldsymbol{F}}^{PC}_{1:q} = \boldsymbol{X}\hat{\boldsymbol{A}}^{PC}_{1:q} \tag{1.8}$$

$$\hat{\boldsymbol{\delta}}^{PC}_{1:q} = (\hat{\boldsymbol{F}}^{PC\prime}_{1:q}\hat{\boldsymbol{F}}^{PC}_{1:q})^{-1}\hat{\boldsymbol{F}}^{PC\prime}_{1:q}\boldsymbol{y} \tag{1.9}$$

$$\hat{y}^{PC}_{T+2|q} = \boldsymbol{x}'_{\boldsymbol{.}T+1}\hat{\boldsymbol{A}}^{PC}_{1:q}\hat{\boldsymbol{\delta}}^{PC}_{1:q} \tag{1.10}$$

Here, $\hat{\boldsymbol{A}}^{PC}_{1:q}$ is the $N \times q$ matrix of the PC factor loading estimates, $\hat{\boldsymbol{F}}^{PC}_{1:K}$ is a $T \times q$ matrix of q PC factors, $\hat{\boldsymbol{\delta}}^{PC}_{1:K}$ is the OLS estimator obtained by regressing $\boldsymbol{y}$ on $\hat{\boldsymbol{F}}^{PC}_{1:q}$, and $\hat{y}^{PC}_{T+2|q}$ denotes the forecast for y_{T+2} obtained by the first q PC factors. Under PA, if both $\boldsymbol{f}_{\boldsymbol{.}T+1}$ and $\boldsymbol{\beta}$ were observable, the best forecast for y_{T+2} is $\hat{y}^*_{T+2} = \boldsymbol{f}'_{\boldsymbol{.}T+1}\boldsymbol{\beta}$ $= \Sigma^J_{j=1}\boldsymbol{f}'_{(j)T+1}\boldsymbol{\beta}_{(j)}$. By Bai and Ng (2006), the forecast $\hat{y}^{PC}_{T+2|K}$ that is obtained using K PC factors is a consistent estimator of the best forecast $\hat{y}^*_{T+2}$.[3]

---

[3] Of course, this does not mean that the principal component analysis always produces the best forecasts in finite samples; see De Mol *et al.* (2008). Other competing alternatives such as Ridge, Bayesian and PLS regressions may produce better forecasting results in finite samples.

Alternatively, the PLS regression can be used to consistently estimate $\hat{y}^*_{T+2}$. For the forecasting with PLS factors, we define the PLS estimate of the $N \times q$ matrix of the factor loadings by

$$\tilde{\boldsymbol{A}}^{PLS}_{1:q} = (\tilde{\boldsymbol{\alpha}}^{PLS}_1, ..., \tilde{\boldsymbol{\alpha}}^{PLS}_q) = (\boldsymbol{b}_{NT}, \mathbf{S}_{NT}\boldsymbol{b}_{NT}, ..., (\mathbf{S}_{NT})^{(q-1)}\boldsymbol{b}_{NT}) \tag{1.11}$$

which is of the form of a Krylov matrix. We also define:

$$\tilde{\boldsymbol{P}}^{PLS}_{1:q} = (\tilde{\boldsymbol{p}}^{PLS}_1, ..., \tilde{\boldsymbol{p}}^{PLS}_q) = \boldsymbol{X}\tilde{\boldsymbol{A}}^{PLS}_{1:q} \tag{1.12}$$

$$\tilde{\boldsymbol{\delta}}^{PLS}_{1:q} = (\tilde{\boldsymbol{P}}^{PLS\prime}_{1:q}\tilde{\boldsymbol{P}}^{PLS}_{1:q})^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:q}\boldsymbol{y} \tag{1.13}$$

$$\tilde{y}^{PLS}_{T+2|q} \equiv \boldsymbol{x}'_{\cdot T+1}\tilde{\boldsymbol{A}}^{PLS}_{1:q}\tilde{\boldsymbol{\delta}}^{PLS}_{1:q} \tag{1.14}$$

Here, $\tilde{\boldsymbol{P}}^{PLS}_{1:q}$ is the $T \times q$ matrix of the first q PLS factors, $\tilde{\boldsymbol{\delta}}^{PLS}_{1:q}$ is the OLS estimator obtained regressing $\boldsymbol{y}$ on $\tilde{\boldsymbol{P}}^{PLS}_{1:q}$, and $\tilde{y}^{PLS}_{T+2|q}$ is the the forecast for y_{T+2} by using the first q PLS factors.

The factors defined in (1.12) are different from the PLS factors that are sequentially generated by the Nonlinear Iterative Partial Least Squares (NIPLS) algorithm. However, as Helland (1988, 1990) has shown, the factors of form (1.12) span the same space as the factors generated by the NIPLS algorithm, and both factors produce the same forecasts. Thus, we refer to the factors of form (1.12) as the PLS factors without distinguishing them from the PLS factors generated by the NIPLS algorithm.

We investigate the asymptotic properties of PLS factors using $\tilde{\boldsymbol{P}}^{PLS}_{1:q}$, because their asymptotic properties are much easier to analyze than those of the factors from the NIPLS algorithm. However, we note that the PLS factors computed by the NIPLS algorithm are better to use for actual data analysis. Krylov matrices are generally highly ill-conditioned matrices and computation of them often generates numerical errors; see Dax (2017). Consequently, the PLS factors computed by (1.8) are more likely to contain serious numerical errors. For actual applications, the PLS factors generated by the NIPLS algorithm are numerically more accurate. For this reason, we use the NIPLS procedure for our simulation experiments and actual data analysis. The NIPLS algorithm is described in Appendix A.

An important issue in using the PLS regression is how to find the optimal $q(q^*)$ for forecasting y_{T+2}. Helland (1990) finds that q^* could be smaller than the optimal number of the PC factors for forecasting y_{T+2}. For an intuition on his result, let us consider the "population versions" of $\hat{\boldsymbol{A}}_{1:K}^{PC}$, $\hat{\boldsymbol{\delta}}_{1:K}^{PC}$, $\tilde{\boldsymbol{A}}_{1:J}^{PLS}$, and $\tilde{\boldsymbol{\delta}}_{1:J}^{PLS}$, which are computed replacing $\boldsymbol{b}_{NT}$ and $\mathbf{S}_{NT}$ by $\mathrm{E}(\boldsymbol{b}_{NT})$ and $\mathrm{E}(\mathbf{S}_{NT})$, respectively. Let us denote them by $\boldsymbol{A}_{1:K}^{PC}$, $\boldsymbol{\delta}_{1:K}^{PC}$, $\boldsymbol{A}_{1:J}^{PLS}$, and $\boldsymbol{\delta}_{1:J}^{PLS}$, respectively. Under PA, we can easily find that

$$\mathrm{E}\left(\boldsymbol{b}_{NT}\right) = \frac{1}{N^{1/2}}\Sigma_{j=1}^{J}\sigma_j^2\boldsymbol{\Phi}_{(j)}\boldsymbol{\beta}_{(j)}; \quad \mathrm{E}\left(\mathbf{S}_{NT}\right) = \frac{1}{N}\Sigma_{j=1}^{J}\sigma_j^2\boldsymbol{\Phi}_{(j)}\boldsymbol{\Phi}_{(j)}'$$

With these, we can easily show

$$\boldsymbol{A}_{1:K}^{PC} \equiv \boldsymbol{\Xi}(\mathrm{E}(\mathbf{S}_{NT})|1:K) = N^{-1/2}\boldsymbol{\Phi};$$

$$\boldsymbol{F}_{1:K}^{PC} \equiv \boldsymbol{X}\boldsymbol{A}_{1:K}^{PC} = N^{1/2}\boldsymbol{F};$$

$$\boldsymbol{\delta}_{1:K}^{PC} \equiv \left[\mathrm{E}(\boldsymbol{F}_{1:K}^{PC\prime}\boldsymbol{F}_{1:K}^{PC})\right]^{-1}\mathrm{E}(\boldsymbol{F}_{1:K}^{PC\prime}\boldsymbol{y}) = N^{-1/2}\left[\mathrm{E}(T^{-1}\boldsymbol{F}'\boldsymbol{F})\right]^{-1}\mathrm{E}(T^{-1}\boldsymbol{F}'\boldsymbol{y}) = N^{-1/2}\boldsymbol{\beta};$$

$$\boldsymbol{f}_{1:K,T+1}^{PC} \equiv \boldsymbol{A}_{1:K}^{PC\prime}\boldsymbol{x}_{\boldsymbol{\cdot}T+1} = N^{1/2}\boldsymbol{f}_{\boldsymbol{\cdot}T+1}$$

By these results, the forecast for y_{T+2} obtained by using the population-versions of the first K PC factors, can be shown to equal the optimal forecast $\hat{y}_{T+2}^*$: $y_{T+2|K}^{PC} \equiv \boldsymbol{f}_{1:K,T+1}^{PC}\boldsymbol{\delta}_{1:K}^{PC} = \boldsymbol{f}_{\boldsymbol{\cdot}T+1}'\boldsymbol{\beta} = \hat{y}_{T+2}^*$. The optimal number of the PC factors for forecasting y_{T+2} is K (the total number of the common factors in $\boldsymbol{f}_{\boldsymbol{\cdot}t}$).

We now consider the population version of the PLS regression using the first J PLS factors. Let

$$\boldsymbol{G}_0^* = (\boldsymbol{F}_{(1)}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{F}_{(J)}\boldsymbol{\beta}_{(J)}); \quad \bar{\boldsymbol{D}}_0^* = \begin{pmatrix} \sigma_1^2 & \sigma_1^4 & ... & \sigma_1^{2J} \\ \sigma_2^2 & \sigma_2^4 & ... & \sigma_2^{2J} \\ \vdots & \vdots & & \vdots \\ \sigma_J^2 & \sigma_J^4 & ... & \sigma_J^{2J} \end{pmatrix}$$

Observe that $\bar{\boldsymbol{D}}_0^*$ is a square Vandermonde matrix which is invertible because all of the σ_j^2 are distinct. Under PA,

$$\boldsymbol{\alpha}_q^{PLS} \equiv [\mathrm{E}(\mathbf{S}_{NT})]^{q-1}\mathrm{E}(\boldsymbol{b}_{NT}) = N^{-3/2}\Sigma_{j=1}^{J}\sigma_j^{2q}\boldsymbol{\Phi}_{(j)}\boldsymbol{\beta}_{(j)};$$

$$\boldsymbol{A}_{1:J}^{PLS} \equiv (\boldsymbol{\alpha}_1^{PLS}, ..., \boldsymbol{\alpha}_J^{PLS}) = N^{-3/2}(\boldsymbol{\Phi}_{(1)}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{\Phi}_{(J)}\boldsymbol{\beta}_{(J)})\boldsymbol{D}_0^*(R);$$

$$\boldsymbol{P}_{1:J}^{PLS} \equiv \boldsymbol{X}\boldsymbol{A}_{1:J}^{PLS} = N^{-1/2}\boldsymbol{G}^*\boldsymbol{D}_0^*$$

It can be also shown that $\boldsymbol{\delta}_{1:J}^{PLS} \equiv [\mathrm{E}(\boldsymbol{P}_{1:J}^{PLS\prime}\boldsymbol{P}_{1:J}^{PLS})]^{-1}\mathrm{E}(\boldsymbol{P}_{1:J}^{PLS\prime}\boldsymbol{y}) = N^{-1/2}[\boldsymbol{D}_0^*]^{-1}\mathbf{1}_J$, where $\mathbf{1}_J$ is the $J \times 1$ vector of ones. With these results, we can show that the forecast for $y_{T+2}$with the population versions of the first J PLS factors is

$$y_{T+2|J}^{PLS} \equiv \boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\boldsymbol{A}_{1:J}^{PLS\prime}\boldsymbol{\delta}_{1:J}^{PLS} = \Sigma_{j=1}^{J}\boldsymbol{f}'_{(j)}\boldsymbol{\beta}_{(j)} = \hat{y}_{T+2}^*$$

The optimal number of the PLS factors for forecasting y_{T+2} is J, which is the number of the distinct factor variances. Thus, unless all the factors in $\boldsymbol{f}_{\boldsymbol{\cdot}t}$ have distinct variances, the forecasting by the PLS method requires a smaller number of factors than the forecasting by the PC method. For an extreme case where all factor variances are the same, using the first PLS factor is sufficient for optimal forecasting.

Even for more general cases in which the predictor variables x_{it} contain idiosyncratic components, the results obtained under PA asymptotically hold if the error groups $\{u_{t+1}\}$ and $\{e_{it}\}$ are independent. Kelly and Pruitt (2015) consider the asymptotic properties of the PLS factors under this assumption and two additional assumptions: all factor variances are distinct ($k(j) = 1$ for all $j = 1, ..., J$) and some of the factors $\boldsymbol{f}_{(j)t}$ are uncorrelated with y_{t+1} (i.e., $\boldsymbol{\beta}_{(j)} = \mathbf{0}_{k(j)\times 1}$ for some j). Under these assumptions, the asymptotically optimal number of the PLS factors for forecasting y_{T+2} equals the number of the factor vectors $\boldsymbol{f}_{(j)t}$ that are correlated with y_{t+1}.

Our study has two novelties compared to Kelly and Pruitt (2015). The first is that we allow some factors to have the same variances. The second is that we investigate the properties of the forecasting results obtained using more than the optimal number of PLS factors used.

Groen and Kapetanios (2016) consider an alternative model in which the predictor variables x_{it} are directly correlated with y_{t+1}, not indirectly through the factors $\boldsymbol{f}_{\cdot t}$. Specifically, they consider a model that consists of equation (1.4) and a forecast model $y_{t+1} = \boldsymbol{x}'_{\cdot t}\boldsymbol{\beta}^x + u_{t+1}$, where $\boldsymbol{\beta}^x$ is an $N \times 1$ coefficient vector. For this case, $\hat{y}^*_{T+2} = \boldsymbol{f}'_{\cdot T+1}\boldsymbol{\beta}$ is no longer optimal forecast even if both $\boldsymbol{f}_{\cdot T+1}$ and $\boldsymbol{\beta}$ are known. With some restrictive assumptions on $\boldsymbol{E}$ and $\boldsymbol{\beta}^x$, Groen and Kapetanios (2016) show that the PLS regression could generate more accurate forecasting results than the PC regression. For the model given in equations (1.4) and (1.5), their finding suggests that the PLS regression could be a powerful forecasting method, particularly when the idiosyncratic components of x_{it} (e_{it}) are correlated with y_{t+1}. For our asymptotic analysis we do not consider such cases. However, it is an interesting case that idiosyncratic components of some predictor variables are correlated with y_{t+1}. Thus, we consider some of such cases in our simulation experiments.

1.2.2 Assumptions

In this subsection, we make formal assumptions for our asymptotic analysis and state the main results. Let $m = \min\{N, T\}$; $M = \max\{N, T\}$; and let η denote a generic positive constant. All of the asymptotic assumptions are made for the cases in which as $m \to \infty$.

Assumption 1 (A.1): (i) The variable sets, $\{\boldsymbol{f}_{\cdot t}\}$, $\{\boldsymbol{\phi}_{\cdot i}\}$, $\{e_{it}\}$, and $\{u_{t+1}\}$ are mutually independent, while the variables within each group could be correlated. (ii) The variables in the 4 groups have finite moments at least up to the 4th order. (iii) $\mathrm{E}(\boldsymbol{f}_{\cdot t}) = \mathbf{0}_{K\times 1}$, $\mathrm{E}(e_{it}) = 0$, and $\mathrm{E}(u_{t+1}) = 0$, for all i and t.

Assumption 2 (A.2): For $j,\ j' = 1, ..., J$ and $j \neq j'$, $T^{-1}\boldsymbol{F'}_{(j)}\boldsymbol{F}_{(j)} \to_p \sigma_j^2\boldsymbol{I}_{k(j)}$ and $T^{-1}\boldsymbol{F'}_{(j)}\boldsymbol{F}_{(j')} \to_p \mathbf{0}_{k(j)\times k(j')}$, where $\sigma_1^2 > \sigma_2^2 > ... > \sigma_J^2 > 0$, $ks(j) = \Sigma_{h=1}^{j}k(h)$, and $K = ks(J)$. That is,

$$\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}} = T^{-1}\boldsymbol{F'F} \to_p \boldsymbol{\Omega}_{\boldsymbol{F}} = \boldsymbol{diag}(\sigma_1^2\boldsymbol{I}_{k(1)}, ..., \sigma_J^2\boldsymbol{I}_{k(J)}).$$

Assumption 3 (A.3): For $j,\ j' = 1, ..., J$ and $j' \neq j$, $N^{-1}\boldsymbol{\Phi'}_{(j)}\boldsymbol{\Phi}_{(j)} \to_p \boldsymbol{I}_{K(j)}$, $N^{-1}\boldsymbol{\Phi'}_{(j)}\boldsymbol{\Phi}_{(j')} \to_p \mathbf{0}_{k(j)\times k(j')}$. That is, $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}} = N^{-1}\boldsymbol{\Phi'\Phi} \to_p \boldsymbol{I}_K$.

Assumption 4 (A.4): For some real number $\gamma \in (0, 1/2]$, $T^{\gamma}\left(\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}} - \boldsymbol{\Omega}_{\boldsymbol{F}}\right) \to_d \boldsymbol{W}_{\boldsymbol{F}}$ and $N^{\gamma}\left(\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}} - \boldsymbol{I}_K\right) \to_d \boldsymbol{W}_{\boldsymbol{\Phi}}$, where $\boldsymbol{W}_{\boldsymbol{F}}$ and $\boldsymbol{W}_{\boldsymbol{\Phi}}$ are some matrices of real or rational random variables.

Assumption 5 (A.5): (i) For all t and N, $\mathrm{E}\left(N^{-1}\boldsymbol{e'}_{\boldsymbol{.}t}\boldsymbol{e}_{\boldsymbol{.}t}\right) < \eta$. (ii)$\lambda_1(\boldsymbol{E'E}/M) = O_p(1)$. (iii) There exists an increasing integer function of m, m^c, such that $0 < \lim_{m\to\infty}m^c/m < 1$ and $\lambda_{m^c}\left(\boldsymbol{E'E}/M\right) \geq \eta + o_p(1)$.

Assumption 6 (A.6): For all i, t, N and T, $\mathrm{E}\left(\left\|T^{-1}\Sigma_{t=1}^{T}\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right\|_2^2\right) < \eta$ and $\mathrm{E}\left(\left\|N^{-1/2}\Sigma_{i=1}^{N}\boldsymbol{\phi}_{\boldsymbol{.}i}e_{it}\right\|_2^2\right) < \eta$.

Assumption 7 (A.7): (i) For all T, $\lambda_1(\mathrm{E}(\boldsymbol{uu'})) < \eta$. (ii) For all N and T, $\mathrm{E}\left(\left\|T^{-1/2}\boldsymbol{F'u}\right\|_2^2\right) < \eta$, and $\mathrm{E}\left(\left\|(NT)^{-1/2}\boldsymbol{E'u}\right\|_2^2\right) < \eta$. (iii) $\hat{\sigma}_u^2 = \boldsymbol{u'u}/T \to_d \sigma_u^2 \in (0, \infty)$.

Assumption 8 (A.8): $\boldsymbol{\beta}_{(j)} = \mathbf{0}_{k(j)\times 1}$ for $j = R+1, ..., J$.

Some comments follow on (A.1) – (A.8). The part (i) of (A.1) rules out the possibility that the idiosyncratic errors in the x_{it} are correlated with the error term in the target variable y_{t+1}. The predictor variables x_{it} are correlated with the target variable y_{t+1} only through the factors $\boldsymbol{f}_{\boldsymbol{.}t}$. Some of the assumptions of independence among the variable groups could be relaxed for our asymptotic analysis. For example, we may allow some weak dependence between $\{\boldsymbol{f}_{\boldsymbol{.}t}\}$ and $\{e_{it}\}$ as long as (A.6) holds. As discussed in the previous subsection, the zero-mean assumption on the $\boldsymbol{f}_{\boldsymbol{.}t}$ in (A.1) is made to save notation.

Assumptions (A.2) and (A.3) are the normalization restrictions that are frequently used for factor model; see, for example, Stock and Watson (2002a). As discussed in the previous subsection, the assumptions are not restrictive ones. Onatski (2012) have considered the factor models with an alternative assumption of $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$ instead of (A.3). He refers as "weak" factors to those whose factor loadings satisfy this alternative assumption and as "strong" factors to those whose factor loadings satisfy (A.3). In this paper we only consider strong factors, leaving up the analysis of the cases with weak factors to a future study.

(A.4) implies that $\hat{\boldsymbol{\Omega}}_F$ and $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}$ are T^{γ}-consistent and N^{γ}-consistent estimators of $\boldsymbol{\Omega}_{\boldsymbol{F}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\Phi}}$, respectively, while the elements in $\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}$and $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}$ need not be normal. It would be reasonable to assume that $\gamma = 0.5$ for (A.4). In fact, restricting γ to be 0.5 does not alter our main asymptotic results. However, using γ instead of 0.5, we can observe what parts of our asymptotic results are affected by (A.4).

Under (A.4), the eigenvalues of $\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}$ and $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}$ could be also T^{γ}-consistent and N^{γ}-consistent for the eigenvalues of $\boldsymbol{\Omega}_{\boldsymbol{F}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\Phi}}$, respectively. For example, Anderson *et al.* (1963) has shown that the eigenvalues of $\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}$ are $T^{1/2}$-consistent if the $\boldsymbol{f}_{\boldsymbol{.}t}$ are *iid* multivariate normal vectors. In fact, the eigenvalues of $\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}$ are $T^{1/2}$-consistent even if the $\boldsymbol{f}_{\boldsymbol{.}t}$ are not normal; See Fang and Krishnaiah (1982). It is too restrictive to assume that $\{\boldsymbol{f}_{\bullet t}\}$ is an *iid* process. Taniguchi and Krishnaiah (1987) have shown that the eigenvalues of $\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}$ are $T^{1/2}$-consistent if $\{\boldsymbol{f}_{\bullet t}\}$ is a Gaussian stationary process. More general results related to the asymptotic distributions of the eigenvalues of sample variance matrices can be found from Eaton and Tyler (1991).

The parts (i) and (ii) of (A.5) can hold even if the idiosyncratic errors e_{it} are cross-sectionally and/or serially correlated. Some sufficient conditions for (ii) can be found from Ahn and Horenstein (2013), and Moon and Weidner (2015). Roughly speaking, the parts (i) and (ii) hold unless too strong cross sectional or serial correlations exist among the errors e_{it} as in the cases in which the errors contain some common factors. The part (iii) of (A.5) means that an asymptotically non-negligible number of the eigenvalues of $M^{-1}\boldsymbol{E'E}$ are bounded away from zero as $m \to \infty$. The condition holds unless the common factors $\boldsymbol{f}_{\bullet t}$ can explain most of the predictors perfectly; see Ahn and Horenstein (2013). Under (iii) of (A.5),

$$\Sigma_{h=1}^{m}\lambda_h((NT)^{-1}\boldsymbol{E'E}) \geq (m^c/m)(c + o_p(1)) > 0.$$

Sufficient conditions for (A.6) are the following:

$$N^{-1/2}\Sigma_{i=1}^{N}\boldsymbol{\phi}_{\bullet i}e_{it} \to_d N(\mathbf{0}_{K\times 1}, \boldsymbol{\Gamma}_t) \text{ as } N \to \infty, \text{ for each t} \tag{1.15}$$

$$T^{-1/2}\Sigma_{t=1}^{T}\boldsymbol{f}_{\bullet t}e_{it} \to_d N(\mathbf{0}_K, \boldsymbol{\Gamma}_i) \text{ as } T \to \infty, \text{ for each i} \tag{1.16}$$

where $\boldsymbol{\Gamma}_i = \lim_{T\to\infty}T^{-1}\Sigma_{t=1}^{T}\Sigma_{t'=1}^{T}\mathrm{E}(\boldsymbol{f}_{\bullet t}\boldsymbol{f}'_{\bullet t'}e_{it}e_{it'})$ and $\boldsymbol{\Gamma}_t = \lim_{N\to\infty}N^{-1}\Sigma_{i=1}^{N}\Sigma_{i'=1}^{N}$ $\mathrm{E}(\boldsymbol{\phi}_{\bullet i}\boldsymbol{\phi}'_{\bullet i'}e_{it}e_{i't})$. Assuming (1.14), Bai (2003) and Bai and Ng (2006) have derived the asymptotic distributions of the principal component factors and factor loadings. Imagine that the factor loading matrix $\boldsymbol{\Phi}$ is observable. For such cases, the factor vector $\boldsymbol{f}_{\bullet t}$ can be consistently estimated by the OLS regression of $\boldsymbol{x}_{\bullet t}$ on $\boldsymbol{\Phi}$. The conditions (A.2) and (1.15) are the sufficient conditions under which the resulting OLS estimators are asymptotically normal. Similarly, for the cases in which the factor matrix $\boldsymbol{F}$ is observable, the conditions (A.2) and (1.15) are the sufficient conditions under which the OLS estimators of $\phi_{\cdot i}$ obtained by regressing $\boldsymbol{x}_{i\cdot} = (x_{i1}, ..., x_{iT})'$ on $\boldsymbol{F}$ are all consistent and asymptotically normal.

In fact, (A.6) is stronger than what is needed for our asymptotic result. The weaker conditions that are sufficient for our results are $\left\|(NT)^{-1/2}\boldsymbol{F'E}\right\| = O_p(1)$, $\left\|(NT)^{-1/2}\boldsymbol{\Phi' E'}\right\| = O_p(1)$, and $\left\|(NT)^{-1/2}\boldsymbol{\Phi' E' F}\right\| = O_p(1)$. It is shown in Appendix (Lemma C.3) that these conditions hold under (A.6) and (A.1). Part (i) of (A.7) holds if the error terms u_{t+1} are not too strongly autocorrelated.

Under (A.7) and (A.8), the optimal forecast for y_{T+2} is $\hat{y}^o_{T+2} = \Sigma^R_{j=1}\boldsymbol{f}'_{(j)T+1}\boldsymbol{\beta}_{(j)}$. Strictly speaking, $\hat{y}^o_{T+2}$ is not optimal unless $\mathrm{E}(u_{t+1}|u_t, u_{t-1}, ..., u_1) = 0$ and $\mathrm{E}(u^2_{t+1}|u_t, ..., u_1) = \sigma^2_u$. However, for expository convenience, we refer to $\hat{y}^o_{T+2}$ as the optimal forecast.

(A.8) assumes that only the factors with larger variances are correlated with the target variable y_{t+1}, and that the other factors with smaller variances have no forecasting power. This assumption is just for expository convenience. The condition we need for our analytical results is that R groups of the factors are correlated with the target variable, while the other $(J - R)$ groups are not. Kelly and Pruitt (2015) have considered the cases in which $k(j) = 1$ for all $j = 1, ..., R$ (i.e., the first R strongest factors have distinct asymptotic variances). Similar to (iii) of PA in the previous subsection, (A.8) allows some factors to have the same asymptotic variances. For each $j \leq R$, not all factors in $\boldsymbol{f}_{(j)t}$ need to be correlated with y_{t+1}. Only a proper subset of the factors may be correlated with y_{t+1}.

1.2.3 Spurious Correlation between PLS Factors and Target Variable

One problem in using the PLS factors for forecasting is that if more than the first R PLS factors are used, the added PLS factors could be spuriously correlated with the target variable: they have in-sample explanatory power for the target variable, while they deteriorate the forecasting power of the regression with them.

To see why, we here consider an extreme case in which no common factors exist in the predictor variables x_{it} so that $K = J = 0$ and $\boldsymbol{X} = \boldsymbol{E}$. For this case, the predictor variables x_{it} have no power to forecast y_{t+1}. Nonetheless, the first PLS factor is positively correlated with the target variable even asymptotically. Observe that

$$\tilde{\boldsymbol{\alpha}}_1^{PLS} = \frac{1}{T^{1/2}}\frac{\boldsymbol{E'u}}{(NT)^{1/2}}; \qquad \frac{\tilde{\boldsymbol{p}}_1^{PLS}}{(NT)^{1/2}} = \frac{1}{T^{1/2}}\frac{\boldsymbol{E}}{N^{1/2}T^{1/2}}\boldsymbol{c}_L; \qquad \frac{\boldsymbol{y'}\tilde{\boldsymbol{p}}_1^{PLS}}{N^{1/2}T} = \frac{1}{T}\boldsymbol{c}_L'\boldsymbol{c}_L$$

where $\|\boldsymbol{c}_L\|_2 \equiv \left\|(NT)^{-1/2}\boldsymbol{E'u}\right\|_2 = O_p(1)$ by (A.7). In addition,

$$\frac{\tilde{\boldsymbol{p}}_1^{PLS\prime}\tilde{\boldsymbol{p}}_1^{PLS}}{NT} = \frac{1}{Tm}\boldsymbol{c}_L^{*\prime}\Lambda_L^*\boldsymbol{c}_L^* \leq \frac{1}{Tm}\lambda_1^*\boldsymbol{c}_L^{*\prime}\boldsymbol{c}_L^* \leq \frac{1}{Tm}\lambda_1^*\boldsymbol{c}_L'\boldsymbol{c}_L$$

where $\Xi_L^* = \Xi(\boldsymbol{E'E}/M|1:N)$, $\Lambda_L^* = \Lambda(\boldsymbol{E'E}/M|1:N)$, $\boldsymbol{c}_L^* = \Xi_L^{*\prime}\boldsymbol{c}_L$, $\lambda_1^* = \lambda_1(\boldsymbol{E'E}/M)$, and the last inequality is by the fact that $\boldsymbol{c}_L^{*\prime}\boldsymbol{c}_L^* = \boldsymbol{c}_L'\mathcal{P}(\Xi_L^*)\boldsymbol{c}_L \leq \boldsymbol{c}_L'\boldsymbol{c}_L$. Then, the R^2 from the regression of $\boldsymbol{y}$ on $\tilde{\boldsymbol{p}}_1^{PLS}$ yields

$$R^2_{PLS,1} \equiv \frac{\boldsymbol{y'}\mathcal{P}(\tilde{\boldsymbol{p}}_1^{PLS})\boldsymbol{y}/T}{\boldsymbol{y'y}/T} = \frac{m}{T}\frac{1}{\hat{\sigma}_u^2}\frac{(\boldsymbol{c}_L'\boldsymbol{c}_L)^2}{\boldsymbol{c}_L^{*\prime}\Lambda_L^*\boldsymbol{c}_L^*} \geq \frac{m}{T}\frac{1}{\hat{\sigma}_u^2}\frac{\boldsymbol{c}_L'\boldsymbol{c}_L}{\lambda_1^*} > 0$$

where $\lambda_1^* > 0$ by (A.5). If $m/T \to 0$, that is, if T is dominantly larger than N, then, $R^2_{PLS,1} \to_p 0$. However, if $m/T = O(1)$, that is, if neither of T and N is dominantly larger than the other, $R^2_{PLS,1}$ is asymptotically positive because $\boldsymbol{c}_L'\boldsymbol{c}_L$ and λ_1^* are positive by (A.5) and (A.7). This indicates that the PLS factor $\tilde{\boldsymbol{p}}_1^{PLS}$ and the target vector $\boldsymbol{y}$ are "spuriously" correlated unless T is dominantly larger than N.

The spurious correlation problem may also produce poor forecasting outcome. Notice that

$$\tilde{\boldsymbol{\delta}}_{1:1}^{PLS} = \frac{\tilde{\boldsymbol{p}}_1^{PLS\prime}\boldsymbol{y}/(NT)}{\tilde{\boldsymbol{p}}_1^{PLS\prime}\tilde{\boldsymbol{p}}_1^{PLS}/(NT)} = \frac{\boldsymbol{c}_L'\boldsymbol{c}_L/(TN^{1/2})}{\boldsymbol{c}_L^{*\prime}\Lambda_L^*\boldsymbol{c}_L^*/(Tm)} = \frac{m}{N^{1/2}}\frac{\boldsymbol{c}_L'\boldsymbol{c}_L}{\boldsymbol{c}_L^{*\prime}\Lambda_L^*\boldsymbol{c}_L^*}$$

Thus, we have

$$\tilde{y}^{PLS}_{T+2|1} = \boldsymbol{x}'_{\cdot T+1}\tilde{\boldsymbol{\alpha}}^{PLS}_{1}\tilde{\boldsymbol{\delta}}^{PLS}_{1:1} = \frac{m^{1/2}}{T^{1/2}}\frac{\boldsymbol{c}'_L\boldsymbol{c}_L}{\boldsymbol{c}^{*\prime}_L\Lambda^*_L\boldsymbol{c}^*_L}\frac{\boldsymbol{e}'_{\cdot T+1}}{N^{1/2}}\frac{\boldsymbol{E}'\boldsymbol{u}}{M^{1/2}}$$

Using the fact that $\boldsymbol{e}'_{\cdot T+1}\boldsymbol{E}'\boldsymbol{u}$ is a scalar, (A.5) and (A.7), we can also obtain

$$\mathrm{E}\left(\left\|\frac{\boldsymbol{e}'_{\cdot T+1}\boldsymbol{E}'\boldsymbol{u}}{N^{1/2}M^{1/2}}\right\|_2^2\right) = O(1)$$

because

$$\mathrm{E}\left(\left\|\frac{\boldsymbol{e}'_{\cdot T+1}\boldsymbol{E}'\boldsymbol{u}}{N^{1/2}M^{1/2}}\right\|_2^2 \middle| \boldsymbol{E}, \boldsymbol{e}_{\cdot T+1}\right) \leq \left\|\frac{\boldsymbol{e}'_{\cdot T+1}}{N^{1/2}}\right\|_2^2\left\|\frac{\boldsymbol{E}}{M^{1/2}}\right\|_2^2\left\|\mathrm{E}(\boldsymbol{u}\boldsymbol{u}')\right\|_2 = O_p(1).$$

These results indicate that $\left\|\tilde{y}^{PLS}_{T+2|1} - \hat{y}^{o}_{T+2}\right\| = \left\|\tilde{y}^{PLS}_{T+2|1}\right\| = O_p((m/T)^{1/2})$, where $\hat{y}^{o}_{T+2} = 0$. Thus, when $N/T \to 0$, $\left|\tilde{y}^{PLS}_{T+2|1} - \hat{y}^{o}_{T+2}\right| \to_p 0$ as $m \to \infty$. In contrast, when $m/T = O(1) > 0$ (that is, when $m = T$ or when neither of N and T is dominantly larger than the other), $\tilde{y}^{PLS}_{T+2|1}$ is not a consistent estimator of $\hat{y}^{o}_{T+2}$.

While this example is a special case in which $K = 0$ and the first PLS factor is used for the prediction of y_{T+2}, it suggests that in general, the forecast for y_{T+2} obtained using more than R PLS factors may have poor asymptotic and finite-sample properties.

1.2.4 Main Results

This subsection reports our main asymptotic results. All of the results hold as $N, T \to \infty$ jointly. We need some notation to state our results. Set $ks(0) = 0$. For $j = 1, ..., J$, we define

$$\boldsymbol{\Lambda}^{\mathbf{S}_{NT}}_{(j)} = \boldsymbol{\Lambda}(\mathbf{S}_{NT}|ks(j-1)+1 : ks(j));$$

$$\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)} = \boldsymbol{\Xi}(\mathbf{S}_{NT}|ks(j-1)+1 : ks(j));$$

$$\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} = \boldsymbol{\Xi}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{b}_{NT}$$

Here, $\mathbf{\Lambda}_{(j)}^{\mathbf{S}_{NT}}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\mathbf{S}_{NT}$ which converge to σ_j^2, the jth largest largest asymptotic factor variance. The matrix $\mathbf{\Xi}_{(j)}^{\mathbf{S}_{NT}}$ is the matrix of the eigenvectors corresponding to the eigenvalues in $\mathbf{\Lambda}_{(j)}^{\mathbf{S}_{NT}}$.

Similarly, we also define

$$\mathbf{\Lambda}_{L}^{\mathbf{S}_{NT}} = \mathbf{\Lambda}(\mathbf{S}_{NT}|K+1:m); \quad \mathbf{\Xi}_{L}^{\mathbf{S}_{NT}} = \mathbf{\Xi}(\mathbf{S}_{NT}|K+1:m); \quad \boldsymbol{c}_{L}^{\mathbf{S}_{NT}} = \mathbf{\Xi}_{L}^{\mathbf{S}_{NT}\prime}\boldsymbol{b}_{NT}.$$

The matrix $\mathbf{\Lambda}_{L}^{\mathbf{S}_{NT}}$ is a diagonal matrix that contains the rest of the eigenvalues of $\mathbf{S}_{NT}$ other than the first K largest ones. The matrix $\mathbf{\Xi}_{L}^{\mathbf{S}_{NT}}$ is the matrix of the eigenvectors corresponding to the eigenvalues in $\mathbf{\Lambda}_{L}^{\mathbf{S}_{NT}}$. A technical point is worth noting related to $\mathbf{\Xi}_{L}^{\mathbf{S}_{NT}}$ and $\mathbf{\Lambda}_{L}^{\mathbf{S}_{NT}}$. When $N > T$, for all integers $h > T$, $\lambda_h(\mathbf{S}_{NT}) = 0$, which in turn implies $(NT)^{-1/2}\boldsymbol{X}\xi_h(\mathbf{S}_{NT}) = \mathbf{0}_{T\times 1}$. For this result, we can have

$$\mathbf{\Xi}(\mathbf{S}_{NT}|K+1:N)[\mathbf{\Lambda}(\mathbf{S}_{NT}|K+1:N)]^{q-1}\mathbf{\Xi}(\mathbf{S}_{NT}|K+1:N)' = \mathbf{\Xi}_{L}^{\mathbf{S}_{NT}}(\mathbf{\Lambda}_{L}^{\mathbf{S}_{NT}})^{q-1}\mathbf{\Xi}_{L}^{\mathbf{S}_{NT}\prime}$$

for both cases with $N > T$ and $T \geq N$.

With the above notation and result, we can show that

$$\tilde{\boldsymbol{\alpha}}_q^{PLS} = (\mathbf{S}_{NT})^{q-1}\boldsymbol{b}_{NT} = \Sigma_{j=1}^{J}\mathbf{\Xi}_{(j)}^{\mathbf{S}_{NT}}(\mathbf{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} + \mathbf{\Xi}_{L}^{\mathbf{S}_{NT}}(\mathbf{\Lambda}_{L}^{\mathbf{S}_{NT}})^{q-1}\boldsymbol{c}_{L}^{\mathbf{S}_{NT}} \qquad (1.17)$$

Thus, the asymptotic property of the qth PLS coefficient vector $\tilde{\boldsymbol{\alpha}}_q^{PLS}$ depends on those of the eigenvalues and eigenvectors of the matrix $\mathbf{S}_{NT}$, the vector $\boldsymbol{b}_{NT}$, and the vectors $\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}$ and $\boldsymbol{c}_{L}^{\mathbf{S}_{NT}}$. The asymptotic properties of these terms are given in the following Lemma.

Lemma 2.4.1: Under (A.1) – (A.8), the following holds.

(i) $\lambda_h(\mathbf{S}_{NT}) = \sigma_j^2 + O_p(m^{-\gamma})$, for $h = ks(j-1)+1, ..., ks(j)$ and $j = 1, ..., J$.

(ii) $\lambda_h(\mathbf{S}_{NT}) = O_p(m^{-1})$, for $h = K+1, K+2, ..., m$.

(iii) $\left\|\boldsymbol{b}_{NT} - \Sigma_{j=1}^{R}\sigma_j^2 N^{-1/2}\boldsymbol{\Phi}_{(j)}\boldsymbol{\beta}_{(j)}\right\|_2 = O_p(T^{-\gamma})$.

For each $j = 1, ..., R$, there exists some orthonormal matrix $\boldsymbol{O}_{jj}^{*}$ such that

(iv) $\left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - N^{-1/2}\boldsymbol{\Phi}_{(j)}\boldsymbol{O}_{jj}^{*}\right\|_F = O_p(m^{-\gamma})$, for $j = 1, ..., J$;

(v) $\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{O}_{jj}^{*\prime}\sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_2 = O_p(m^{-\gamma})$, for $j = 1, ..., R$.

For $j = R + 1, ..., J$,

(vi) $\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right\| = O_p(m^{-\gamma})$.

Let $\boldsymbol{H}_{NT} = (NT)^{-1/2}\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime}\mathcal{Q}(\boldsymbol{\Phi})\boldsymbol{E}'\mathcal{Q}(\tilde{\boldsymbol{F}})$ and $\tilde{\boldsymbol{F}} = \boldsymbol{F} + \boldsymbol{E}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}$. Let $\boldsymbol{r}_{NT}$ be an $m \times 1$ random vector with $\mathrm{E}\left(\|\boldsymbol{r}_{NT}\|_2\right) = O_p(1)$ which is independent of $\boldsymbol{u}$. Then,

(vii) $\left\|\boldsymbol{c}_L^{\mathbf{S}_{NT}} - T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u}\right\|_2 = O_p(m^{-3/2})$;

(viii) $\left\|T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u}\right\|_2 = O_p(T^{-1/2})$.

(ix) $\left\|T^{-1/2}\boldsymbol{r}_{NT}'\boldsymbol{H}_{NT}\boldsymbol{u}\right\|_2 = O_p((Tm)^{-1/2})$.

Some remarks follow on (vii) – (ix) of Lemma 2.4.1. First, (vii) and (viii) of Lemma 2.4.1 imply that $\left\|\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2 = O_p(T^{-1/2} + m^{-3/2})$. Second, the convergency speed of $\boldsymbol{c}_L^{\mathbf{S}_{NT}}$ depends on the term $T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u}$, which is a function of the error terms in $\boldsymbol{E}$ and $\boldsymbol{u}$. As it turns out later, the term $T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u}$ is the major source of the spurious correlation problem discussed in the previous subsection. While individual error terms in $\boldsymbol{e}_{\boldsymbol{\cdot}t}$ are uncorrelated with the error u_{t+1}, linear combinations of the N error terms in $\boldsymbol{e}_{\boldsymbol{\cdot}t}$ could appear to be spuriously correlated with u_{t+1} when N is large. An intuition on this result is that for a regression estimation, using more regressors for a dependent variable increases the R-square measure even if the regressors have no explanatory power.

We now consider the properties of the PLS coefficient vectors $\tilde{\boldsymbol{\alpha}}_q^{PLS}$. In order to make our asymptotic analysis easier, we need to modify equation (1.17). Define

$$\begin{aligned}
&\mu_j^{\mathbf{S}_{NT}} = \lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}}, \text{ for } j = 1, ..., R;\\
&\boldsymbol{d}_0(q) = ((\mu_1^{\mathbf{S}_{NT}})^{q-1}, (\mu_2^{\mathbf{S}_{NT}})^{q-1}, ..., (\mu_R^{\mathbf{S}_{NT}})^{q-1})';\\
&\boldsymbol{D}_0(q) = (\boldsymbol{d}_0(1), \boldsymbol{d}_0(1), ..., \boldsymbol{d}_0(1))
\end{aligned}$$

Notice that $\mu_j^{\mathbf{S}_{NT}}$is the largest one in the jth group of the eigenvalues, $\lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}}$, ··· , $\lambda_{ks(j)}^{\mathbf{S}_{NT}}$. Notice also that $\boldsymbol{D}_0(q)$ is a Vandermonde matrix. By construction, $\boldsymbol{D}_0(R)$ is a square matrix which is invertible because the $\mu_j^{\mathbf{S}_{NT}}(j = 1, ..., R)$ are all distinct even asymptotically. By Lemma 2.4.1, $\mu_j^{\mathbf{S}_{NT}} \to_p \sigma_j^2$ as $m \to \infty$.

With the terms defined above, we can easily show that

$$\tilde{\boldsymbol{\alpha}}_q^{PLS} = \boldsymbol{V}_0\boldsymbol{d}_0(q) + \boldsymbol{v}_{H1}(q) + \boldsymbol{v}_{H2}(q) + \boldsymbol{v}_L(q), \tag{1.18}$$

where

$$\begin{aligned}
&\boldsymbol{V}_0 = (\boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}, ..., \boldsymbol{\Xi}_{(R)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}});\\
&\boldsymbol{v}_{H1}(q) = \Sigma_{j=1}^{R}\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}}\left[(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\right]\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}};\\
&\boldsymbol{v}_{H2}(q) = \Sigma_{j=R+1}^{J}\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}}(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}};\\
&\boldsymbol{v}_L(q) = \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}}(\boldsymbol{\Lambda}_L^{S_{NT}})^{q-1}\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime}\boldsymbol{b}_{NT} = \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}}(\boldsymbol{\Lambda}_L^{S_{NT}})^{q-1}\boldsymbol{c}_L^{\mathbf{S}_{NT}}
\end{aligned}$$

where $\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}} = \mu_j^{\mathbf{S}_{NT}}\boldsymbol{I}_{k(j)}$. Thus,

$$\tilde{\boldsymbol{A}}_{1:q}^{PLS} = \boldsymbol{V}_0\boldsymbol{D}_0(q) + \boldsymbol{V}_{H1}(q) + \boldsymbol{V}_{H2}(q) + \boldsymbol{V}_L(q) \tag{1.19}$$

where $\boldsymbol{V}_{H1}(q) = (\boldsymbol{v}_{H1}(1), ..., \boldsymbol{v}_{H1}(q))$, and $\boldsymbol{V}_{H2}(q)$ and $\boldsymbol{V}_L(q)$are defined similarly.

The asymptotic property of each term in $\tilde{\boldsymbol{\alpha}}_q^{PLS}$ and $\tilde{\boldsymbol{A}}_{1:q}^{PLS}$ is stated in the following lemma and corollary. It is shown that $\boldsymbol{V}_0$ is the asymptotically dominant term in $\tilde{\boldsymbol{\alpha}}_q^{PLS}$.

Lemma 2.4.2: Define

$$\boldsymbol{\Pi}_{NT} = N^{-1/2}[\boldsymbol{\Phi}_{(1)}\boldsymbol{\beta}_{(1)}, \boldsymbol{\Phi}_{(2)}\boldsymbol{\beta}_{(2)}, ..., \boldsymbol{\Phi}_{(R)}\boldsymbol{\beta}_{(R)}];$$

$$\boldsymbol{\Sigma}_R = \boldsymbol{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_R^2)$$

Under (A.1) – (A.8), the following holds.

(i) $\|\boldsymbol{V}_0 - \boldsymbol{\Pi}_{NT}\boldsymbol{\Sigma}_R\|_F = O_p(m^{-\gamma})$;

(ii) $\|\boldsymbol{v}_{H1}(q)\|_2 = O_p(m^{-\gamma})$; $\|\boldsymbol{v}_{H2}(q)\|_2 = O_p(m^{-\gamma})$;

(iii) $\|\boldsymbol{v}_L(q)\|_2 = O_p\left(m^{-(q-1)}(T^{-1/2} + m^{-3/2})\right)$

Corollary 2.4.2: Under (A.1) – (A.8),

(i) $\|\boldsymbol{V}_0\boldsymbol{D}_0(q) - \boldsymbol{\Pi}_{NT}\boldsymbol{\Sigma}_R\boldsymbol{D}_0(q)\|_F = O_p(m^{-\gamma})$;

(ii) $\|\boldsymbol{V}_{H1}(q)\|_F = O_p(m^{-\gamma})$; $\|\boldsymbol{V}_{H2}(q)\|_F = O_p(m^{-\gamma})$;

(iii) $\|\boldsymbol{V}_L(q)\|_F = O_p\left(T^{-1/2} + m^{-3/2}\right)$

Lemma 2.4.2 and Corollary 2.4.2 imply our first main result. Stated formally:

Theorem 1: Define $\boldsymbol{g}_{T+1} = (\boldsymbol{f}'_{(1),T+1}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{f}'_{(R),T+1}\boldsymbol{\beta}_{(R)})'$. Under (A.1) – (A.8), for $q = 1, \ ... \ , R$,

(i) $\left\|\tilde{\boldsymbol{A}}_{1:q}^{PLS} - \boldsymbol{\Pi}_{NT}\boldsymbol{\Sigma}_R\boldsymbol{D}_0(q)\right\|_F = O_p(m^{-\gamma})$;

(ii) $\left\|N^{-1/2}\tilde{\boldsymbol{A}}_{1:q}^{PLS\prime}\boldsymbol{x}_{\boldsymbol{\cdot}T+1} - \boldsymbol{D}_0(q)'\boldsymbol{\Sigma}_R\boldsymbol{g}_{T+1}\right\|_F = O_p(m^{-\gamma})$.

The first part of Theorem 1 implies that $\tilde{\boldsymbol{A}}^{PLS}_{1:q}$ and $\boldsymbol{\Pi}_{NT}\boldsymbol{\Sigma}_R\boldsymbol{D}_0(q)$ span the same linear space asymptotically. When $q = R$, the matrix $\boldsymbol{D}_0(R)$ is invertible as we discussed above. Thus, $\tilde{\boldsymbol{A}}^{PLS}_{1:R}$ and $\boldsymbol{\Pi}_{NT}$ span the same space asymptotically. When $q = R$, the second part of Theorem 1 implies that $\left|\boldsymbol{1'}_R\boldsymbol{\Sigma}_R^{-1}[\boldsymbol{D}_0(R)\boldsymbol{'}]^{-1}N^{-1/2}\tilde{\boldsymbol{A}}^{PLS\boldsymbol{'}}_{1:q}\boldsymbol{x}_{\boldsymbol{\cdot}T+1} - \hat{y}^o_{T+2}\right|$ $= O_p(m^{-\gamma})$, because $\hat{y}^o_{T+2} = \boldsymbol{1'}_R\boldsymbol{g}_{T+1}$.

We now consider the asymptotic properties of the PLS factors. Define

$$\boldsymbol{G}_0 = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{V}_0;$$

$$\boldsymbol{g}_{H1}(q) = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{v}_{H1}(q); \quad \boldsymbol{g}_{H2}(q) = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{v}_{H2}(q)$$

$$\boldsymbol{g}_L(q) = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{v}_L(q)$$

With this notation, we have

$$(NT)^{-1/2}\tilde{\boldsymbol{p}}^{PLS}_q = (NT)^{-1/2}\boldsymbol{X}\tilde{\boldsymbol{\alpha}}^{PLS}_q = \boldsymbol{G}_0\boldsymbol{d}_0(q) + \boldsymbol{g}_{H1}(q) + \boldsymbol{g}_{H2}(q) + \boldsymbol{g}_L(q) \qquad (1.20)$$

Because $\boldsymbol{g}_{H1}(q) = \mathcal{P}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q) + \mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q)$, equation (1.20) is equivalent to

$$(NT)^{-1/2}\tilde{\boldsymbol{p}}^{PLS}_q = (NT)^{-1/2}\boldsymbol{X}\tilde{\boldsymbol{\alpha}}^{PLS}_q = \boldsymbol{G}_0\hat{\boldsymbol{d}}_0(q) + \boldsymbol{g}^c_H(q) + \boldsymbol{g}_L(q), \qquad (1.21)$$

where

$$\hat{\boldsymbol{d}}_0(q) = \boldsymbol{d}_0(q) + (\boldsymbol{G'}_0\boldsymbol{G}_0)^{-1}\boldsymbol{G'}_0\boldsymbol{g}_{H1}(q);$$

$$\boldsymbol{g}^c_H(q) = (\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q), \boldsymbol{g}_{H2}(q))$$

By (1.21), we also have

$$(NT)^{-1/2}\tilde{\boldsymbol{P}}^{PLS}_{1:q} = \boldsymbol{G}_0\hat{\boldsymbol{D}}_0(q) + \boldsymbol{G}^c_H(q) + \boldsymbol{G}_L(q) \qquad (1.22)$$

where

$$\hat{\boldsymbol{D}}_0(q) = (\hat{\boldsymbol{d}}_0(1), ..., \hat{\boldsymbol{d}}_0(q)) = \boldsymbol{D}_0(q) + (\boldsymbol{G'}_0\boldsymbol{G}_0)^{-1}\boldsymbol{G'}_0\boldsymbol{G}_{H1}(q);$$

$$\boldsymbol{G}^c_H(q) = \mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{G}_{H1}(q) + \boldsymbol{G}_{H2}(q)$$

and $\boldsymbol{G}_L(q)$ is similarly defined.

Two remarks follow on equation (1.22). First, by construction, the matrices $\boldsymbol{G}_0$, $\boldsymbol{G}_H^c$, and $\boldsymbol{G}_L$ are mutually orthogonal. This structure facilitates our asymptotic analysis. Second, we merge $\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{G}_{H1}(q)$ and $\boldsymbol{G}_{H2}(q)$ into $\boldsymbol{G}_H^c(q)$ because the Frobenius norms of the two matrices are both $O_p(m^{-\gamma})$.

Consider the case in which $R = K$; that is, all of the factors $\boldsymbol{f}_{\boldsymbol{.}t}$ have distinct asymptotic variances and are correlated with the target variable y_{T+2}. For the case, $\boldsymbol{G}_H^c(q) = 0$ and $\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(q) = \boldsymbol{G}_0\boldsymbol{D}_0(q)$. Thus, the asymptotic property of $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$ depends on $\boldsymbol{G}_0\boldsymbol{D}_0(q)$ and $\boldsymbol{G}_L(q)$. In contrast, when $R < K$, that is, when $k(j) > 1$ for some $j = 1, ..., R$ and/or $R < J$, the asymptotic property of $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$ also depends on $\boldsymbol{G}_H^c(q)$.

The asymptotic properties of the terms that appear in the PLS factors are stated in the following lemma and corollary.

Lemma 2.4.3: Under (A.1) – (A.8), the following holds for $q \geq 1$.

(i) $\left\|\boldsymbol{G}_0 - T^{-1/2}(\boldsymbol{F}_{(1)}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{F}_{(R)}\boldsymbol{\beta}_{(R)})\boldsymbol{\Sigma}_R\right\|_F = O_p(m^{-\gamma})$;

(ii) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_0 - (\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{\beta}'_{(R)}\boldsymbol{\beta}_{(R)})\boldsymbol{\Sigma}_R^2\right\|_2 = O_p(m^{-\gamma})$;

(iii) $\|\boldsymbol{g}_H^c(q)\|_2 = O_p(m^{-\gamma})$; $\|\boldsymbol{g}_L(q)\|_2 = O_p\left(m^{-(q-1/2)}(T^{-1/2} + m^{-3/2})\right)$;

(iv) $\left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0(q)\right\|_2 = O_p(m^{-\gamma})$;

(v) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_H^c(q)\right\|_2 = O_p(m^{-2\gamma})$; $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(q)\right\|_2 = O_p\left(m^{-(q-1)}(T^{-1/2} + m^{-3/2})^2\right)$.

Corollary 2.4.3: Under (A.1) – (A.8), the following holds for $q \geq 1$.

(i) $\|\boldsymbol{G}_H^c(q)\|_F = O_p(m^{-\gamma})$; $\|\boldsymbol{G}_L(q)\|_F = O_p(m^{-1/2}(T^{-1/2} + m^{-3/2}))$; $\left\|\hat{\boldsymbol{D}}_0(q) - \boldsymbol{D}_0(q)\right\|_F = O_p(m^{-\gamma})$;

(ii) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_H^c(q)\right\|_2 = O_p(m^{-2\gamma})$; $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_L(q)\right\|_2 = O_p\left((T^{-1/2} + m^{-3/2})^2\right)$

Lemma 2.4.3 and Corollary 2.4.3 indicate that the asymptotically dominant term in $\tilde{\boldsymbol{P}}^{PLS}_{1:q}$ is $\boldsymbol{G}_0$. For $q \leq R$, the asymptotic properties of the q PLS factors in $\tilde{\boldsymbol{P}}^{PLS}_{1:R}$ are determined by $\boldsymbol{G}_0\boldsymbol{D}_0(q)$. Thus, we can obtain the following results.

Lemma 2.4.4: Assume that (A.1) – (A.8) hold. When $R < K$,

(i) $\left\|(NT)^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R} - \hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right\|_F = O_p(m^{-\gamma})$;

(ii) $\left\|N^{-1/2}T^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y} - \hat{\boldsymbol{D}}_0(R)'T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 = O_p(m^{-2\gamma})$.

When $R = K$,

(iii) $\left\|(NT)^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R} - \boldsymbol{D}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\boldsymbol{D}_0(R)\right\|_F = O_p\left(m^{-1}(T^{-1/2} + m^{-3/2})^2\right)$;

(iv) $\left\|N^{-1/2}T^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y} - \boldsymbol{D}_0(R)'T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 = O_p\left((T^{-1/2} + m^{-3/2})^2\right)$

With Lemma 2.4.4, we can obtain our second main result:

Theorem 2: Under (A.1) – (A.8),

(i) $\left\|N^{1/2}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2 = O_p(m^{-\gamma})$;

(ii) $\left\|\tilde{y}^{PLS}_{T+2|R} - \hat{y}^o_{T+2}\right\|_2 = O_p(m^{-\gamma})$;

(iii) $R^2_{1:R} \equiv \dfrac{\boldsymbol{y}'\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} \to_p R^2_{\max} \equiv \dfrac{\Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)}}{\Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)} + \sigma^2_u}$

Two remarks on Theorem 2 follow. First, the theorem indicates that the forecast for y_{T+2} obtained using the first R PLS factors, $\tilde{y}^{PLS}_{T+2|R}$, is a consistent estimator of the optimal forecast, $\hat{y}^o_{T+2} = \Sigma^R_{j=1}\boldsymbol{f}'_{(j)T+1}\boldsymbol{\beta}_{(j)}$. We can show that the forecast by a fewer number of PLS factor is not consistent for $\hat{y}^o_{T+2}$. Thus, the minimum number of the PLS factors that can produce a consistent estimator of $\hat{y}^o_{T+2}$ is R, the number of distinct asymptotic variances of the common factors in $\boldsymbol{f}_{\boldsymbol{.}t}$that are correlated y_{t+1}. For example, if all the factors have the same asymptotic variances, then the first PLS factor is sufficient to produce a consistent estimator of $\hat{y}^o_{T+2}$. Given this finding, we from now on refer to the R factors as "informative" PLS factors.

Second, in (iii) of Theorem 2, $R^2_{\max}$ is the probability limit of the in-sample R^2 from the regression of $\boldsymbol{y}$ on the $ks(R)$ unobservable common factors in $\boldsymbol{F}_{(1)}$, ... , $\boldsymbol{F}_{(R)}$. Interestingly, the result in (iii) of Theorem 2 indicates that the in-sample fit of the regression of $\boldsymbol{y}$ on R PLS factors is as good as that of the regression of $\boldsymbol{y}$ on $ks(R)$ relevant latent factors.

We now consider the forecasting with more than R PLS factors. Specifically, we consider the cases in which the first $(R+1)$ PLS factors are used. Observe that

$$\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}) = \mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R}) + \mathcal{P}(\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1})$$

This implies that the asymptotic properties of the forecast by the first $(R+1)$ PLS factors depend on $\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})$ and $\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1}$. More specifically, the asymptotic property of $\hat{y}^{PLS}_{T+2|R+1}$ depends on the following three terms:

$$\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R})^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1} \tag{1.23}$$

$$\mathcal{Y}_{1,NT} = \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1}/(NT) \tag{1.24}$$

$$\mathcal{Y}_{2,NT} = \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}/(N^{1/2}T) \tag{1.25}$$

The following Lemma states the asymptotic properties of $\mathcal{Y}_{1,NT}$ and $\mathcal{Y}_{2,NT}$:

Lemma 2.4.5: Assume that (A.1) – (A.8) hold. When $R < K$,

(i) $\left\|\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 = O_p(m^{-\gamma})$;

(ii) $\mathcal{Y}_{1,NT} = O_p(m^{-2\gamma})$; $\mathcal{Y}_{2,NT} = O_p(m^{-2\gamma})$.

When $R = K$,

(iii) $\left\|\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 = O_p\left(m^{-1}(T^{-1/2} + m^{-3/2})^2\right)$;

(iv) $\mathcal{Y}_{1,NT} = O_p\left(m^{-1}(T^{-1/2} + m^{-3/2})^2\right)$; $\mathcal{Y}_{2,NT} = O_p\left((T^{-1/2} + m^{-3/2})^2\right)$.

Some remarks on Lemma 2.4.5 follow. The R^2 from the regression of $\boldsymbol{y}$ on the first $(R+1)$ PLS factors $\tilde{\boldsymbol{P}}^{PLS}_{1:R}$ depends on both $\mathcal{Y}_{1,NT}$ and $\mathcal{Y}_{2,NT}$ because

$$\frac{\boldsymbol{y}'\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R+1})\boldsymbol{y}}{T} = \frac{\boldsymbol{y}'\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}}{T} + \frac{\boldsymbol{y}'\mathcal{P}(\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1})\boldsymbol{y}}{T} = \frac{\boldsymbol{y}'\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}}{T} + \frac{(\mathcal{Y}_{2,NT})^2}{\mathcal{Y}_{1,NT}}.$$

When $R < K$, $m^{2\gamma}\mathcal{Y}_{1,NT}$ and $m^{2\gamma}\mathcal{Y}_{2,NT}$ are positive random variables, while $(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = O_p(m^{-2\gamma}) =o_p(1)$. Thus, using the $(R+1)$th PLS factor additionally does not change the asymptotic goodness of fit of the PLS regression. In contrast, when $R = K$,

$$(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = O_p\left(m(T^{-1/2} + m^{-3/2})^2\right) = O_p(m/T)$$

If $m/T \to 0$ as $m \to \infty$, then $(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = o_p(1)$. Thus, once again, the asymptotic goodness of fit of the PLS regression is unaltered when the $(R+1)$th PLS factor is added. However, if $m/T = O(1) > 0$, the ratio $(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT}$ becomes a positive $O_p(1)$ variable, so that $R^2_{1:R+1} = R^2_{\max} + O_p(1) > R^2_{\max}$.

In short, when $R = K$ and $T$is not dominantly larger than N, use of the $\tilde{\boldsymbol{p}}^{PLS}_{R+1}$ in addition to $\tilde{\boldsymbol{P}}^{PLS}_{1:R}$ makes the part of the PLS factors spuriously correlated with the target variable asymptotically important. We state this result formally:

Theorem 3: Assume that (A.1) – (A.8) hold. When $R = K$ and $N/T = O(1) > 0$,

(i) $T^{-1}\boldsymbol{y}'\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R+1})\boldsymbol{y} = \Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)} + |O_p(1)| > \Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)}$.

When $R < K$, or when $R = K$ and $N/T \to 0$,

(ii) $T^{-1}\boldsymbol{y}'\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R+1})\boldsymbol{y} \to_p \Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)}$.

Theorem 3 is for the cases in which the $(R+1)$th PLS factor is additionally used. In fact, it could be shown that the PLS regressions using at least the first R and up to K PLS factors produce asymptotically the same R^2, $R^2_{\max}$. In contrast, use of more than K PLS factors may trigger the spurious correlation problem. However, this asymptotic result does not necessarily imply that using K PLS factors is a safe bet in case in which R is unknown. Our simulation results reported in section 1.3 indicates that use of more than R PLS factors often increases in-sample R^2 sharply while producing poor forecasting results.

Finally, we investigate the performance of the forecast for y_{T+2} obtained by using the first $(R+1)$ PLS factors. Observe that

$$\begin{aligned} N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R+1} &= \left(\frac{\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}{}'\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}{}'\boldsymbol{y}}{N^{1/2}T} \\ &= \begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}}{N^{1/2}T} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\boldsymbol{y}}{N^{1/2}T} \end{pmatrix} \end{aligned} \tag{1.26}$$

By the inversion rule for partitioned matrix

$$\begin{aligned} \begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \end{pmatrix}^{-1} &= \begin{pmatrix} \left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT}\right)^{-1} & \boldsymbol{0} \\ \boldsymbol{0}' & 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix} \left(\frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT}\right)^{-1} \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}' \end{aligned} \tag{1.27}$$

where $\tilde{\boldsymbol{\theta}}$ is defined in (1.23). In addition,

$$\begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}' \begin{pmatrix} \tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}/(N^{1/2}T) \\ \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\boldsymbol{y}/(N^{1/2}T) \end{pmatrix} = -\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}/(N^{1/2}T) \tag{1.28}$$

Substituting (1.27) and (1.28) into (1.26), we can obtain

$$N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R+1} = \begin{pmatrix} N^{1/2}\tilde{\boldsymbol{\delta}}'_{1:R} & 0 \end{pmatrix}' - \begin{pmatrix} \tilde{\boldsymbol{\theta}}' & -1 \end{pmatrix}' \mathcal{Y}_{NT} \tag{1.29}$$

where $\mathcal{Y}_{NT} = \mathcal{Y}_{1,NT}/\mathcal{Y}_{2,NT}$. Using this result and Lemmas 2.4.5, we can obtain our final main result.

Theorem 4: Under (A.1) – (A.8), the following holds.

(i) $\left\|\tilde{y}^{PLS}_{T+2|R} - \hat{y}^{o}_{T+2}\right\| = O_p(m^{-\gamma})$, if $R < K$.

(ii) $\left\|\tilde{y}^{PLS}_{T+2|R} - \hat{y}^{o}_{T+2}\right\| = o_p(1)$, if $R = K$ and $N/T = o(1)$;

(iii) $\left\|\tilde{y}^{PLS}_{T+2|R+1} - \hat{y}^{o}_{T+2}\right\| = O_p(1)$, if $R = K$ and either $N \geq T$ or $N/T = O(1) > 0$.

Some remarks follow on Theorem 4. First, the asymptotic property of $\tilde{\boldsymbol{\delta}}^{PLS}_{1:R+1}$ (the OLS estimator from the regression of $\boldsymbol{y}$ on $\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}$) depends on $\mathcal{Y}_{NT}$. When $R < K$, both $m^{2\gamma}\mathcal{Y}_{1,NT}$ and $m^{2\gamma}\mathcal{Y}_{2,NT}$ are positive $O_p(1)$ variables and $\mathcal{Y}_{NT} = O_p(1)$. For this case, $N^{1/2}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R+1}$ is asymptotically a random variable. In addition, it can be shown that $m^{\gamma}(\tilde{y}^{PLS}_{T+2|R+1} - \hat{y}^{o}_{T+2})$ depends on $\mathcal{Y}_{NT}$, whose mean is not zero. That is, the forecasts $\tilde{y}^{PLS}_{T+2|R+1}$ are asymptotically biased estimators of the $\hat{y}^{o}_{T+1+s}$. This result suggests that the finite-sample property of $\tilde{\boldsymbol{y}}^{PLS}_{T+2|R+1}$ may not be as good as that of $\tilde{\boldsymbol{y}}^{PLS}_{T+2|R}$, even if $R < K$.

Second, while not shown here, it could be shown that part (i) of Theorem 4 holds for the regression with more than R PLS factors and up to K PLS factors. Given that these factors do not contribute to improve the accuracy of the PLS forecasting, we from now on refer to them as “uninformative” PLS factors.

Third, when $R = K$, the PLS forecast $\tilde{y}^{PLS}_{T+2|R+1}$ is expected to have poor finite-sample properties if $N \geq T$ and/or $N/T = O(1)$. The parts of the PLS factors that are spuriously correlated with the target variable is no longer asymptotically negligible, and they hurt the accuracy of the PLS forecast. This result does not necessarily imply that when $R < K$, use of more than K PLS factors must produce an inconsistent estimator of $\hat{y}^o_{T+2}$. However, as shown in the next section, the regressions with more than K PLS factors almost always produce poor forecasting results unless T is dominantly larger than N or the variance of the error term in the target variable is small (that is, the common factors in predictor variables have strong forecasting power for the target variable). For this reason, we refer to the factors other than the first K factors as "spurious" PLS factors.

In the next section, we consider the finite sample properties of the "informative," the "uninformative," and the "spurious" PLS factors.

1.3 Simulation Results

In this section, we report our simulation results. Our simulation setups are designed to investigate the following. First, we examine how the finite-sample in-sample and out-of-sample performances of the PLS regression changes as the number of PLS factors used increases to the asymptotically optimal number (R), and as the more than the optimal number of PLS factors is used. Second, we compare the performances of the forecasts produced by the regressions with PLS factors, principal component (PC) factors, and all of predictor variables. Third, we examine whether the actual number of PLS factors that maximizes forecasting power in finite sample is close to the asymptotically optimal number (R) of PLS factors that our asymptotic analysis suggests. Fourth, we consider in-sample and out-of-sample performances of the R informative, the (K–R) uninformative, and the spurious PLS factors.

1.3.1 Simulation Setup

We simulate data following Kelly and Pruitt (2015) and Stock and Watson (2002a). Specifically, we generate data with the following equations:

$$
\begin{aligned}
&y_{t+1} = a_y^{1/2}(\Sigma_{h=1}^{K} f_{ht}^{*}\beta_h^{*}) + (1-a_y)^{1/2}u_{t+1};\\
&x_{it} = a_x^{1/2}(\Sigma_{h=1}^{K} f_{ht}^{*}\phi_{hi}) + (1-a_x)^{1/2}e_{it}^{*};\\
&f_{ht}^{*} = \rho_f f_{h,t-1}^{*} + w_{ht};\\
&e_{it}^{*} = \rho_e e_{i,t-1}^{*} + \tilde{e}_{it};\ \tilde{e}_{it} = (1+\rho_c^2)\varepsilon_{i+1,t} + \rho_c(\varepsilon_{i,t} + \varepsilon_{i+2,t})
\end{aligned}
$$

where the u_{t+1} $(t = 2, ..., T+2)$, ε_{it} $(i = 1, ..., N, N+1, N+2)$, and ϕ_{hi} $(h = 1, ..., K, i = 1, ..., N)$ are all random draws from $N(0,1)$.

All of the factors f_{ht}^{*} are generated with the same AR(1) coefficient ρ_f. The initial values of the K factors f_{h0}^{*} $(h = 1, ..., K)$ are zeros, while the error terms w_{ht} are independently and identically drawn from $N(0, (1-\rho_f^2)v_h)$. Under this setup, $\text{var}(f_{ht}) \approx v_h$ for most of different t.

All the idiosyncratic error components in x_{it}, e_{it}^{*}, are generated with the same AR(1) coefficient ρ_e. The initial values of the ε_{i0} are independently drawn from $N(0,1)$. The idiosyncratic components e_{it}^{*} are cross-sectionally correlated. We control the degree of cross-section correlations by changing the value of the parameter ρ_c. The value of β_h^{*} equals one (zero) if the corresponding factor f_{ht}^{*} is correlated (uncorrelated) with the target variable y_{t+1}.

After we generate the sum of the common components in x_{it} $(\Sigma_{h=1}^{K} f_{ht}^{*}\phi_{hi})$, the part of y_{t+1} explained by the common factors $(\Sigma_{h=1}^{K} f_{ht}^{*}\beta_h^{*})$, and the idiosyncratic error components in x_{it} (e_{it}^{*}), we normalize them such that they have unit variances. By this normalization, we can use the two parameters a_x and a_y to control for the explanatory power of the common factors $\boldsymbol{f}_{.t}^{*} = (f_{1t}^{*}, ..., f_{Kt}^{*})'$ for the predictors x_{it} and the target variable y_{t+1}, respectively. Notice that the parameter a_x equals the probability limit of the average R^2 from individual regressions of x_{it} on the common factors $\boldsymbol{f}^*_{\boldsymbol{.}t}$, while a_y equals the probability limit of the R^2 from the regression of y_{t+1} on $\boldsymbol{f}^*_{\boldsymbol{.}t}$.

We use $\boldsymbol{\Omega}^*_{\boldsymbol{F}}$ to denote $\mathrm{Var}(\boldsymbol{f}^*_{\boldsymbol{.}t}) = \boldsymbol{diag}(v_1, ..., v_K)$. The variables with superscripted star, f^*_{ht}, e^*_{it}, β^*_h and $\boldsymbol{\Omega}^*_{\boldsymbol{F}}$ are not the same as the variables, f_{ht}, e_{it}, β_h and $\boldsymbol{\Omega}_{\boldsymbol{F}}$, that are used in section 1.2. However, they are related roughly as follows:

$$\begin{aligned}
f_{ht} &\approx a_x^{1/2} f^*_{ht} / \sqrt{\mathrm{var}(\Sigma^K_{h=1} f^*_{ht}\phi_{hi})};\\
e_{it} &\approx (1 - a_x) e^*_{it} / \sqrt{\mathrm{var}(e^*_{it})};\\
\beta_h &\approx \beta^*_h \frac{a_y^{1/2} \sqrt{\mathrm{var}(\Sigma^K_{h=1} f^*_{ht}\phi_{hi})}}{a_x^{1/2} \sqrt{\mathrm{var}(\Sigma^K_{k=1} f^*_{ht}\beta^*_h)}};\\
\boldsymbol{\Omega}_{\boldsymbol{F}} &\approx \frac{a_x}{\mathrm{var}(\Sigma^K_{h=1} f^*_{ht}\phi_{hi})} \boldsymbol{\Omega}^*_F
\end{aligned}$$

For each set of the parameter values chosen (T, N, K, $\boldsymbol{\Omega}_{\boldsymbol{F}}$, a_x, a_y, ρ_f, ρ_e, and ρ_c), we generate 1,000 different samples. Each sample contains $(T + 1)$ observations. The first T observations are used to estimate the parameters that are needed to forecast y_{T+2}. The PLS factors are computed by the NIPLS algorithm introduced in Appendix A. The last observation is used to compute the forecasting error by a forecast $\hat{y}_{T+2}$. Using the forecasting errors from the 1,000 samples, we compute the following out-of-sample R^2 of a forecast:

$$R^2_{OS} \equiv 1 - \frac{\Sigma^{1000}_{s=1} \left(y^{[s]}_{T+2} - \hat{y}^{[s]}_{T+2}\right)^2}{\Sigma^{1000}_{s=1} \left(y^{[s]}_{T+2} - \bar{y}^{[s]}\right)^2}$$

where $\bar{y}^{[s]} = T^{-1} \Sigma^{T+1}_{t=2} y^{[s]}_{t+1}$ and s indexes simulated samples. The second term of R^2_{OS} is a ratio of the mean square error (MSE) of the forecast and the MSE of the target variable's historical mean. When the forecast is more accurate than the historical mean, the out-of-sample R^2_{OS} must be a positive number. In contrast, when the historical mean outperforms, the measure becomes negative. The R^2_{OS} measure is also used in Kelly and Pruitt (2015).

Our benchmark case is the case in which data are generated with $N = T = 100$, $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)' = (1, 0, 1, 0)'$, $\boldsymbol{\Omega}_{\boldsymbol{F}}^*= \boldsymbol{diag}(3, 3, 5, 5)$, $a_x = 0.2$, $a_y = 0.7$, and $\rho_f= \rho_e= \rho_c = 0.5$. Under this setup, the asymptotically optimal number of PLS factors for forecasting (R) equals two, because there are two groups of factors the same variance (the factors whose asymptotic variances equal to 3 and the factors whose asymptotic variances equal to 5) and at least one factor from each of the two groups is correlated with the target variable. This is the case in which $R = 2 < K = 4$ in the notation used in section 1.2. That is, there are two informative and two uninformative PLS factors. The rest of the PLS factors are spurious factors.

1.3.2 Simulation Results from the Benchmark Case

We begin by examining how the performances of the forecast by the PLS regression change as the number of PLS factors used increases. To save space, we denote the number of factors (PLS or PC factors) used for forecasting by q.

Table A.1 reports the results from our benchmark case. The table shows how the in-sample fits and out-of-sample forecasting performances of the PLS regression change as different numbers of factors are used: from one to ten. For each regression with a different number of PLS factors, the table reports the average and standard error of the in-sample R^2's and the R^2_{OS}'s from individual PLS regressions with 1,000 different samples. We use the adjusted R^2 instead of the usual R^2 for the in-sample R^2. We do so because the usual R^2 always increases with the number of regressors used, while the adjusted R^2does not. Figure 1 depicts the changes in average in-sample R^2 and R^2_{OS} as the number of PLS factors used (q) increases.

Table A.1 and Figure A.1 show that the in-sample R^2 from the PLS regression always increases as more factors are used. In contrast, the R^2_{OS} from the PLS regression is always peaked at $q = 2 = R$, the asymptotically optimal number of PLS factors for forecasting. As q increases, the R_{OS}^2 keeps falling. For example, as q increases to 10, the R_{OS}^2 falls to 18 percent points while the in-sample R^2 increases to 90 percent points. Table A.1 and Figure A.1 clearly show that a PLS regression with higher in-sample R^2does not guarantee a more accurate forecasting result.

For our benchmark case, our asymptotic results predict that the forecast obtained using 2 to 4 PLS factors are consistent estimators of the optimal forecast $\hat{y}_{T+2}^o = \Sigma_{h=1}^K f_{h,T+1}^* \beta_h^*$. Interestingly, however, the simulation results reported in Table A.1 and Figure A.1 indicate that using 3 or 4 PLS factors would rather produce less accurate forecasts. The results in Table A.1 and Figure A.1 suggest that the PLS regression with more than R and up to K PLS factors would produce less precise forecasts.

Our asymptotic results also predict that the regressions using more than 4 PLS factors would produce spuriously high in-sample R^2's and low R_{OS}^2's. The results reported in Table A.1 and Figure A.1 are also consistent with this prediction.

1.3.3 Comparisons of the Forecasting Powers of PLS and PC Factors

We here compare the forecasting performances of the regressions with PLS factors, principal component (PC) factors, and all of the predictors. For this comparison, we generate data with five common factors with $\boldsymbol{\Omega}_{\boldsymbol{F}}^* = 5 \times \boldsymbol{I}_5$ and $\boldsymbol{\beta}^* = (1, 0, 0, 0, 0)'$. For these data, $R = 1 < K = 5$. That is, the asymptotically optimal number of PLS factors equals one, while the number of PC factors to be used for optimal forecasting is five.

Tables A.2 and A.3 report the out-of-sample forecasting performances of the PLS regressions with the first PLS factor only (PLS1), the PC regression with first five PC factors (PC5) and the usual OLS regression with all predictor variables (OLS). Table A.2 reports the results obtained from the data with $(N, T) = (80, 100)$, while Table

A.3 reports the results from the data with $(N,T) = (160, 200)$. For this simulation exercise, N is chosen to be smaller than T to make the regression with all available predictors possible. Data are simulated with many different combinations of the parameters, a_x, a_y, ρ_f, ρ_e, and ρ_c. To save space, we only report the results obtained using the data generated with $\rho_f = \rho_e = \rho_c$. For each combination of data generating parameters, the highest R^2_{OS} is marked in bold.

Tables A.2 and A.3 show that the forecasting performance of the OLS regression is always dominated by those of the PLS1 and PC5 regressions. The R^2_{OS} from the OLS regression is always negative, indicating that the historical mean of the target variable is a better forecast than the OLS forecast. This finding is consistent with the well-known fact that the MSE of the OLS forecast increases with the number of predictors used; see, for example, Carrasco and Rossi (2016) and Stock and Watson (2006), among many.

Tables A.2 and A.3 show that when the common factors' explanatory power for the predictors is low ($a_x = 0.1$ or 0.2) and their explanatory power for the target variable is relatively high ($a_y = 0.5$ or 0.7), the PLS1 forecast outperforms the PC5 forecast. This pattern remains the same even if different AR(1) coefficients (ρ_f and ρ_e) and the cross-section correlation parameter (ρ_c) are used. In general, the PC5 regression produces more accurate forecasts when the factors are more weakly autocorrelated and predictor variables' idiosyncratic components are less serially and cross-sectionally correlated.

One interesting observation from Tables A.2 and A.3 is that when the PLS1 regression outperforms the PC5 regression, it does so by a relatively greater margin. For example, in Table A.2, the R^2_{OS} from the PLS1 regression is almost twice larger than that from the PC5 regression when $a_x = 0.1$, $a_y = 0.7$, and $\rho_c = \rho_e = \rho_f = 0.5$: the R^2_{OS}'s from the PLS1 and PC5 regressions are 39.9 percent points and 20.5 percent points, respectively. As shown in Table A.3, when the sample size is doubled while other parameter values remain unchanged, the R^2_{OS} from the PLS1 regression is still higher than that from the PC5 regression by 15.3 percent points: the out-of-sample R^2's from the PLS1 and PC5 regressions are 48.5 percent points and 33.2 percent points, respectively.

Tables A.2 and A.3 also report the number of common factors ($\tilde{K}$) estimated by the method of Ahn and Horenstein (2013) (AH, 2013). The tables show that when a_x is low, the AH tends to underestimate the number of common factors in predictor variables. Not surprisingly, the PC regression with the estimated number of factors ($\tilde{K}$) significantly underperforms the PL5 regression, especially when a_x is low, although these results are unreported here to save space. When a_x is low, the PLS1 regression significantly outperforms the PC regression with the estimated number of factors more than it does the PC5 regression.

The main findings from Tables A.2 and A.3 can be summarized as follows. First, the PLS1 regression produces more accurate forecasts than the PC5 regression when the common factors in predictor variables are relatively weak factors. Second, when the predictors have stronger factors, the PC5 regression outperforms the PLS1 regression in forecasting, but generally by a small margin. These results indicate that the PLS regression is a viable forecasting tool which is particularly useful when the factor structure in predictor variables is weak.

1.3.4 Forecasting with Asymptotically Optimal Number of PLS Factors

We now consider the finite-sample properties of the PLS regression when the asymptotically optimal number of PLS factors for forecasting (R) is greater than one. Tables A.4 – A.7 report the forecasting performances of the PLS regressions with three different numbers of PLS factors. The R^2_{OS}'s from the PLS regressions with one, two, and three are reported in the PLS1, PLS2, and PLS3 columns, respectively. All of the data used for the results reported in Tables A.4 – A.7 are generated with $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 5, 7)$ and $a_y = 0.7$, while different parameter values are used for a_x, ρ_c, ρ_e, and ρ_f. Notice that for all of the cases considered in Tables A.4 to A.7, the optimal number of PLS factors for forecasting is three ($R = 3$).

Table A.4 reports the forecasting results from the data with $N = T = 100$ and $N = T = 200$. Differently from what our asymptotic results predict, the R^2_{OS} from the PLS3 regression is lowest for all cases. When the common factors' explanatory power for predictor variables is low (e.g., $a_x = 0.1$), the PLS1 regression more often outperforms the PLS2 regression. In contrast, as the factors' explanatory power becomes stronger ($a_x = 0.2$ or 0.3), the PLS2 regression more often outperforms the PLS1 regression.

Table A.5 reports the forecasting results obtained using larger data: $N = T = 1,000$ and $N = T = 2,000$. Even for these large data, the R^2_{OS} from the PLS3 regression is highest only once (when $a_x = 0.3$, $\rho_c = \rho_e = 0.3$, $\rho_f = 0$, and $N = T = 2,000$). For other cases, the PLS2 regression produces the highest R^2_{OS}. As shown in Table A.6, for the unusually large data with $N = T = 7,000$, we can observe that the PLS3 regression outperforms the PLS1 and PLS2 regressions for some cases. When we have extremely large data with $N = T = 10,000$, as shown in Table A.7, the PLS3 regression outperforms the PLS1 and PLS2 regressions for all different data specifications. However, even for the cases in which the PLS3 regression outperforms the PLS1 and PLS2 regressions, the prediction gain by the PLS regression is marginal.

The three main implications from Tables A.4 – A.7 are the following. First, when the asymptotically optimal number of PLS factors for forecasting (R) is greater than one, the PLS regressions using a fewer number of PLS factors very often produce more accurate forecasts than the PLS regression using R factors, unless the data are exceptionally large. Second, the PLS1 regression often produces a more accurate forecast than the regressions with PLS factors, especially when the sample size is small and the common factors in predictor variables are weak.

Third and finally, when larger data are used and $R = 1$, using more than one PLS factor could produce more accurate forecasts. However, the accuracy gains by using additional factors are not substantial. The gains are generally very marginal. This result indicates that when the optimal number of PLS factors (R) is unknown, using only one PLS factor for forecasting could be a useful alternative. This is so because, as shown in Table A.1, using more than R PLS factors can produce much poorer forecasts than the PLS regression with only one factor.

Why then could the regression with a fewer than R PLS factors produce more accurate forecasts that the regression with R PLS factors does? There are two possible answers. The first possible answer is that for the simulated data used for Tables A.4 – A.7, the variances of the three factors (f_{ht}) are not sufficiently distinct for PLS regressions unless exceptionally large data are used. For example, when $a_x = 0.1$ is chosen, the three factors' variances are 0.3, 0.5, and 0.7, respectively. It is possible that in small samples, these differences in factor variances may not be sufficient to make all of the three PLS factors have independent forecasting power for the target variable. In unreported experiments, we have tried to use more dispersed variances for the three factors. However, under our data generating setting, we need to assign very small variance to one factor to assign much greater variances to two other factors. For that case, the factor with the smallest variance has too weak explanatory power for both predictor variables and the target variable. Unless the sample is exceptionally large, the factor models constructed with such factors are more or less similar to two or one factor models. For this reason, in the unreported experiments, the PLS1 and PLS2 regressions very often outperform the PLS3 regression.

The second possible answer is the following. While the PLS factors used for our simulation exercises are generated by the NIPLS algorithm, they are the orthogonalized versions of the PLS factors examined in section 1.2. The asymptotically dominant term in the first R PLS factors ($\tilde{\boldsymbol{P}}_{1:R}^{PLS}$) is $\boldsymbol{G}_0\boldsymbol{D}_0(R)$, where $\boldsymbol{D}_0(R)$ is a Vandermonde matrix. It is well known that Vandermonde matrices are highly ill conditioned matrices in the sense that the columns of a Vandermonde matrix are highly collinear; see Dax (2017). Thus, the first one or two columns of the matrix $\boldsymbol{G}_0\boldsymbol{D}_0(R)$, and correspondingly, the first and second PLS factors may contain most of the forecasting power for the target variable vector $\boldsymbol{y}$.

1.3.5 Spurious Correlation Problem and Relative Sizes of N and T

Our asymptotic results suggest that depending on whether T is dominantly larger than N or not, use of more than K PLS factors for forecasting could exaggerate in-sample goodness of fit of the PLS regression and produce poor forecasting outcomes. Thus, we now examine how sensitive the finite-sample performances of the regressions with more than K PLS factors to the $N - T$ ratio. We generate data using the parameter values for the benchmark case. we investigate how the performances of the PLS regression change as the ratio N/T varies.

Figure A.2 shows how the out-of-sample forecasting performances of the PLS regressions with different numbers of PLS factors change as N increases while T is fixed at 100. The figure for the case with $N = T = 100$ is identical to Figure A.1. Figure A.2 indicates that when N/T is low, the regressions with more than 4 PLS factors do not significantly underperform the regressions with smaller number of PLS factors. For example, when $N = 20$, use of more than 4 PLS factors does not incur seriously inflated in-sample R^2 nor deteriorated R^2_{OS}. It appears that the problem of spurious correlations between PLS factors and the target variable is not severe when N is substantially smaller than T. However, Figure A.2 also shows that the spurious correlation problem becomes substantial as N increases. For the cases with N closer to or greater than T, the PLS regression produces more highly inflated in-sample R^2's and lower R^2_{OS}'s more PLS factors are used.

Figures A.3 and A.4 report the simulation results obtained using different N with $T = 200$ and $T = 500$. All other data generating parameter values are the same as those which are used for the benchmark case. While greater T values are used, the $N - T$ ratios used for the two figures are the same as those which are used for Figure A.2. The reported results in Figures A.3 and A.4 are not materially different from those in Figure A.2. Overall, the results reported in Figures A.2 – A.4 are consistent with the notion that the severity of the spurious correlation problem and the $N - T$ are inversely related.

1.3.6 Spurious Correlation Problem and Explanatory Power of Latent Factors

Our asymptotic results indicate that the spurious correlation problem occurs by the interaction of the error terms in the target variable and predictor variables. Consequently, we can expect that the spurious correlation problem would be mitigated as the variances of the errors decrease, or equivalently as the explanatory power of the latent factors for the target variable and predictor variables. Thus, we now examine how the forecasting performances of the PLS regression would change as a_y or a_x increases.

Figure A.5 shows how the significance of the spurious correlation problem of the PLS regression changes as the value of a_y (explanatory power of latent factors for the target variable) changes. The values of other data generating parameters used for Figure A.5 are the same as those that are used for Figure A.1. Figure A.5 shows that the significance of the spurious correlation problem falls as a_y increases (the variance of the error term in the target variable falls). When $a_y = 0.1$, the regression with 10

PLS factors yields about negative 100 percent points R^2_{OS}. This means that the MSE of the forecast from the PLS regression is twice as large as the MSE of the historical mean of the target variable. In contrast, when $a_y = 1$ (no error in the target variable), the R^2_{OS} is peaked when two PLS factors are used and it remains little changed as more PLS factors are used. It is clear that the degree of spurious correlation between PLS factors and the target variable is strongly negatively related to the explanatory power of the common factors for the target variable (a_y).

Figures A.6 – A.7 report the results obtained replicating the simulation exercises used for Figure A.5, but with greater values of a_x (0.5 and 0.7, respectively). The patterns of the PLS forecasting performance reported in Figures A.6 and A.7 are virtually identical to those that are reported in Figure A.8.

We now examine how the significance of the spurious correlation problem is related to the explanatory power of the common factors for predictor variables (a_x). To do so, we generate data with many different values of a_x (from 0.1 to 0.99), but with the same values for other data generating parameters that are used for Table A.1 and Figure A.1. Figure A.8 reports the results for the cases with $a_y = 0.7$. When $a_x = 1$, that is, when the four common factors can perfectly explain predictor variables, the 5th PLS factor is a perfect linear combination of the first 4 PLS factors. For this reason, the maximum value of a_x we use is 0.99. Since $a_y = 0.7$ is used, the regression with two PLS factors is expected to produce the in-sample and out-of-sample R^2's of about 70%.

The main findings from Figure A.8 are the following. First, when a_x is small (the explanatory power of the latent factors for predictor variables is weak), the forecasting power of the regression with the first 2 PLS factors is somewhat lower than what our asymptotic results suggest. Although it is not clear from the figure, the R^2_{OS} from the regression with 2 PLS factors is always lower than the expected level of 70%.

However, as a_x increases, the R^2_{OS} from the regression with the two PLS factors rises close to 70%.

Second, when a_y is low, the in-sample R^2's from the regressions with 3 and 4 PLS factors are higher than 70%, while the R^2_{OS}'s from the same regressions are lower than 70%. This result contradicts our asymptotic results predicting that the third and fourth PLS factors do not have additional in-sample explanatory power and additional out-of-sample forecasting power. The result seems to be consistent with the notion that the uninformative PLS factors (the third and fourth factors) may also suffer from the spurious correlation problem in finite samples, especially when a_x is low. The spurious correlation effect on the third and fourth PLS factors weakens as a_x increases. For the extreme case with $a_x = 0.99$, the third and fourth PLS factors perform as our asymptotic results predict: use of the two factors does not decrease the forecasting power of the PLS regression. In addition, use of the two factors does not inflate the in-sample goodness of fit of the regression.

Figure A.8 shows that the regressions with more than 4 PLS factors suffer from the spurious correlation effect, even if a_x is near to one: the average in-sample R^2 are inflated and the R^2_{OS} deteriorates as more PLS factors are used.

As Figures A.9 and A.10 show, the results from Table A.8 remain unaltered even if different values are used for a_y (0.5 and 0.3). The patterns of the changes in in-sample and out-of-sample performances of the PLS regressions by using different numbers of factors used are similar across Figure A.8 to A.10.

1.3.7 Forecasting with Uninformative and Spurious PLS Factors

We here consider how the uninformative and spurious PLS factors would influence the quality of the PLS forecast. Figures A.11 and A.12 highlight the performances of the uninformative and spurious factors in finite samples. For the figures, we generate the data using the benchmark parameter values. For the benchmark case, there are two informative PLS factors and two uninformative factors, and the rest of the PLS factors are spurious factors. We focus on how use of the two uninformative factors and other spurious factors would influence the quality of the PLS forecasts. We have seen from Figure A.1 and other figures that using more than the informative PLS factors decreases the accuracy of the PLS forecast.

Figure A.11 zooms up how the patterns of the in-sample and out-of-sample performances of the regressions using uninformative and spurious factors change as the explanatory power of the latent factors for predictor variables (a_x) increases from 0.2 to 0.995. The average in-sample R^2's from the regressions with different numbers of PLS factors are marked by red squares connected with dotted line. The R^2_{OS}'s are marked by blue circles connected with solid line. For both lines, the lighter color is associated with the greater value of a_x. The average in-sample R^2's and the R^2_{OS}'s for the case with $a_x = 0.2$ are identical to those that are reported in Figure A.1.

When predictors have weak factor structure (low a_x), using the two uninformative factors increases the average in-sample R^2 and decreases the R^2_{OS} from the PLS regression. Using spurious factors additionally inflates the in-sample $R^2$and decreases the R^2_{OS} even more. When a_x is extremely high (0.995), the two uninformative PLS factors do not inflate the in-sample R^2 and do not hurt the forecasting accuracy. Both the average in-sample R^2 and R^2_{OS} match the value of a_y (0.7) that is used to generate data. In contrast, using the spurious PLS factors additionally still inflates the in-sample $R^2$and deteriorate the forecasting accuracy of the regression. The case of $a_x = 0.995$ is, of course, an extreme case. For more empirically plausible cases, using the uninformative PLS factors tends to inflate the in-sample R^2 while decreasing the forecasting power of the regression.

For Figure A.12, we experiment the same simulations conducted for Figure A.11, but with larger data. The data are generated with $N = T = 2000$. In Figure A.12, using the two uninformative factors no longer hurts the forecasting power of the regression, even when a_x is low. However, using the two uninformative PLS factors tends to inflate the in-sample fit of the regression unless a_x is very high. For any value of a_x, using a larger number of spurious PLS factors inflates the in-sample fit and weakens the forecasting power of the regression.

In order to check how the results from Figures A.11 and A.12 would change if more uninformative factors are added to predictor variables, we conduct the same simulation exercises used for Figures A.11 and A.12, but with a six-factor model with $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3,3,3,5,5,5)$ and $\boldsymbol{\beta}^{*} = (1,0,0,1,0,0)'$. The results from this additional experiment are reported in Figures A.13 and Figure A.14. For the factor model used for the figures, there are two informative and four uninformative PLS factors.

Figure A.13 reports the results obtained using the data with $N = T = 100$ as in Figure A.11. From Figure A.13, we can see that the regression using the 6th PLS factor, which is the fourth uninformative factor, produces inflated in-sample R^2's and decreased R^2_{OS}'s, even when the 6 latent factors have extremely strong explanatory power for predictor variables ($a_x = 0.995$). When $a_x < 0.5$, all of the four uninformative factors perform as spurious factors do: they inflate the in-sample R^2 and deteriorate the R^2_{OS} from the regression.

Figure A.14 reports the result from the data with $N = T = 2,000$ as in Figure A.12. 3 out of 4 uninformative factors perform more consistently with what our asymptotic results predict. However, the last uninformative factor (the 6th PLS factor) behaves more like a spurious factor, especially when a_x is low. The main point from Figures A.11 – A.12 and A.13 – A.14 is that using uninformative PLS factors can significantly lower the accuracy of the PLS forecast, unless data are unusually large.

1.3.8 Summary

The main messages from our simulation results so far can be summarized as follows. First, the forecasting with PLS factors could be a viable alternative to the forecasting with PC factors, especially when the common factors in predictor variables have strong explanatory power for the target variable while having weak power for predictor variables.

Second, the regressions using spurious factors substantially inflate in-sample goodness of fit results while producing significantly poorer out-of-sample forecasting results. Consistent with our asymptotic results, the negative effect of using the spurious factors is weaker when the data with T substantially larger than N are used for the regression, and/or when the common factor in predictor variables have strong explanatory power for the target variable.

Third, the asymptotically optimal number of PLS factors for forecasting is R, the number of the factor groups sharing the same asymptotic variances that are correlated with the target variables. However, the number of the PLS factors that achieves the maximum forecasting power in finite samples is often smaller than R, especially when R is large. This problem does not disappear even if very large data are used (e.g., data with $N = T = 2000$). The optimal number of PLS factors for forecasting in finite samples is close to the asymptotically optimal number, when T is substantially larger than N or explanatory power of the common factors in predictor variables for the target variable is very strong. Interestingly, these cases are precisely the cases in which the effects of the spurious correlations between PLS factors and the target variable are weak. It appears that under the environment in which the spurious correlation between PLS factors and the target variable is not asymptotically negligible, using the asymptotically optimal number of PLS factors would rather produces poorer forecasting results than using a fewer number of PLS factors.

Fourth, using uninformative PLS factors can decrease the forecasting power of the regressions with PLS factors, especially when the spurious correlation between PLS factors and the target variable is strong. One important implication is the following. The total number (K) of factors in predictor variables can be estimated by numerous estimation methods, e.g., Bai and Ng (2002), Onatski (2010), Alessi *et al.* (2010), and Ahn and Horenstein (2013), among many. However, our simulation results indicate that using the estimates from these methods for the number of the PLS factors for forecasting may not be a good practice. Many of the K PLS factors could be uninformative factors for the target variables and using the large number of uninformative PLS factors can produce poorer forecasting results.

Fifth and finally, using the first PLS factor only may not be a bad alternative when the optimal number of the PLS factors for forecasting is not readily available. Our simulation results indicate that a large portion of the information for the target variable contained in PLS factors is in the first PLS factor. When the asymptotically optimal number of factors for forecasting is more than one, the forecasting gain by using more PLS factors in addition to the first PLS factor is not substantial. Also, the regression using a fewer than the asymptotically optimal number of PLS factors, often produces more accurate forecasts than the regression using the asymptotically optimal number of PLS factors. The forecasting loss by using only the first PLS factor seems to exceed the loss by using too many PLS factors.

1.3.9 Cross-Validation Estimation for the Optimal Number of PLS Factors

One important question we have not addressed yet is how we can determine the optimal number of PLS factors for forecasting. In our asymptotic analysis, the number of informative PLS factors (R) is the optimal number. However, our simulation results indicate that the optimal number of the PLS factors in finite samples is often smaller than R. As an alternative to determine the optimal number of PLS factors for forecasting in finite sample, we examine the finite-sample performances of a cross-validation method.

For the cross-validation method we consider, we divide the whole available data (with $T+1$ observations) into two parts, training and test data. Let us use int($\centerdot$) to denote the integer part of the inside of the parenthesis. The initial training data consist of the observations from $t = 2$ to $t = \text{int}((0.7)(T+1)) \equiv T^{*}+1$, while the test data set consists of the observations from $t = \text{int}((0.7)(T+1)) \equiv T^{*}+2$ to $t = T+1$. For a given time $s \in [T^{*}+2, T+1]$, we forecast y_s using a given number of PLS factors and the parameter estimates obtained from the training data from $t = 2$ to $t = s$–1. Let MSE(q) be the MSE of the forecasts for y_s obtained using q PLS factors. The cross-validation estimate of the optimal number of PLS factors, which we denote by $\hat{R}_{CV}$, is the value of q that minimizes MSE(q).

In Tables A.8 to A.11, we compare the forecasting performances of the regressions with different numbers of PLS factors ($q = 1, 2, ..., 10$) and the regression using the estimated number of PLS factors by the cross-validation method. We refer to the regression with q PLS factors as "PLSq" regression. To save space, we only report the forecasting results from the PLS1 to PLS6 regressions, while up to 10 PLS factors were calculated, and cross-validation were conducted over the 10 PLS factors in all experiments. For the results reported in Tables A.8 to A.10, we use a five-factor model with $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5, 7)$ and $\boldsymbol{\beta}^{*} = (1, 0, 1, 0, 1)'$. For this model, the first 3 PLS factors are informative ones and the next 2 PLS factors are uninformative ones: $R = 3$ and $K = 5$. The other parameters are set at their benchmark values: $a_x = 0.2$, $a_y = 0.7$, and $\rho_f = \rho_e = \rho_c$= 0.5.

The main findings from the results reported in Table A.8 – A.10 are as follows.

First, consistent with the results reported in Tables A.4 and A.5, the PLS2 regression very often outperforms the PLS3 regression despite that $q = 3$ is the asymptotically optimal number of PLS factors for forecasting. Second, the forecasting performance of the cross-validation augmented PLS (CV-PLS) regression is generally comparable to that of the PLS2 regression. Third and finally, the performance of the PLS1 regression is not significantly dominated by that of the CV-PLS regression. In fact, the PLS1 regression often outperforms the CV-PLS regression, especially when the explanatory power of the factors for the target variable is low, as Table A.10 shows. When the CV-PLS regression outperforms the PLS 1 regression, the gain by using the CV-PLS regression instead of the PLS1 regression is generally marginal. In addition, the out-of-sample performance of the PLS1 regression is not far behind that of the PLS2 regression.

Lastly, we consider a special case that is inspired by Groen and Kapetanios (2016). They have considered the cases in which all of the predictor variables x_{it} are individually directly correlated with the target variables, not just indirectly through the latent factors $\boldsymbol{f}_{\boldsymbol{.}t}$. Our asymptotic analysis does not consider such cases. However, it would be interesting to see how the CV-PLS regression would perform for such cases.

We here consider a special case in which some predictors have some direct forecasting power for the target variable. Specifically, we consider a case in which the first predictor has some direct forecasting power for the target variable y_{t+1}:

$$x_{1t} = \Sigma_{h=1}^{5} \phi_{hi} f_{ht}^{*} + e_{1t}^{*}; \quad e_{1t}^{*} = \rho_{eu}^{1/2} u_{t+1}^{*} + (1 - \rho_{eu})^{1/2} v_{1t}^{*} \tag{1.30}$$

where $u_{t+1}^{*} = (1 - a_y)^{1/2} u_{t+1}$ and the v_{1t}^{*} are random draws from $N(0, 1)$. All other predictors and the target variable are generated by the process explained in subsection 1.3.1. Observe that when $\rho_{eu} = 1$, the idiosyncratic component of x_{1i}, e_{1t}^{*}, has perfect information about the error term of the target variable, u_{t+1}. While we only consider the case in which only one predictor variable has some direct forecasting power for the target variable, our simulation results would have some implications for more general cases in which a small number of predictor variables have some direct forecasting power for the target variable.

Even if some predictor variables have direct forecasting power for the target variable, the PC factors do not convey such information because they are extracted without using the information about correlations between predictor variables. However, the PLS factors may contain the information generated by the correlations between individual predictors and the target variable.

We generate data using the same benchmark data generating parameter values used for Tables A.8 – A.10, except that the first predictor variable is generated by (1.30). Table A.11 reports some of the simulation results. When ρ_{eu} is low, the PLS2 regression outperforms other PLS regressions including the CV-PLS regression. This result is consistent with the results reported in Tables A.8 – A.10. However, one interesting observation from Table A.11 is that the spurious correction problem by using the sixth PLS factor, which is a spurious factor, mitigates as ρ_{eu} increases. In fact, for the cases with $\rho_{eu} \geq 0.8$, the PLS6 regression is the best performer for forecasting, among PLS1 to PLS6. For the cases with $\rho_{eu} \geq 0.9$, the PLS6 regression significantly outperforms the PLS1 – PLS3 regressions.

The following conjecture seems to be reasonable for these results. When predictors do not have strong direct forecasting power (forecasting power conditional on the common factors) for the target variable, some parts of the PLS factors become spuriously correlated with the target variables. Using too many PLS factors amplifies the effect of the spurious correlation and hurts forecasting accuracy. However, when some predictors have strong direct forecasting power, the spurious components of the PLS factors are asymptotically dominated by the informative parts of the PLS factors.

Consequently, the effect of the spurious correlation is no longer prevalent.

Another interesting observation from Table A.11 is that when $\rho_{eu} \geq 0.9$, the CV-PLS regression outperforms the PLS6 regression, especially when larger data are used. In addition, the mean of $\hat{R}_{CV}$ exceeds $K = 5$ (the total number of latent factors in predictor variables). These results indicate that cross-validation methods are most useful for the PLS regression when some predictors have strong direct forecasting power of which economists are not aware. The gain by using the cross-validation method could be substantial. Our simulation results indicate that PLS users should be advised to estimate the optimal number of PLS factors by some cross-validation methods.

1.4 Empirical Application

In this section, we conduct a typical empirical study to demonstrate applicability of our results. We use actual macroeconomic data. Total 178 monthly variables were collected from FRED-MD data of McCracken and Ng (2016), FRED and ISM (Institute for Supply Management) to closely mimic the dataset Stock and Watson (2002b) used. The data have 732 time series observations, from 1959:01 to 2019:12. Following Stock and Watson (2002b) and McCracken and Ng (2016), we categorize the variables in the data into eight major groups: output and income; labor market; housing; consumption, orders and inventories; money and credit; interest and exchange rates; prices; and stock market.

We conduct 12-month-ahead forecasting exercises. To do so, we transform the data to make them stationary. The transformation methods are first or second differencing (in log form). The detailed information is listed in the appendix. We also standardize the transformed variables so that they have unit variances and zero means. Finally, we screen the data for any possible outliers. We drop the outliers from the data and treat them as missing values. The final data set contains a balanced panel of 108 variables and an unbalanced panel of 70 variables. The missing values are estimated by the EM algorithm of the PC method with the number of common factors estimated by the method of Ahn and Horenstein (2013).

The following forecasting equation is used for our data analysis:

$$\hat{y}^{12}_{T+12|T} = \hat{a} + \hat{\boldsymbol{b}}' \hat{\boldsymbol{f}}_{\boldsymbol{.}T} + \Sigma^{p}_{h=1} \hat{c}_h y_{T-h+1} \tag{1.31}$$

where $\hat{\boldsymbol{f}}_{\boldsymbol{.}T}$ is the $K \times 1$ factor vector estimated by the PC or PLS methods and $\hat{a}$, $\hat{\boldsymbol{b}}$ and $\hat{c}_h$ are OLS estimates. The maximum number of the AR coefficients and the maximum number of the factors in $\hat{\boldsymbol{f}}_{\boldsymbol{.}T}$ are restricted to be 6 and 12, respectively.

The number of factors used matters for predictive power. We tried several experiments with different choices of K. For both PLS and PCA, the forecast with given k, were $k = 1, 2, ..., 12$ will be tested. (Denoted as PLS k and PCA k) This exercise always chooses the same number of factors in all time series. Tables A.12 and A.13 display results for PLS k and PCA k, with k = 1, 2, 3, 4. Also, the Bayesian Information Criteria of Stock and Watson (2002b) will be used for PLS and PCA to determine k (denoted as PLS BIC and PCA BIC). Finally, the consistent estimator for the number of true factors, Ahn and Horenstein (2013)'s eigenvalue test will be applied to PCA forecasts (denoted as PCA AH) and Cross-Validation method is applied to PLS forecast (denoted as PLS CV). The lag of the dependent variable, p is decided by BIC, following Stock and Watson (2002b) for all factor estimation methods, except for PLS forecast with cross-validation. For PLS CV, both k and p were chosen by cross-validation which will be discussed more in detail later.

The target variables y^{12}_{T+12} are generated as following. We treat real and price variables as $I(1)$ and $I(2)$ variables in logarithms, respectively, following Stock and Watson (2002b). For instance, industrial production (IP) is a real variable and hence the target variable is generated by the following equation:

$$y_{T+12}^{12} = (1200/12)\ln(\mathrm{IP}_{T+12}/\mathrm{IP}_T) \text{ and } y_T = 1200\ln(\mathrm{IP}_T/\mathrm{IP}_{T-1})$$

On the other hand, the Consumer Price Index (CPI) is price variable and the target variable is

$$y_{T+12}^{12} = (1200/12)\ln(\mathrm{CPI}_{T+12}/\mathrm{CPI}_T) - 1200\ln(\mathrm{CPI}_{T+12}/\mathrm{CPI}_T)$$
$$\text{and } y_T = 1200\Delta\ln(\mathrm{CPI}_T/\mathrm{CPI}_{T-1})$$

The y_{T+12}^{12} in the above equations is the target variable being forecasted and y_T and y_{T-h} where $h = 1, 2, ..., p$ are used as lagged dependent variables in the main forecasting framework (1.31). Since PLS estimate factors by using target variables, we should have enough time series observations of the target variable when PLS factors are estimated. Target variables with too much missing values may lead to unstable PLS factors and hence makes the empirical analysis inaccurate. Therefore, only the variables whose time-series observations are more than 80% of the first factor estimation period, are forecasted in the practice. Due to this reason, we have 144 different target variables.

At time T, the target variables are generated, factors are estimated and a model with k and p is chosen. Then the forecasting equation (1.31) is estimated by regressing y_{t+12}^{12} onto the estimated factors $\hat{\boldsymbol{f}}_{\boldsymbol{.}t}$ and other observables, $t = 1, 2, ..., T$. This procedure yields the estimated parameters, $\hat{a}$, $\hat{\boldsymbol{b}}$ and $\hat{c}_h$. Given the estimated factors, observables, selected model and parameters, the forecast for y_{T+12}^{12} is made, which is denoted as $\hat{y}_{T+12|T}^{12}$ in (1.31). The same procedure is repeated at $T + 1$, forecasting y_{T+13}^{12}. Therefore, this exercise is simulated real-time forecasting practice.

More specifically, the first forecasting starts from $T =$1970:01 to forecast the target variable at 1971:01 ($T+12$). Using predictors from 1959:03 to 1970:01, the factors $\hat{\boldsymbol{f}}_{\boldsymbol{.}T}$

are estimated by PLS and PCA, where T =1970:01. The first two-month observations are dropped due to possible second difference (in log) in the transformation procedure. Then the model is chosen for k and p. The first 12-months data is used for initial conditions, hence the parameters $\hat{a}$, $\hat{\boldsymbol{b}}$, and $\hat{c}_h$ are estimated using target variable y^{12}_{t+12}, estimated factors and observable variables, from t =1960:01 to 1970 : 01. Then the final forecast $\hat{y}^{12}_{T+12|T}$ is made, following (1.31) to forecast y^{12}_{T+12}. This procedure repeats until 2018:12. After all the 144 number of $\{\hat{y}^{12}_{t+12|t}\}^{2018:12}_{t=1970:01}$ are generated by all methods, the mean squared errors (MSE) is calculated by comparing $\{y^{12}_{t+12}\}^{2018:12}_{t=1970:01}$ with $\{\hat{y}^{12}_{t+12|t}\}^{2018:12}_{t=1970:01}$.

For PLS, cross-validation is also used for model selection. The training and test set are 70% and 30% of the available data, respectively. For instance, the first forecast was made in 1970:01. In this case, since y^{12}_{t+12} is defined as a function of raw target variables at $t+12$ and t as the above equation show, the available information of the target variable is $\{y^{12}_{t+12}\}^{1969:01}_{t=1959:01}$, while the available predictors are $\{x_t\}^{1970:01}_{t=1959:01}$, at 1970:01. The first 70% of these time series is close to 1966:04. Therefore, we pretend only the data, $\{x_t\}^{1966:04}_{t=1959:01}$ and $\{y^{12}_{t+12}\}^{1965:04}_{t=1959:01}$ are available to us (training set) and estimate factors, select the model and forecast $\{\hat{y}^{12}_{t+12|t}\}_{t=1966:04}$ using only those available data. Repeat this process at 1966:05 and continue until 1969:01 is reached (test set). Then compare the predicted value $\{\hat{y}^{12}_{t+12|t}\}^{1969:01}_{t=1966:04}$ with the actual value available to us, $\{y^{12}_{t+12}\}^{1969:01}_{t=1966:04}$ and pick the pair $(p^*_{CV},\ k^*_{CV})$ that gives the best out-of-sample R^2 in this test set. With the pair of $(p^*_{CV},\ k^*_{CV})$, make a forecast $\hat{y}^{12}_{t+12|t=1970:01}$, using the estimated $k^*_{CV}\times 1$ dimensional PLS factors $\hat{F}_{1970:01}$ with lagged value $y_{1970:01}$, ..., $y_{1970:01-p^*_{CV}+1}$. This process will give PLS forecasts with cross-validation.

The results are presented in the following two tables. The entries are the percentage out-of-sample R^2, which is

$$100 \times [1 - RMSE(method)] \; = 100 \times \; \left[1 - \frac{MSE(method)}{MSE(by\, mean)}\right]$$

where the relative mean squared error (RMSE) of the given factor estimation method is the mean squared errors relative to that of a forecast based on a naïve historical mean of the target variable.

Table A.12 displays out-of-sample R^2 of different factor estimation methods for forecasting eight important variables that Stock and Watson (2002b) focused. Table A.13 is the forecasting results for the whole 144 target variables. For Table A.13, after all the target variables are forecasted, they are collected according to the category they belong to. Then the median out-of-sample R^2 of each category is reported in Table A.13. Table A.12 and A.13 reveal interesting findings. First, consistent with the simulation results, incorporating more PLS factors deteriorates forecasting power significantly. For some variables, incorporating even the third PLS factor yields worse forecasting performance than a naïve forecast based on the historical mean of the target variable, such as Personal Income. Some target variables show improvement when we use more PLS factors, such as Producer Price Index. However, the predictive improvement even in this case is marginal and PLS1 still gives a fairly good result.

Second, PLS CV, which is the PLS forecast with cross-validation, does not dominate PLS1. Rather, the PLS CV is very often dominated by forecasting performance of PLS1. This result is again consistent with the simulation results. Third, PLS BIC show significantly worse performance. Even forecasts based on a historical mean strictly dominates PLS BIC forecast in many cases. This is not surprising, because the BIC chooses$k$, to maximize the in-sample fit with some penalties. Therefore, the number of PLS factors that explains the in-sample movements may not necessarily forecast the future variable well. Rather, it is most likely that the PLS factors that explain the in-sample variation well, would give a bad forecasting performance, due to spurious correlation property of PLS factors. As the tables confirm, PLS forecasts where the number of PLS factors chosen by BIC, actually poorly performs in the data. Finally, PLS1 outperforms other alternative methods, from PLS2 to PCA AH. Sometimes PLS1 does not give the best results, but the predictive power between the best forecasts and PLS1 is similar.

1.5 Conclusion

This paper has considered the PLS regression to forecast a single target variable using many predictors. Asymptotic and finite-sample properties of the PLS factors are derived. Our main findings from our asymptotic analysis are the following. First, the number of the necessary PLS factors for the asymptotically optimal forecasting crucially depends on the covariance structure of the common factors in predictor variables. Previous studies routinely assume that all of the factors have distinct asymptotic variances. However, our results indicate that the asymptotical optimal number of the PLS factors for forecasting is determined by the number of distinct asymptotic variances of the common factors. If all factors have the same asymptotic variances, the optimal number of PLS factor is one. Second, the regression with more than the total number of factors could substantially poor forecasting results.

The main findings from our simulation exercises are the following. First, use of more than the asymptotically optimal number of PLS factors generally reduces forecasting power of the PLS factors. Second, the actual optimal number of PLS factors is often smaller than the asymptotically optimal number, unless unrealistically large data are used. Third, the first PLS factor contains the most predictive information about the target variable in finite samples. The additional explanatory power that can be obtained by the second or more PLS factors is not substantial. Fourth and finally, our simulation results indicate that the regression with the number of PLS factors determined by some cross-validation methods can dramatically increase forecasting power, when some predictor variables have strong direct power for the target variable.

Chapter 2

FACTOR-AUGMENTED FORECASTING IN BIG DATA

2.1 Introduction

Recent technological improvements lead economists to gain access to larger datasets. Given that the traditional econometric models, such as Ordinary Least Squares, may not work properly in a large data set, factor analysis is commonly used to analyze big data. Factor model assumes that all variables are generated by a few latent reference variables, or factors. Factor-augmented forecasts, also known as diffusion index forecasts, estimate the latent factors from many predictors and then augment them to the forecasting equations, along with other observable variables. Many studies have found empirical evidence that factor-augmented forecasts may produce better forecasting performance than traditional forecasting models using autoregression, VAR, or structural models.[1]

In this sense, estimating factors accurately is crucial for better forecasting results. Many factor estimation methods are proposed, but it is difficult to compare them comprehensively. Theoretically, they rely on different assumptions and it may be hard to expect how these factor estimators behave when these assumptions are not satisfied. Empirically, they often forecast different target variables, using different dataset, under different forecasting framework.

The goal of this paper is to analyze comprehensively the predictive performance of various, widely used factor estimations, in a coherent forecasting framework.[2] This

---

[1] Stock and Watson (1999) and Stock and Watson (2002b), among many, show empirical evidence that supports this finding.

[2] D'Agostino and Giannone (2012) and Boivin and Ng (2005) among many, compare the forecast-

paper has three contributions to the existing literature. First, it provides a comprehensive predictive evaluation of many factor estimation techniques under the same forecasting framework. Since different factor estimation methods often estimate factors in different ways, incorporating them in the same data and forecasting scheme is not easy. More specifically, seven different factor estimation methods are tested: Principal Component Analysis, Weighted Principal Component by Boivin and Ng (2006), One-sided Estimation from Forni *et al.* (2005), Targeted Predictors from Bai and Ng (2008a), Partial Least Squares from Kelly and Pruitt (2015) and Ahn and Bae (2020), Two-step Estimation of Doz *et al.* (2011) and Quasi-Maximum Likelihood Estimator of Doz *et al.* (2012). To provide a coherent analysis, I construct big data that contain major U.S. macroeconomic and finance variables. 148 target variables are forecasted, under three forecasting equations, across three different forecasting horizons, using seven factor estimation methods with 11 information criteria that determine the number of estimated factors for forecasting.

The second contribution of this paper is to investigate the common properties of various factor estimation methods and information criteria. I analyze weaknesses and strengths of each factor estimation method in comprehensive environments. Finally, this paper contributes to the existing literature by providing an empirical guidance for forecasting practice. In particular, incorporating different factor estimation methods and information criteria gives 101 possible combinations. Among all the combinations, I find that the first Partial Least Squares factor (PLS1) often outperforms other methods.

This paper provides four novel findings. First, the number of factors used in fore-

---

ing performances of different factor estimation methods. They focus on comparison of the static and dynamic principal component methods. On the other hand, this paper aims to provide comprehensive predictive evaluations of commonly used factor estimation methods.

casting is important for predictive power. Incorporating more factors may not yield better forecasting performance. Rather, forecasting power often deteriorates after a certain number of factors are used. The consistency of certain estimated factors to the true factor space has been proven by many studies.[3] The first finding of this paper contributes to the existing literature by providing an empirical evidence that the forecasting power may deteriorate if we estimate more than those factors.

Second, I find that consistently estimated number of factors in data, may not lead to the best result in empirical forecasting practice. I investigate the forecasting performance of 11 information criteria that determine the number of factors used in forecasting. Inspired by the first finding, this experiment includes number of factors estimation methods popularly used in practice: Bai and Ng (2002), Bai and Ng (2007), Onatski (2010), Alessi *et al.* (2010) and Ahn and Horenstein (2013), as well as Bayesian Information Criteria implemented by Stock and Watson (2002b). For PLS, Ahn and Bae (2020) find from simulations that the first PLS factor (PLS1) very often yields better forecasting performance than forecasting with more PLS factors, even when PLS1 is not theoretically optimal to achieve the maximum forecasting power asymptotically. Inspired by this finding, PLS1 is tested as well. Overall, the forecasts obtained by these information criteria perform well, except for PLS. PLS1, which is not consistent, often outperforms other forecasts. On the other hand, PLS-augmented forecasts with other information criteria may have worse predictive power. This finding implies that consistent estimators for the true number of factors in data, may not lead to the best predictive results in empirical forecasting practice.

---

[3] Connor and Korajczyk (1986), Stock and Watson (2002a) and Bai (2003) for Principal Component Analysis, Kelly and Pruitt (2015) and Ahn and Bae (2020) for Partial Least Squares, Doz *et al.* (2011) for Two-step estimator, Doz *et al.* (2012) for Quasi-Maximum Likelihood estimator and Forni *et al.* (2005) for One-sided estimation.

Third, the best forecasting performance of each factor estimation, chosen across different information criteria is similar. However, forecasting power varies significantly across different information criteria, even when the same factor estimation method is used. Therefore, the choice of factor estimation method, as well as information criteria, is crucial in the empirical forecasting exercise. More specifically, for each factor estimation method, I choose the best forecasting performance out of the 11 information criteria. These 7 best results are similar.[4] However, there is no dominant information criterion that gives the best results for all the target variables, other than PLS1 for PLS. Therefore, the combination of information criteria and factor estimation method is important for forecasting performance.

Finally, I find that PLS often outperforms many factor estimation methods. More specifically, PLS1 usually shows significant improvement upon other factor estimations that involve more discretion about the number of factors and parameter values. Related to the third finding, PLS1 yields the forecasting performance close to the best result from all combinations of 7 factor estimation methods and 11 information criteria. The strong predictive power of PLS comes from its factor estimation strategy. PLS estimates factors using not only predictors but also a target variable, which can explain the significant forecasting improvement of PLS.

This paper is organized as follows. Section 2 discusses econometric framework and introduces factor estimation methods used in this article. Section 3 explains the data and forecasting experiment design. Section 4 reports and interprets empirical results and Section 5 concludes.

---

[4] This result may be explained by the fact that most of the seven methods have been proven to be consistent estimators to the true factor space.

2.2 Forecasting Model and Factor Estimation Methods

2.2.1 Approximate Dynamic Factor Model

Let y_{t+1} be the one-period ahead future value of y_t. y_{t+1} is a variable being forecasted, or referred to as a target variable in this paper. x_t is $N \times 1$ vector of predictors. For all the time series observations, $t = 1, \cdots, T$, denote the $T \times N$ matrix of predictors as X, where $X = [x_1, x_2, \cdots x_T]'$. Both forecast target y_{t+1} and x_t have mean zero. If (y_{t+1}, x_t) follow a dynamic factor model, we can write the model as following.

$$\begin{aligned} y_{t+1} &= \beta(L) f_t + \gamma(L) y_t + \varepsilon_{t+1} \\ x_{it} &= \lambda_i(L) f_t + e_{it} \end{aligned} \tag{2.1}$$

where the dynamic factor f_t is a $q \times 1$ vector and $\beta(L),\ \gamma(L)$ and $\lambda_i(L)$ are lag polynomials. Unlike static factor model where only the contemporaneous factor f_t affects x_{it} and y_{t+1}, the dynamic factor model allows the past factors to generate predictors and the target variable. The approximate dynamic factor model allows weak correlation among idiosyncratic errors, while the exact factor model assumes no correlation between idiosyncratic errors. It is assumed to satisfy $E(\varepsilon_{t+1} | f_t, y_t, x_t, f_{t-1}, y_{t-1}, x_{t-1}, ...) = 0$. Therefore, if the factors and parameters, $\{f_t\},\ \beta(L)$ and $\gamma(L)$ are known, the best forecast for y_{T+1} is $\beta(L) f_T + \gamma(L) y_T$.

Suppose further that the lag polynomials $\beta(L),\ \gamma(L)$ and $\lambda_i(L)$ have finite orders of at most s, such that $\lambda_i(L) = \sum_{j=0}^{s} \lambda_{ij} L^j$ and $\beta(L) = \sum_{j=0}^{s} \beta_j L^j$. Let $F_t = (f_t',\ f_{t-1}',\ ...,\ f_{t-s}')'$ be a $r \times 1$ vector, where $r = (s+1)q$, and Λ be a $N \times r$ matrix whose i-th row is $(\lambda_{i0},\ \lambda_{i1},\ ...,\ \lambda_{is})$. Then we can represent the above dynamic factor model (2.1) into a static representation.

$$\begin{aligned} y_{t+1} &= \beta F_t + \gamma(L) y_t + \varepsilon_{t+1} \\ x_t &= \Lambda F_t + e_t \end{aligned} \tag{2.2}$$

where $\beta = (\beta_0, \beta_1, ..., \beta_s)$ and e_t is a $N \times 1$ vector of $e_t = (e_{1t}, e_{2t}, ..., e_{Nt})'$. The above model (2.2) is referred to as a static representation of dynamic factor model because the model (2.2) does not involve lag expression of factors.[5] A dynamic factor model with q dynamic factor can be represented by a static factor model with r static factors. However, it should be noted that the dimension of f_t is different from that of static factor, F_t. Even though those two models (2.1) and (2.2) are identical, empirical estimation of the two often involves different estimation strategies. Empirically, the static framework in (2.2) estimates factors by time domain analysis and the dynamic factor models in (2.1) are estimated by frequency domain analysis.[6] Most of the factor models that are implemented in this paper estimate the static representation of the dynamic factor model, (2.2). On the other hand, One-sided estimation of Forni *et al.* (2005) estimates the dynamic factors using (2.1).

The main empirical experiments of this paper focus on h-step forecasts. Following Stock and Watson (2002b), I adopt a multistep forecast approach which is assumed to be linear in F_t and y_t with lags, and the h-step-ahead projection is used directly to make the forecast. Therefore, changing the one-step ahead forecast of (2.2) to multistep ahead version gives

$$y^h_{t+h} = \delta_h + \beta_h(L)F_t + \gamma_h(L)y_t + \varepsilon^h_{t+h} \tag{2.3}$$

where y^h_{t+h} is the target variable, or the h-step-ahead variable to be forecasted. The subscript h implies the projection changes according to the forecasting horizon h. Therefore, the forecasting equation (2.3) is the main econometric forecasting framework in this paper.

---

[5] Refer to Stock and Watson (2002b) for detailed explanation.

[6] Refer to Bai and Ng (2008b) for more information.

2.2.2 Factor Estimation Methods

Since the true factor F_t in the main forecasting equation (2.3) is not directly observable in the data, we need to estimate factors. This paper analyzes comprehensively the forecasting power of different estimated factors $\hat{F}_t$ proposed by various literature. Therefore, this chapter briefly introduces the history of factor analysis and explains the factor estimation methods used in the empirical section.

Following Stock and Watson (2010), this paper categorizes the history of factor estimations into three generations. The earliest generation of factor estimation calculates the Gaussian likelihood and estimate factors with the Kalman filter by representing the factor structure into state space model. This method was used by early literature of factor analysis: Engle and Watson (1981), Watson and Engle (1983), Stock and Watson (1989), and Quah and Sargent (1993), among many. However, by its nature, the number of parameters estimated in this model is increasing with N. Due to this reason, this method is not commonly used to analyze big data. Therefore, this method won't be considered in this paper.

The next generation is nonparametric averaging methods. Nonparametric model estimates factors directly from (2.2). As the name implies, the nonparametric models do not require additional assumptions on distribution of disturbance or a model for the factors. To simplify the problem, suppose that the number of the static factors r is known. This approach finds $N \times r$ weight matrix $\alpha = [\alpha_1,\ \alpha_2,\ ...,\ \alpha_r]$ for a given r. The weight vector $\alpha_j,\ j = 1, 2,\ ...,\ r$, is a $N \times 1$ vector and the $r \times 1$ estimated factor $\hat{F}_t$ is calculated by $\hat{F}_t = \alpha' x_t = [\alpha_1,\ \alpha_2,\ ...,\ \alpha_r]' x_t$. For all the time series observations, $t = 1, ..., T$, the factor estimator $\hat{F}$ is calculated as $\hat{F} = X\alpha = X[\alpha_1,\ \alpha_2,\ ...,\ \alpha_r]$. The weight vector α_j works as a cross-sectional average weight of predictors when the factors are estimated. Intuitively, nonparametric factor estimations use cross-sectional averaging of predictors to filter out the effects of idiosyncratic errors and leave only the variations from the factors. Since the comovements from factors are stronger than the cross-sectional correlation among idiosyncratic disturbances, averaging predictors would remove the effects of idiosyncratic errors by the law of large numbers.[7] Hereafter, an arbitrary factor estimator is denoted as $\hat{F}$, and $\hat{F}$ is a general notation for factor estimator, not restricted to any specific factor estimation method.

Principal Component Analysis (PCA)

Principal Component Analysis (hereafter PCA) is one of the most commonly used factor estimation techniques in economics, as discussed by Stock and Watson (2002a,b, 2006), Bai and Ng (2002, 2006), Bai (2003), Bernanke *et al.* (2005), and Ahn and Horenstein (2013), among many. PCA estimates factor loading Λ and factors F_1, F_2, ..., F_T by solving the following least-squares problem.

$$\min_{\Lambda,\,\{F_t\}_{t=1}^{T}} \ \frac{1}{NT}\sum_{t=1}^{T}(x_t-\Lambda F_t)'(x_t-\Lambda F_t) \tag{2.4}$$

subject to a normalization. The solution of this problem, $\hat{\Lambda}$ boils down to the scaled r eigenvectors of the sample covariance matrix of predictors $\hat{\Sigma}_X = T^{-1}X'X$, corresponding to the largest r eigenvalues. It is noteworthy that $N \times r$ weight matrix $\alpha = [\alpha_1, \alpha_2, ..., \alpha_r]$ is the estimated factor loadings, $\hat{\Lambda}$, with the restriction of $\Lambda'\Lambda/N = I_r$. The estimated factor $\hat{F}_t$ is $N^{-1}\hat{\Lambda}x_t$, which is simply obtained by regressing predictors x_t on the estimated factor loading $\hat{\Lambda}$. The estimated factors for all time series are $\hat{F} = N^{-1}X\hat{\Lambda}$. As equation (2.4) shows, the objective of PCA is to estimate factors that explains variance of predictors most.

PCA is widely used, because it is easy to estimate factors and the estimated factors by PCA, $\hat{F}$, are consistent estimators for the true factor space up to rotation. (Connor

---

[7] Refer to Stock and Watson (2016) for more information.

and Korajczyk (1986), Stock and Watson (2002a) and Bai (2003)) However, many factor estimation methods are proposed to improve upon PCA. This paper will focus on two major improvements upon PCA among nonparametric factor estimations.

The first methods aim to improve efficiency, which is a Generalized Principal Component. The Generalized Principal Component and generalized least squares (GLS) share the same intuition. If the variance of the idiosyncratic error is different across predictors, generalized principal component can improve efficiency by adjusting weighting matrix in the least square problem of (2.4). Boivin and Ng (2006) and Forni *et al.* (2005) among many, will be considered. On the other hand, the second group aims to improve forecasting performance. Even though PCA is versatile, PCA may not give the best forecasting result since PCA factors are only obtained from predictors. If there is a forecast target variable of interest, the second group incorporates the information of the target variable when factors are estimated. Targeted Predictors proposed by Bai and Ng (2008a) and Partial Least Squares investigated by Kelly and Pruitt (2015), Groen and Kapetanios (2016) and Ahn and Bae (2020) among many, will be considered.

Generalized Principal Components (Generalized PC)

PCA estimates the factor loading Λ and factors F by solving the least-square problems in (2.4). The intuition of the Generalized Principal Components (hereafter Generalized PC) is similar to that of Generalized Least Squares (GLS) that can possibly improve efficiency of Ordinary Least Squares (OLS) problem. If idiosyncratic errors have different variance across predictors, or have cross-correlation, there can be an efficiency gain by modifying the problem (2.4) as follows. Let Σ_e be the true covariance matrix of idiosyncratic errors and solve the following weighted version of (2.3),

$$\min_{\Lambda, \{F_t\}_{t=1}^{T}} \quad \frac{1}{NT} \sum_{t=1}^{T} (x_t - \Lambda F_t)' \Sigma_e^{-1} (x_t - \Lambda F_t) \tag{2.5}$$

subject to a normalization. The estimated factor loadings obtained from this problem are similar with (2.4), which is the r scaled eigenvectors of $\Sigma_e^{-1/2} \hat{\Sigma}_X \Sigma_e^{-1/2'}$, corresponding to the r largest eigenvalues. The factors can be estimated by $\hat{F} = N^{-1} X \hat{\Lambda}$. Similar with (2.4), the $N \times r$ weight matrix $\alpha = [\alpha_1, \alpha_2, ..., \alpha_r]$ in the Generalized PC is $\hat{\Lambda}$. Therefore, the cross-sectional weight matrix α for the Generalized PC is obtained by the weight-corrected variance matrix, $\Sigma_e^{-1/2} \hat{\Sigma}_X \Sigma_e^{-1/2'}$.

However, this above solution is infeasible because the true variance matrix of idiosyncratic errors, Σ_e is unknown. The feasible Generalized PC estimator is obtained by replacing Σ_e with the estimated $\hat{\Sigma}_e$. This paper will consider two studies that propose different $\hat{\Sigma}_e$, Boivin and Ng (2006) and Forni *et al.* (2005).

Boivin and Ng (2006) propose to generate two-step diagonal weight matrix. In the first step, PCA factors are estimated. In the second step, Generalized PC factors are estimated that solves (2.5) with a diagonal matrix of $\hat{\Sigma}_e$, whose diagonal elements are the sample variance of estimated idiosyncratic error, $\hat{e}_t$ in the first step. On the other hand, Forni *et al.* (2005) propose One-sided Estimation that uses the decomposition of variance matrix of predictors, $\Sigma_X = \Sigma_{\Lambda F} + \Sigma_e$, where $\Sigma_{\Lambda F}$ is the variance of the common component ΛF_t in (2.2). This equality gives $\hat{\Sigma}_e = \hat{\Sigma}_X - \hat{\Sigma}_{\Lambda F}$, where $\hat{\Sigma}_{\Lambda F}$ is estimated by the dynamic principal component analysis of Forni *et al.* (2000) which involves frequency domain analysis. Since the weight matrix is estimated by dynamic principal component analysis that estimates (2.1), the decomposition of the number of the dynamic factors (q) from that of the static factors (r) is needed for this estimation.

Targeted Predictors

On the other hand, several factor estimation methods are propoosed to improve upon PCA for forecasting performance. PCA does not use the information of the target variable while factors are estimated and only predictors are used to estimate factors. There are several factor estimation methods that incorporate the information of the target variable in factor estimation process. They aim to improve PCA upon its forecasting power by estimating factors to maximize explanatory power for a certain target variable. Since this type of factor models estimate factors using the information of the target variable, the estimated factors are different across a target variable of interest. In this sense, I will refer to these estimation methods as 'target specific factor estimation'. Two studies among target specific factor estimations will be investigated in this paper: Targeted Predictors of Bai and Ng (2008a) and Partial Least Squares investigated by Kelly and Pruitt (2015) and Ahn and Bae (2020).

First, Bai and Ng (2008a) propose Targeted Predictors. They argue PCA does consider the predictability of each predictor x_{it} for the target variable y_{t+h} while PCA factors are estimated. In this spirit, they propose to select a group of predictors that have a strong predictable power for the target variable and estimate PCA factors only from this subset of predictors. Targeted Predictors estimate factors by two steps. First, a group of predictive predictors are selected by LASSO (Least Absolute Shrinkage and Selection Operator;) or LARS (Least-Angle Regression). Then PCA is applied over the selected subset of predictors, and the estimated PCA factor is the targeted predictor factors. Therefore, cross-sectional weight matrix α for targeted predictors comes from the selected predictors that have higher predictive power for the target variable than the rest data.

Partial Least Squares (PLS)

The second method of the target specific factor estimation is Partial Least Squares (PLS) that estimates factors which has the maximum covariance with the target variable. PLS is an algorithm that estimates one factor each step, such that the estimated factors have the maximum covariance with the target variable.

For instance, in the first iteration, PLS estimates the $N \times 1$ weight vector α_1 that solves

$$\max_{\alpha_1} \ \frac{1}{T} \sum_{t=1}^{T} \left(\alpha_1' x_t y_{t+1}\right)^2 \tag{2.6}$$

subject to a normalization of $\alpha_j{}' \alpha_j / N \ = 1$. As the problem (2.6) shows, PLS estimates the weight vector α_j to maximize the absolute value of the sample covariance between the estimated PLS factor $\hat{F}_{1,t} = \alpha_1' x_t$ and the forecast target y_{t+1}.

The solution α_1 is the eigenvector of $Xyy'X$, which is $\alpha_1 = X'y/\sqrt{y'XXy}$. Correspondingly, the first estimated PLS factor $\hat{F}_1$ is $\hat{F}_1 = XX'y/\sqrt{y'XX'y}$. Since a scalar $\sqrt{y'XX'y}$ is just a normalization to make α_1 unit norm, we can eventually denote the factor as $\hat{F}_1 = XX'y$. This is because a scalar $\sqrt{y'XX'y}$ does not change the projection matrix of the factors and hence does not affect the predictive power of $\hat{F}_1$.

The above equation (2.6) and the estimated factor, $\hat{F}_1$ are important in this paper. For the rest of the paper, 'the first PLS factor' stands for the PLS factor estimated in the first iteration, $\hat{F}_1 = XX'y$. When we augment the first PLS factor $\hat{F}_1$ in the forecasting equation (2.3), it is referred to as 'the PLS-augmented forecast with the first PLS factor (PLS1)'. It is noteworthy that PLS1 uses only the first PLS factor, $\hat{F}_1$, so $k = 1$ for all time series, $t = 1, ..., T$. In this sense, PLS1, $\hat{F}_1$ may not be a consistent estimator for the true factor space with r dimension, F. However, this simple estimation, or PLS1, often outperforms other factor estimations that involves more discretion about the number of factors and specifications. Moreover, it will be shown that PLS forecast with information criteria that estimates r consistently, may not necessarily give better forecasting results than PLS1. Also, $\hat{F}_1$ provides strong forecasting performance very close to the best results from the total combinations of 7 factor estimation methods and 11 information criteria. More detailed results are presented in the empirical results section.

Since $X'yy'X$ has rank of 1, there exists only one eigenvector that corresponds to non-zero eigenvalue. Put differently, $N \times N$ matrix $X'yy'X$ has only one non-zero eigenvalue and the rest is all zeros. That is why we are able to estimate only one weight vector $\hat{\alpha}_j$ and the corresponding factor estimator $\hat{F}_j$, at each j-th algorithm.

For an arbitrary j-th PLS iteration, similar problems are defined and similar solutions can be obtained. More specifically, denote $1, ..., (j-1)$-th PLS factors as $\hat{F}_1, ..., \hat{F}_{j-1}$. The j-th PLS iteration uses the j-th updated predictors $X_j^* = Q(\hat{F}_1, ..., \hat{F}_{j-1})X$, where $Q(A) = I - A(A'A)^{-1}A'$ for arbitrary matrix A. Therefore, the j-th updated predictors X_j^* is the part of predictors X, which is not explained by $\hat{F}_1, ..., \hat{F}_{j-1}$. Then j-th PLS estimates $N \times 1$ weight vector $\hat{\alpha}_j$ that solves

$$\max_{\alpha_j} \ \frac{1}{T} \sum_{t=1}^{T} \left(\alpha_j' x_{j,t}^* y_{t+1}\right)^2 \tag{2.7}$$

subject to a normalization, where $X_j^* = [x_{j,1}^*, x_{j,2}^*, ..., x_{j,t}^*, ..., x_{j,T}^*]'$. Similar to PLS1 from equation (2.6), j-th PLS is obtained as $\hat{F}_j = X_j^* X_j^{*'} y$. To proceed the next algorithm, we should first update the predictors such that $X_{j+1}^* = Q(\hat{F}_1, ..., \hat{F}_{j-1}, \hat{F}_j)X$ which filter out all the movement of X explained by the PLS factors $\hat{F}_1$, ..., $\hat{F}_{j-1}$, $\hat{F}_j$. Then repeat the same process with the updated predictors, X_{j+1}^*.

Therefore, proceeding to r-th PLS iteration produces the $N \times r$ cross-weight matrix α for PLS is obtained by $\alpha = [X'y,\ X_2^{*'}y,\ ...,\ X_r^{*'}y]$. Contrary to PCA and other nonparametric factor estimations, the cross-sectional weight matrix for PLS factors is obtained as the covariance between (updated) predictors and the target variable. In this sense, PLS estimates factor such that the covariance with the target variable is maximized at each iteration. This is the key difference between PCA or their derivatives with PLS. While PCA estimates the weight vector α_j to maximizes the variance of predictors most, the estimated factors by PLS have the maximum covariance with the target. Therefore, PLS is design to explain the target variables most, not the predictors. While Targeted Predictors use the information of the target variable, it is different from PLS because Targeted Predictors selects a subset of predictors first to explain the target variables, and then estimate PCA factors from this subset of predictors.

PLS was first developed by Wold (1966, 1973, 1982). The properties of the PLS factors under large-N and large-T environment are studied by Kelly and Pruitt (2015), Groen and Kapetanios (2016) and Ahn and Bae (2020), among many. Groen and Kapetanios (2016) investigate the forecasting performance of PLS factors under the model where predictors are directly correlated with a target variable, not through the latent factors. Since this paper considers factor-augmented forecasting framework, the model of Groen and Kapetanios (2016) is not considered in this paper.[8]

Kelly and Pruitt (2015) study the theoretical and finite-sample properties of PLS, under factor model. They defined ‘relevant factors’, by allowing only a subset of the total r factors that generates predictors, actually explains the target variables. Suppose the number of relevant factors is D, where $D \leq r$. They showed that when all the true relevant factors have different variances, the first D PLS factors estimates all the relevant factors up to rotation.

Ahn and Bae (2020) investigate the theoretical and finite-sample properties of PLS, under the model more general than Kelly and Pruitt (2015). The main difference

---

[8] PLS also can be understood as a shrinkage estimation method. For instance, Jong (1993) and Phatak and de Hoog (2002) consider PLS as a shrinkage estimation method.

the two studies lies on the asymptotic variance of factor structure.[9] While Kelly and Pruitt (2015) assume that all factors have asymptotically distinct variances, Ahn and Bae (2020) allow some of their asymptotic variances to identical.

Ahn and Bae (2020) show three important facts about PLS. First, they prove that the optimal number of PLS factor can be reduced further than D. They show we only need the same number of PLS factors with that of distinct variances of the relevant factors. Therefore, the optimal number of PLS factors to achieve the maximum forecasting power to predict a target variable can be reduced further than D. Similarly, Kelly and Pruitt (2015) also show that if all the factors have the identical asymptotic variances, only one PLS factor is needed to span all the factor space needed for forecasting a target variable. Ahn and Bae (2020) extend this finding in more general setting.

Second, the number of PLS factors used for forecasting is important. Ahn and Bae (2020) show that if we use more than the theoretically optimal number of PLS factors, incorporating those PLS factors deteriorates forecasting performance. If we use the exact same number of PLS factors with that of the total factors or relevant factors when some factors share the identical asymptotic variance, it may hurt the predictive power.

Finally, Ahn and Bae (2020) find by empirical evidence and simulations that often,

---

[9] From equation (2.2), $X = F\Lambda' + E$ holds. Note that $F\Lambda' = FBB^{-1}\Lambda' = F^{*}\Lambda^{*'}$ for any $r \times r$ non-singular matrix B. Therefore, F and Λ are not separately identified without any restriction. In factor analysis literature, B is often chosen to normalize F^{*} and Λ^{*} such that the asymptotic variance of F^{*} and Λ^{*} are $r \times r$ diagonal matrix and identity matrix, respectively. Often, B can be also chosen to satisfy that the asymptotic variance of F^{*} and Λ^{*} be $r \times r$ identity matrix and diagonal matrix, respectively. Kelly and Pruitt (2015) and Ahn and Bae (2020) assume the first normalization. However, their conclusions are also valid under the normalization assumption of F^{*} and Λ^{*} whose asymptotic variance are identity and diagonal matrix, respectively.

the first PLS factor (PLS1, which is $\hat{F} = XX'y$) achieves the maximum forecasting power. They find that even when we theoretically may need more PLS factors to asymptotically span the true relevant factor space, unless the sample size is very large, the first PLS often gives the best predictive power and incorporating more decreases forecasting performance.

Their results can explain why PLS1 tends to show strong predictive power in this paper. Also, their results also explain why consistent number of factor estimations that estimate r, might not produce the best forecasting results when they are used to PLS. Often, PLS-augmented forecasts with these consistent number of factor estimation, tend to underperform than a simple PLS1.

Two-Step Estimation

The hybrid method is a third generation of factor estimation. (Stock and Watson (2010)) They aim to estimate factors by combining the efficiency improvement of the state space approach with the principal component analysis that can be used when a large number of predictors are available. One of the benefits, along with possible efficiency gain, is that this approach can be conducted in real time, updating newly released data, since the Kalman filter can handle missing data easily. Also, unlike PCA that implement only cross-sectional average, the Kalman filter and Kalman smoother estimate factors by both of cross-sectional and time-series average. However, this type of factor estimations relies on assumptions on factor structure and distribution of idiosyncratic errors, contrary to nonparametric factor estimation, since this method involves state space representation of factor model. Also, since this method incorporates PCA estimation, the Kalman filter and smoother, factor estimation may take more time for computation than the nonparametric estimation methods. In this paper, two estimations in the hybrid method are considered, Doz *et al.* (2011) and Doz *et al.* (2012).

Doz *et al.* (2011) propose two-step estimation. In the first step, PCA factors are estimated and the model parameters are obtained using the PCA factors. In the second step, the factors are updated using the Kalman smoother. In this sense, the PCA factors and the corresponding parameters are used to initiate the maximum likelihood estimation algorithm. They prove that the two-step estimation is consistent for the true factor space.

Quasi-Maximum Likelihood Estimation (QMLE)

As mentioned earlier, maximum likelihood estimation is not feasible for large-dimensional data since it involves estimation of too many parameters. Doz *et al.* (2012) propose to estimate the factors by MLE, assuming the model follows the exact factor model, where correlation between idiosyncratic errors is zero. Under this assumption, the number of parameters shrinks and MLE is feasible. They show that the factor estimates obtained in this process are consistent for the true factor space, even when the true model is an approximate factor model where the cross-sectional correlation among idiosyncratic errors is allowed. In this sense, their model is Quasi-Maximum Likelihood Estimation (hereafter QMLE) of White (1982). QMLE repeats the process of the above two-step estimation further. It uses PCA as the initial points but repeats the Kalman filter and Kalman smoother until convergence.

2.2.3 Parameter Estimation and Factor-Augmented Forecasting

The main forecasting framework, (2.3) involves the estimation of factors $\{F_t\}$ and parameters, $\delta_h, \beta_h(L)$ and $\gamma_h(L)$, since none of them are directly observable or known. Due to this reason, forecasts of $y^h_{\bar{T}+h}$ in (2.3) are conducted by a two-step procedure as Stock and Watson (2002b) suggest, at given time period $\bar{T}$. First, use all the available sample data $\{x_t\}_{t=1}^{\bar{T}}$ to estimate a sequence of factors, $\{\hat{F}_t\}_{t=1}^{\bar{T}}$. Contrary to Stock and Watson (2002b) whose factor estimators $\{\hat{F}_t\}_{t=1}^{\bar{T}}$ are obtained only by PCA, this paper estimates the factors by PCA, Generalized PC by Boivin and Ng (2006), One-sided estimation of Forni *et al.* (2005), Targeted predictors of Bai and Ng (2008a), Partial Least Squares of Kelly and Pruitt (2015) and Ahn and Bae (2020), Two-step estimation of Doz *et al.* (2011) and QMLE of Doz *et al.* (2012). Hence the seven different factor estimation methods are used to estimate $\{\hat{F}_t\}_{t=1}^{\bar{T}}$. In the second step, the estimators for parameters, $\hat{\delta}_h, \hat{\beta}_h(L)$ and $\hat{\gamma}_h(L)$ are obtained by regressing y_{t+h} on a constant, the estimated factors $\hat{F}_t$ and the lagged dependent variables, $y_t, y_{t-1}, ..., y_{t-p}$. The forecast for $y^h_{\bar{T}+h}$, $\hat{y}^h_{\bar{T}+h}$ is constructed as $\hat{y}^h_{\bar{T}+h} = \hat{\delta}_h + \hat{\beta}_h(L)\hat{F}_{\bar{T}} + \hat{\gamma}_h(L)y_{\bar{T}}$. Since the seven different factor estimation methods yield different estimated factors, the estimated parameters and forecast of the target variables become different across factor methods.

2.3 Data and Forecasting Procedure

2.3.1 Data Description and Transformation

This paper follows forecasting strategies of Stock and Watson (2002b) to make a comprehensive comparison between factor-augmented forecasts, since their forecasting framework is one of the most conventional benchmarks. The forecasting experiment mimics real-time forecasting for 148 monthly macroeconomic target variables in the United States. This paper simulates real-time forecasting in the sense that factors and parameters are estimated, and forecasts are made recursively at every time period. However, it is a pseudo real-time forecasting since the real-time data revision process is not considered. My forecasting experiment may not mimic the actual real-time forecasting completely. But this practice provides us the relative forecasting performance between factor estimation methods and investigates common properties of various factor-augmented forecasts, which are the main interest of this paper.

The main forecasting framework follows that of Stock and Watson (2002b). For each target variable, three forecasting models are compared at the 6-, 12-, and 24-moth forecasting horizons. The forecasts are made by regressing h-step-ahead target variable y^h_{t+h} on regressors at t, which may include factors, lagged factors and lagged target variables.[10] Some variables are transformed to be stationary, following Stock and Watson (2002b). More specifically, the real variables are assumed to be I(1) in logarithms. For instance, since the industrial production (IP) is a real variable, the target variable and the lagged dependent variables are obtained by

$$y^h_{t+h} = (1200/h)\ln(\mathrm{IP}_{t+h}/\mathrm{IP}_t) \text{ and } y_t = 1200\ln(\mathrm{IP}_t/\mathrm{IP}_{t-1}) \tag{2.8}$$

The y^h_{t+h} and y_t obtained from the above transformation (2.8) are used as a target variable being forecasted and lagged dependent variable, respectively, in the main forecasting equation, (2.3). On the other hand, the price-related variables are assumed to be I(2) in logarithms. For example, CPI, along with other nominal variables, are transformed to construct the target variable and the transformed lagged variables as

---

[10] This framework is used to make a consistent comparison across different factor estimation methods. However, it comes at a cost, since it entails modification of some factor methods from their original works. For instance, One-sided estimation of Forni *et al.* (2005) uses two-step estimation. In the first step, the covariance matrices of the common and idiosyncratic components are estimated by the dynamic principal components of Forni *et al.* (2000). In the second step, they use the estimated lag-h covariance matrix of the common components, to construct a forecast. This procedure may lead One-sided estimation to be more efficient than PCA. However, by adopting Stock and Watson (2002b)'s h-step ahead projection framework, the second step cannot be implemented in this paper. Therefore, the possible improvement of One-sided estimation may not be reflected completely in this paper, due to h-step ahead projection framework.

following.

$$\begin{gathered} y^h_{t+h} = (1200/h)\ln(\mathrm{CPI}_{t+h}/\mathrm{CPI}_t) - 1200\ln(\mathrm{CPI}_t/\mathrm{CPI}_{t-1}) \\ \text{and } y_t = 1200\Delta\ln(\mathrm{CPI}_t/\mathrm{CPI}_{t-1}) \end{gathered} \tag{2.9}$$

2.3.2 Factor-Augmented Forecasts

Based on the equation (2.3), the basic factor-augmented forecasts at time t are

$$\hat{y}^h_{t+h|t} = \hat{\delta}_h + \sum_{j=1}^{m} \hat{\beta}'_h \hat{F}_{t-j+1} + \sum_{j=1}^{p} \hat{\gamma}_{hj} y_{t-j+1} \tag{2.10}$$

where $\hat{F}_t$ is k dimensional estimated factors. As mentioned earlier, the estimated factors, $\hat{F}_t$ is obtained from different factor estimation: PCA, Partial Least Squares, Targeted predictors, Two-step estimation, QMLE and One-sided estimation. Given all the other conditions same, different factor estimation leads to different forecasts.

Following Stock and Watson (2002b), three different forecasting equations will be tested. First, the equation denoted as DI, includes only the k number of contemporaneous factors in the forecasting, where DI stands for diffusion index. The DI forecasts involve a choice of k, where $1 \leq k \leq 12$, $m = 1$ and $p = 0$. The second forecasting equation is denoted as DIAR. As the name implies, it combines DI forecast with AR process, which incorporates k dimensional contemporaneous factors as well as p lagged dependent variables. In this sense, $m = 1$, $0 \leq p \leq 6$ is chosen by BIC, and $1 \leq k \leq 12$. The last forecast is DIAR-LAG, where m lagged estimated factors are included in DIAR forecast. Therefore, it includes m lagged factors, the lagged target variables and the contemporaneous factors, where $1 \leq m \leq 3$, $0 \leq p \leq 6$ are chosen by BIC, and $1 \leq k \leq 4$.

It is noteworthy that all the three forecasting equations, DI, DIAR and DIAR-LAG involve the choice of the number of contemporaneous factors, k, under a certain range. The rest parameters, such as m and p are chosen by BIC, if they are allowed to be included. Stock and Watson (2002b) implement BIC to choose k as well, along with other parameters. However, I allow the number of factors ,k, to be estimated by the estimation methods for the number of latent factors in predictors. In this sense, two experiments to determine k are conducted. First, given factor estimation and forecasting equations, the number of factors k will be fixed for all time series, where $k = 1$, ..., $k_{\max}$. This experiment shows how the number of factors affects the predictive power of factor-augmented forecasts. Second, the forecasts with information criteria for k will be conducted. The second experiment implements estimation methods for the number of factors in the data. Therefore, for each time period, the number of factors are estimated by information criteria and updated recursively. This experiment shows the predictive power of information criteria.

2.3.3 Information Criteria

In this paper, number of factors estimation methods will be referred to information criteria. Information criteria indicate model selection methods only for k, if not mentioned otherwise. This practice incorporates important number of factors estimation methods that are commonly used in practice: Bai and Ng (2002) (hereafter "BN"), Onatski (2010) (hereafter "ON"), Alessi *et al.* (2010) (hereafter "ABC") and Ahn and Horenstein (2013) (hereafter "AH"), as well as Bayesian Information Criteria implemented by Stock and Watson (2002b) (hereafter "BIC").

Since some factor estimations involve distinct information criteria, two information criteria are applied only to certain factor estimations. "PLS1" is the PLS-augmented forecasts only with the first PLS factor ($k = 1$ for all time series), suggested by Ahn and Bae (2020). One-sided estimation is applied over two different versions, static with $k = q$ and dynamic with $q(s + 1) \leq k$. Dynamic One-sided estimation needs information about both the number of dynamic factors, q, and that of static factors, k. Bai and Ng (2007) is one of the methods that estimate both q and k at the same time, hence Bai and Ng (2007) is only applied to dynamic One-sided forecasts. Therefore, BIC, BN, ON, ABC, and AH are applied to all the factor estimations: PCA, PLS, Targeted Predictors, Two-step estimation, QMLE and static One-sided estimation. PLS1 is applied only to PLS, and Bai and Ng (2007) method is used to dynamic version of One-sided estimation.

2.3.4 Data transformation and Factor Estimation

Some variables are transformed to be stationary. First, depending on the variable, possible transformation is imposed such as taking log, first or second difference, or log difference. The decision of a proper transformation for a given variable mostly follows previous literature, such as Stock and Watson (2002b) or McCracken and Ng (2016). The list of variables and corresponding transformation is displayed in Appendix, Section **??**. Second, the transformed variables are standardized to have a unit variance and mean zero. Third, the data is screened for outliers. Following Stock and Watson (2002b), any observation whose values are exceeding 10 times the interquartile range from the median is treated as missing values.

Then two sets of empirical factors are constructed. For all the factor estimation methods, the first strategy is to estimate factors, $\hat{F}_t$, only from the balanced panel with 108 time series. The second is to estimate factors from the full data set. Since the full data set involves 70 unbalanced panel, the missing values are estimated by EM algorithm of PCA. After the whole panel is recovered, factors are estimated from this data set. Stock and Watson (2002b) also estimate factors from stacked panel, which stacks 108 current balanced panel and their first lag. However, stacked panel is not used in this paper, because the factors from stacked panel in Stock and Watson

(2002b) tend to have weak forecasting power. The empirical section of this paper only reports the forecasting results based on factor estimation from the balanced panel. The forecasting performance from the full data set is similar with the balanced panel, and the results will be provided upon request.

It is noteworthy that target specific factor estimation requires target variables while factors are estimated, such as Targeted Predictors or PLS. However, some variables have missing values in their initial observations. Therefore, only the variables which have more than 80% of observations in the first factor estimation period, are chosen to be forecasting targets. If we do not have enough target variable observations, target specific factor estimation may not estimate factors precisely, which results in inaccurate analysis. Due to this reason, I have 144 target variables for 6- and 12-months forecasting, and 148 for 24-months predictions.

2.3.5 Recursive Estimation for Simulated Real-Time Forecasting

This paper adopts real-time forecasting strategy, in the sense that the data only available until t will be used to forecast a variable at $t + h$. This includes all the process needed for forecasting, such as factor estimation, parameter estimation, choice of model by BIC and information criteria, among many.

More specifically, the first out-of-sample forecast is made at 1970:01 for the h-period ahead target variable at 1970:01+h, where $h = 6, 12, 24$. Factors are estimated according to 7 different factor estimation methods, using data only from 1959:03 to 1970:01. The first two months are removed due to possible transformation using second difference (in logarithm). With the estimated factors and target variable y^h_{t+h} (and their lags), the model is selected by BIC (m and p) and 11 information criteria (k). After the model is chosen for each of 11 different information criteria, the parameters in (2.10) are estimated by regressing the target variable y^h_{t+h} on the regressors including factors and intercept, from t =1960:01 to 1970:01-h. For model selection and regression, the first 12 months observations from 1959:01 to 1960:01 are not used, since they are used for initial conditions. The final forecast, $\hat{y}^h_{1970:01+h}$ is generated by the intercept, the factor $\hat{F}_{1970:01}$, the lag $y_{1970:01}$, ..., $y_{1970:01-p}$ and possibly the lagged factors, $\hat{F}_{1970:01}$, ..., $\hat{F}_{1970:01-m}$, multiplied by the estimated parameters.

Assuming the true future value $y^h_{1970:01+h}$ is not known, repeat this process for the next month, 1970:02. The factors are estimated by using data from 1959:03 to 1970:02, according to seven different factor estimation methods. Then model is selected by 11 different information criteria and the parameters in (2.10) are estimated by factors and regressors from 1960:01 to 1970:02-h. Using the factor at 1970:02 with other regressors and parameters, make a forecast $\hat{y}^h_{1970:02+h}$ for the unknown future value $y^h_{1970:02+h}$. This recursive estimation and forecasting repeats and the last out-of-sample forecast is made at 2019:12-h. After all the out-of-sample forecasting practice is conducted, the forecast $\hat{y}^h_{t+h}$ and the actual value y^h_{t+h} are compared, from t =1970:01 to 2019:12-h, to evaluate the forecasting performance of all the 7 factor estimations and 11 different information criteria.

2.4 Empirical Results

2.4.1 RMSE and Specifications of Some Factor Estimation Methods

This section reports and interprets the empirical results. The forecasting power of all forecasting practice is measured by relative mean squared errors (RMSE) of the given factor-augment forecasts. The RMSE in this paper is defined as mean squared error (MSE) of the method relative to that of a naive forecast based on a historical mean of the target variable.

$$RMSE(\text{method}) = \frac{MSE(\text{method})}{MSE(\text{mean})}$$

Hence the RMSE larger than one implies that a forecast with a historical mean of the given target variable gives better forecasting result than the factor-augmented forecast.

When those estimations are implemented, I try to follow the original work, on condition that the econometric model in this paper allows. For instance, Bai and Ng (2008a) used soft-thresholding using Elastic Net estimator of Zou and Hastie (2005), to incorporate the case where time-series observation at time t is smaller than cross-sectional predictor observation N. Elastic Net estimator is a convex combination of ridge and LASSO regression,

$$\underset{\beta}{min} \;\; RSS + \lambda_1 \sum_{j=1}^{N} |\beta_j| + \lambda_2 \sum_{j=1}^{N} {\beta_j}^2$$

If λ_1 is larger, the regression becomes closer to the LASSO and if λ_2 is larger, it becomes closer to the ridge regression. Bai and Ng (2008) report three sets of λ_2, where $\lambda_2 = 0.25, 0.5, 0.75$. Following the original work, the three λ_2 are implemented in this article, using Zou and Hastie (2005).

On the other hand, Boivin and Ng (2006) try several weighting matrices used in (2.5). Among them, the two weighting matrices that outperform others are chosen in this paper: Rule SWa and SWb. For Rule SWa, the weight matrix $\hat{\Sigma}_e$ is a diagonal matrix whose elements are the estimated variances of idiosyncratic errors. Rule SWa is also considered in Jones (2001). For Rule SWb, the weight matrix is a diagonal matrix whose elements are the average of estimated covariance of idiosyncratic errors. Those rules are focused on the effects of residuals. Boivin and Ng (2006) also test to estimate PCA factors only from a specific groups of predictors, not from the whole data. Four rules are tested, and the most predictive rule among them, which is Rule B, is used in this paper. Rule B estimates PCA factors only from nominal variables, which corresponds to Category 5 (Money and Credit), 6 (Interest and Exchange

Rates) and 7 (Prices) in this article.[11] Therefore, the most predictive three rules in the original work of Boivin and Ng (2006) are implemented: Rule SWa and SWb for weight matrix and Rule B that estimates PCA factors only from nominal variables.

For One-sided Estimation of Forni *et al.* (2005), two versions are tested. Since One-sided Estimation involves estimation of (2.1), the first is a static version, which imposes $q = r$. In this set up, the number of dynamic factors, q is identical to that of the static factors, r, and it is assumed that no lagged factors in (2.1) generate data. The second is a dynamic version, which imposes $q \leq r$. However, we should disentangle q from r in this case and there are many possible combinations for fixed q and r, given that DI and DIAR forecasts allow at maximum $r = k = 12$ factors used in forecasting. Allowing q and r fixed for $r = 1, 2, ..., 12$ generates 12! possible combinations of q and r. Hence, for the dynamic version of One-sided Estimation, the information criterion of Bai and Ng (2007) is used since it estimates q and r at the same time.

2.4.2 Major Findings

Number of Factors and Forecasting Performance

The 12-month-ahead DIAR forecast ($h = 12$) will be focused in this paper unless otherwise mentioned. The rest results not discussed in the article will be provided upon request. Table B.1 and B.2 display DIAR 12-month ahead forecasts for the eight important variables forecasted in Stock and Watson (2002b), four real and four nominal variables. Table B.1 and B.2 report factor-augmented forecasts of all the factor estimation methods by given $k = 1, 2, ..., 12$. The number of lagged dependent variables, p, is determined by BIC for all factor-augmented forecasts. For instance, PCA forecast for industrial production with $k = 1$ is 0.893, which means the mean

---

[11] The detailed list of those categories are presented in Appendix.

squared of error of PCA forecast with $k = 1$ is 89.3% of that of forecast by a historical mean of industrial production. Therefore, lower RMSE indicates larger forecasting improvements upon a forecast based on mean. If a naive forecast by mean outperforms the given method, the RMSE of the corresponding method is greater than one.

Table B.1 and B.2 show the number of contemporaneous factors included in forecasting equation, k, is usually important for predictive power for almost all types of factor-augmented forecasts. Some factor estimation leads to relatively robust forecasts to the number of factors used, compared to other factor-augmented forecasts, such as PLS. However, predictive performance of even these methods still often varies significantly with the number of contemporaneous factors used in forecasting. Often, the forecasting power deteriorates as more than a certain number of factors are used.

Table B.3 and B.4 summarize 12-month-ahead DIAR forecasts for the entire 144 target variables.[12] All target variables belong to one of the eight categories and Table B.3 and B.4 first forecast all target variables with the same procedure conducted in Table B.1 and B.2, with a given k. Then all the 144 forecasts are sorted into the eight categories according to their target variables. Finally, for each k, the entries show the median RMSE of the given factor estimation method in the corresponding category. Consistent with Table B.1 and B.2, the number of contemporaneous factors incorporated in forecasting equation often matter for the predictive power, for most of the target variable categories and many factor estimation methods. Furthermore, incorporating more factors after certain number of factors may increase RMSE, which

---

[12] The target variables are constructed from h-period ahead future values, as the equation 2.8 and 2.9 show. Therefore, missing values of the first h months can eventually decide the number of target variables. Since target-specific factor estimations require target variables when factors are estimated, only the variables that have more than 80% of time-series observations in the first factor estimation period, are considered as target variables. Hence there exists 144 target variables for 6 and 12-months ahead forecasting, and 148 for 24-months ahead forecasting.

implies lower forecasting improvement, consistent with Table B.1 and B.2.

The PLS forecast by given k from Table B.1 to B.4 is an extreme example of this tendency: incorporating more PLS factors in the forecasting equation (2.10) drastically deteriorates forecasting power after a certain k. For some target variables, even including more than three PLS factors lead PLS-augmented forecasts to perform worse than a naïve forecasts by historical mean. In many cases, the predictive loss resulted from incorporating more than the necessary factors in forecasting, is most serious for PLS among all the factor estimation methods.

Especially the four real variables show very drastic increase of RMSE as the number of PLS factors used in forecasting, k increases. For instance, 12-month-ahead DIAR PLS forecasts for Manufacturing and Trade Sales have RMSE over 2, which implies significantly worse forecasting result compared to a naïve forecast by simple mean. As the tables display, PLS1 usually gives the lowest RMSE. This drastic RMSE increase of PLS forecasts across k is prevalent, for many forecasting target, forecasting models and forecasting horizons.

This finding is also consistent with Ahn and Bae (2020). They find that PLS may have a significant predictive gain compared to PCA which does not use information of target and estimate factors only from predictors. However, they theoretically show that incorporating more than the necessary PLS can hurt the forecasting power. More specifically, suppose that 1, ..., C number of PLS factors asymptotically span the true factor space that is needed for forecasting the target variable. They show that incorporating $C + 1$ and more PLS factors can deteriorate the forecasting results. They also show from simulations and empirical results that PLS until C factors, which is theoretically consistent estimators for the true factor space for the target variable, are not often needed in the actual data, unless the data size is incredibly large. They find that PLS1 often tends to give the best results in relatively small sample and actual data. Therefore, the empirical results from Table B.1 to B.4 are consistent with their findings that the best forecasting performance comes from PLS1 and RMSE drastically increases when more PLS factors are used.

Information Criteria and Forecasting Performance

We may not know how many contemporaneous factors should be used in the actual forecasting practice. The number of factors, k, that minimizes RMSE and gives the optimal predictive gain differs across factor estimation methods, target variables, as well as forecasting horizons as Table B.1 to B.4 show. Therefore, analyzing the results from Table B.1 to B.4 and deciding what a fixed number of factors k should be used for forecasting, may not be helpful for actual forecasting experiment in practice. This is because researchers should often decide the number of factors according to her research design, target variables, forecasting horizons as well as factor estimation methods, only using available data.

Since some factor estimation involves distinct information criteria, two information criteria are only applied to two factor estimations. “PLS1” is the PLS-augmented forecasts with $k = 1$, suggested by Ahn and Bae (2020). It is noteworthy that the number of factors by BN, ON, ABC, AH and BIC estimations are estimated at every t and be used to forecast the target variable at $t+h$, for all $t =$1970:01 to 2019:12$-h$. However, since PLS1 uses only the first PLS factors, it does not involve this recursive process to estimate k every time series. Another distinct information criterion is Bai and Ng (2007) applied to the dynamic version of One-sided Estimation. By the nature of One-sided Estimation, it requires information about the number of dynamic factors, q, as well as the number of static factors k. There are many estimation methods for estimating the number of dynamic factors, q, such as Hallin and Liška (2007). However, in order to implement the main forecasting framework of this paper, we need information of both of q and k. Bai and Ng (2007) developed a method that estimates both of q and k at the same time, hence their method is applied to the forecasts by One-sided Estimation.

The results are presented in Table B.5 to B.10. $\hat{k}_{BIC}$ denotes RMSE of the given factor-augmented forecasts, based on BIC estimator used in Stock and Watson (2002b). Since BN estimator involves choice of penalty function, four penalty functions are tested. $\hat{k}_{BN-p1}$, $\hat{k}_{BN-p2}$, $\hat{k}_{BN-p3}$ and $\hat{k}_{BN-BIC}$ stand for RMSE obtained by the number of factors estimated by BN estimator, using IC_{p1}, IC_{p2}, IC_{p3} and BIC_3 of Bai and Ng (2002), respectively. $\hat{k}_{AH}$ and $\hat{k}_{ON}$ are AH and ON estimators, $\hat{k}_{ABC-L}$ and $\hat{k}_{ABC-S}$ are ABC estimators with large and small windows, respectively. PLS1 is the PLS forecasts with $k = 1$ for all time series, and BN2007 is the dynamic One-sided forecasts, obtained by Bai and Ng (2007)'s information criterion.

If the information criterion can't be applied to a given factor estimation method, the corresponding result stays blank. For instance, since PLS1 is the PLS forecasts with $k = 1$, all the other factor estimation than PLS, such as PCA, Targeted Predictors, Two-Step estimation, QMLE, and One-sided estimation, remain blank for the PLS1 information criteria. The 'Mean' row indicates the mean of the given factor-augmented forecasts over all the information criteria applied. The 'Best, Given k' show the minimum RMSE or the best predictive power of the corresponding factor estimation by given k, denoted in bold in Table B.1 to B.4.

For example, PCA column for forecasting industrial production in Table B.5 has a 'Mean' row of 0.713, which is the average RMSE of all the PCA forecasts applied by available information criteria, from $\hat{k}_{BIC}$ to $\hat{k}_{ABC-S}$. The 'Best, Given k' result is 0.699, which is the PCA forecasts for industrial production with $k = 3$, in Table B.1.

Table B.5 to B.6 display all factor-augmented forecasts by information criteria for the eight target variables. For Table B.7 to B.8, the whole 144 target variables are forecasted by seven factor estimations with 11 different information criteria, $\hat{k}_{BIC}$ to $\hat{k}_{ABC-S}$, along with PLS1 and BN2007. Then the forecasts are sorted into the eight categories according to their target variables and the median RMSE of each category is reported, similar to Table B.3 to B.4.

Inspection of the Table B.5 to B.8 reveals that the minimum RMSE of all information criteria estimation is often lower than the lowest RMSE of given k presented in Table B.1 to B.4. It implies that the forecasts obtained by an appropriate number of factor estimation method may improve upon the ex-post best forecast results with given k. For instance, Stock and Watson (2002b) show that in most cases, incorporating BIC factor and lag order selection to PCA provides little or no improvement over just using two factors, DI with $k = 2$ for predicting real variables. The above comparisons imply this conclusion holds only for BIC estimator of PCA-augmented forecasts. Often the forecasts obtained by BIC estimator for number of factors ($\hat{k}_{BIC}$) does not outperform DI with $k = 2$. However, other information criteria such as ON estimator may outperform the naïve DI forecasts with $k = 2$ for real variables.[13]

The 'Mean' rows of Table B.5 to B.8 show that the mean of forecasts based on information criteria is generally slightly worse than the ex-post best forecasts of a given k. It implies on average, forecasts obtained by information criteria for k, using only the data available at t, gives forecasting performance close to the ex-post best factor-augmented forecasts with given k. In this sense, number of factor estimation methods generally perform well for many factor estimation methods for forecasting practice.

However, some factor-augmented forecasts by information criteria may not yield a good predictive performance. PLS can be the example for this case. Unlike other

---

[13] More detailed results about PCA, including DI PCA forecast with given k and by information criteria, are presented in Table B.15 to B.26.

factor forecasts with information criteria, the ‘Mean’ of PLS forecasts by information criteria are often far from ‘Best, Given k’ forecast. It implies that PLS may be very sensitive to the choice of information criteria. For instance, information criteria, such as BIC by Stock and Watson (2002b) ($\hat{k}_{BIC}$), Information criteria by Bai and Ng (2007)($\hat{k}_{BN}$) or ABC estimator of Alessi *et al.* (2010)($\hat{k}_{ABC}$), often tend to yield RMSE larger than one for PLS-augmented forecasts. This tendency is more serious for the real variables.

It is noteworthy that PLS1 very often yields the best forecast result among all the information criteria for PLS forecasts. The strong predictive power of PLS1 tend to be prevalent in various target variables and different forecasting horizons. PLS-augmented forecasts usually have the best predictive gain when $k = 1$, which is PLS1. Due to this reason, PLS1 forecast is usually identical to ‘Best, Given k’ forecast. This finding again reconfirms the empirical findings of Ahn and Bae (2020), that find PLS1 often shows strong predictive performance in empirical forecasting practice.

This finding also shows that consistently estimated number of factors, may not always lead to the best result in empirical forecasting practice. For instance, PLS forecasts by consistent estimators for the true number of factors, such as BN, AH, ON and ABC estimators, may often give worse predictive power than PLS1. Ahn and Bae (2020) explain this finding by following three steps. First, suppose that there are k number of total factors that generates predictors. Assume furthermore that only D number of the total factors govern the target variable. This incorporates the case where not all the factors that explain predictors are able to forecast the target variable. Among the D relevant factors for the target variable, they have C number of different or distinct variances, where $C \leq D$. Then only C PLS factors are needed to forecast the target variable asymptotically, where $C \leq D \leq k$.[14] Therefore, for

---

[14] Notation in Ahn and Bae (2020) has been modified to be consistent with this paper.

PLS, we may not need the whole number of factors that generates the predictors.

Second, they furthermore show that if we estimate more than C PLS factors, incorporating $C+1$ and more PLS factors can hurt the predictive power. Finally, they find that even though 1, ..., C PLS factors asymptotically give the best forecasting results, PLS1 often gives the best forecasting power in actual empirical forecasting practice and simulation. This is a finite property of PLS, and as sample size increases, we need the whole $1, ..., C$ PLS factors to achieve the maximum predictive power. However, unless the sample size is incredibly huge, PLS1 often gives better results in actual forecasting. Therefore, first, we need only $C \leq k$ number of PLS, theoretically. Second, adding $C+1$ and more PLS factors can hurt the predictive power. Finally, in finite sample, PLS1 tends to give the best result and often, we don't need all C PLS factors. The results of Table B.1 to B.8, as well as Ahn and Bae (2020) confirm that consistent estimator for the true number of factors in the predictors, may not always lead to good forecasting performance in practice.

Large Variations on Forecasting Performances, Across Factor Estimation Methods and Information Criteria

Empirically, there are large variations in terms of factor estimation methods and information criteria used. Therefore, I compare different factor-augmented forecasts by information criteria comprehensively. The first analysis is to compare the average predictive power of the given factor estimation, across different information criteria. Table B.9 and B.11 report the mean RMSE over information criteria of respective factor estimation method for the eight target variables and the entire 144 variables by categories, respectively. They simply collect the 'Mean' rows from Table B.5 to B.8. On the other hand, the second analysis is presented in Table B.10 and B.12. These tables report the minimum RMSE, or the best forecasting results of the given factor estimation method, chosen across all the information criteria, for eight variables and the whole 144 variables by categories, respectively. They gather the lowest RMSE in bold in Table B.5 to B.6 and Table B.7 to B.8, respectively.

For example, PCA forecasts for industrial production in Table B.9 is 0.713, identical to the 'mean' row of PCA forecasts for industrial production in Table B.5. For the same example of PCA forecasts for industrial production in Table B.10, it shows 0.659. This is the minimum RMSE of PCA forecasts for industrial production, chosen across different information criteria, denoted in bold in Table B.5 by $\hat{k}_{ON}$.

It should be emphasized that since PLS1 is PLS-augmented forecasts with k=1 always, PLS1 does not involve any other information criteria. Due to this reason, the minimum RMSE of PLS1 in Table B.10 and B.12 are actually identical to the mean RMSE of PLS1 in Table B.9 and B.11. This is because the minimum in Table B.10 and B.12 and mean in Table B.9 and B.11 are chosen across information criteria.

Table B.9 to B.12 demonstrate the common patterns of all the factor-augmented forecasts by information criteria. First, the best forecasts of all the seven factor estimation methods in Table B.10 and B.12 are very similar. Even though the best results are similar, some factor estimations such as PLS, Targeted predictors, and Weighted Principal Components often achieve the minimum RMSE among other factor estimation methods, hence in bold. This finding that the best forecasting performances are similar across all the factor estimation methods, may be explained by the fact that most of the seven methods have been proven consistent estimators to the true factor space.

Second, the average RMSE over information criteria of respective factor estimation method, as presented in Table B.9 and B.11, varies significantly across factor estimation methods. This exercise shows how the corresponding factor estimation method is sensitive to the choice of information criteria, in the actual forecasting practice. The average predictive power varies more significantly than the best forecasting results in Table B.10 and B.12. This is because often, it is hard to find a dominant information criterion that gives the best results for all the target variables, other than PLS1 for PLS. This finding implies the choice of factor estimation method is crucial in empirical forecasting exercise.

Strong Forecasting Performance of PLS1

Overall, PLS1 often yields a forecasting performance very close to the best result from the whole combinations of 7 factor estimation methods and 11 information criteria. The lowest RMSE of PLS1 in Table B.9 and B.11 supports this finding. The strong forecasting power of PLS1 can be also found from Table B.10 and B.12. PLS1 often yields forecasting performance very close to the best results in bold in those tables, which are equivalent to the best performance from the whole combinations of all the factor estimation methods and information criteria. Strong predictive power of PLS comes from its factor estimation strategy. PLS estimates factors, using not only predictors but also a target variable, which can explain significant forecasting improvement of PLS.

To reconfirm the forecasting performance of PLS1, I analyze the forecasting power of all factor-augmented forecasts by information criteria more in detail. Table B.13 and B.14 are 25, 50, and 75 percentiles of mean and minimum RMSE of each factor-augmented forecast across different information criteria, in a given category, respectively. The two tables are generated by following procedure. First, the entire 144 target variables are forecasted by 7 factor estimation methods with 11 information criteria. Then for the respective factor estimation method, the mean and minimum RMSE over different information criteria are calculated. If the factor estimation method involves some parameter choice such as λ_2 of Targeted Predictors, the mean and minimum over these specifications are also considered. Therefore, one RMSE is calculated for the given factor estimation methods for Table B.13 (mean) and Table B.14 (minimum), respectively, for a target variable. Then the entire 144 forecasts are divided into the eight categories according to their target variables. Finally, the 25, 50 and 75 percentiles of forecast distribution of each category are presented.

In this sense, Table B.13 is the expected RMSE of a respective factor estimation method for various target variables, for a given percentile. To reconfirm the findings from previous tables, two ranges are reported: the first is range (the difference between the best and worst forecast), excluding PLS with other information criteria ('PLS, other IC'). This is because PLS with other information criteria tend to have high RMSE on average. The second is the range of all factor estimation methods, including PLS with other information criteria. On the other hand, Table B.14 shows the possible minimum RMSE or the best forecasting results of the factor estimation method for all the target, for a given percentile.

The above findings are also supported in these two tables. First, when we consider the best possible forecasts by all the factor-augmented methods, the range is very small as denoted by Table B.14. The factor estimation methods that gives the lowest RMSE and in bold is rather dispersed, such that all the seven factor estimation methods achieve the lowest RMSE at least once. Second, when we consider the average or expected forecasts in Table B.13, the range becomes significantly larger than that of Table B.14, even excluding PLS with other IC forecasts. It is noteworthy that PLS1 tends to give the best results most frequently and shows the largest predictive improvement, across different categories and percentiles. As discussed above, PLS1 have the identical values in Table B.13 and B.14. Therefore, these tables show that PLS1 forecasts often gives the forecasting performance which is very close to the best results based on all the 7 factor estimation methods with 11 information criteria.

Given this conclusion, more detailed description about each factor estimation method will be discussed in the following sections.

2.4.3 PCA

The Table B.15 to B.20 reconfirm some important results from Stock and Watson (2002b). First, the predictive gain of using lagged target variables is not significant for real variables. The forecasting improvement in DI and DIAR forecasts are very similar or DI forecasts even show slightly lower RMSE than DIAR for some real variables. On the other hand, predictive improvement of incorporating lagged target variable is notably drastic for price-related variables. This finding is consistent with Stock and Watson (2002b).

Consistent with Stock and Watson (2002b), this experiment also finds that the forecasting improvement of DIAR-LAG upon DIAR is not significant for all the important 8 variables. As they point out, this result may indicate that all the predictive dynamics of target variables are explained by the estimated factors. Especially for the four price-related variables, forecasts under DIAR with $k = 1$ tend to be even slightly better than the best forecasts of DIAR-LAG.[15]

It is noteworthy that overall, PCA forecasts are relatively robust to the number of factors used, k, compared to PLS. Given lagged factors or target variables, incorporating more contemporaneous factors does not deteriorate forecasting power drastically as PLS. However, the predictive performance of PCA forecasts still tend to vary significantly with number of contemporaneous factors used in forecasting.

Predictive improvements of all available information criteria for PCA are similar for many cases. However, the forecasting power of PCA-augmented forecasts still

---

[15] There is a predictive gain of DIAR-LAG forecast compared to DIAR with $k = 1$ in forecasting horizon of $h = 24$, but the improvement is not significantly large.

varies across information criteria, for certain target variables and specifications. For instance, predictive power across different information criteria varies significantly in forecasting price-related variables in DI specification. Due to this reason, choice of appropriate information criteria for each target variable is important. However, as Table B.21 to B.26 show, it may be hard to find a information criterion for PCA that dominates in all categories and forecasting equations, DI, DIAR and DIAR-LAG.

2.4.4 PLS

Table B.27 to B.32 show 12-month-ahead PLS-augmented forecasts by given k, for the eight important variables and whole 144 target variables by categories and forecasting models, DI, DIAR and DIAR-LAG. For DI forecasts, the number of contemporaneous factors in forecasting model, k, that gives ex-post best forecasting result is smaller for the real variables compared to the nominal variables. For instance, DI PLS1 forecast gives the lowest RMSE for the real variables. However, DI PLS forecasts often require more PLS factors to forecast the price-related variables better.

However, PLS forecasts in DIAR and DIAR-LAG forecasting model show that a lot less PLS factors, k, are needed to achieve the lowest RMSE even for the nominal variables. Especially PLS1 usually yields the best forecasting improvement for both four real and nominal variables. Even when PLS1 does not give the best predictive gain for some target variables in some forecasting, the improvement of PLS-augmented forecast with more k is not significant from that of PLS1. This is also consistent with empirical findings of Ahn and Bae (2020), which find PLS1 often gives the best forecasting outcome in empirical data.

Consistent with the previous findings, incorporating more PLS factors in forecasting equation deteriorates forecasting power after a certain point. It implies that RMSE of the PLS forecasts, regardless of forecasting models such as DI, DIAR and

DIAR-LAG, increases after certain k. This tendency with lower forecasting gain from more factors is prevalent in other factor estimations as well. However, the forecasting loss resulted from incorporating more than the necessary factors in forecasting is more serious for PLS. The predictive loss from incorporating many PLS factors is more significant for the real variables. For nominal variables, RMSE does not increase with k as drastically as the real variables.

Table B.33 to B.38 compare the predictive power of PLS1 and PLS forecasts with other information criteria. There are several features of PLS forecasts by information criteria, which is distinct from other factor-augmented forecasts with the same criteria. First, unlike other factor forecasts, the 'mean' of those PLS forecasts by IC often tend to be far from 'Best, Given k' forecast. It implies that PLS may be very sensitive to the choice of information criteria. For instance, BIC ($\hat{k}_{BIC}$), BN estimator ($\hat{k}_{BN}$) or ABC estimator ($\hat{k}_{ABC}$), often yield RMSE larger than one, which indicates that PLS forecasts with those information criteria are worse than a naïve mean forecasts. This tendency is more obvious for the real variables. The AH and ON estimator ($\hat{k}_{AH}$, $\hat{k}_{ON}$) gives similar forecasts with PLS1. This is because the number of factors estimated by those estimators are often one, at most two.

Second, overall, PLS1 often yields the best forecast result among all the information criteria. For other factor estimation such as PCA forecasts, the number of PCA factors k that gives the ex-post best predictive outcome varies across forecasting model, forecasting target and forecasting horizons. It is noteworthy that PLS-augmented forecasts usually tend to achieve the best predictive gain when k=1, which is PLS1. Due to this reason, PLS1 forecast is usually identical to 'Best, Given k' forecast. Even when other information criteria yield better predictive performance than PLS1, the predictive gain is often marginal and forecast improvement of the two tend to be similar.

It is important to determine the best information criteria for PLS-augmented forecasting, but another important issue is whether PLS1 has better predictive performance than the traditional factor-augmented forecasting, PCA. Since PCA is one of the most popular factor estimation techniques used for forecasting, PCA-augmented forecast is set as the benchmark. Given this, I investigate whether PLS1 has the predictive gain over PCA-augmented forecasts.

Table B.39 to B.44 compare the forecasting performances of PLS1 and PCA forecasts by given k. Results from these tables do not involve information criteria for PCA forecasts, and compare forecasting power of PLS1 with that of PCA1, PCA2, until PCA12. The lowest RMSE among PCA forecasts with given k is in bold. When PLS1 has lower RMSE than any of the PCA forecasts, the PLS1 forecast is shaded with green, underlined and bold italic. It is noteworthy that even though PCA forecast with certain k gives the lowest RMSE, it is the number of factors that show ex-post best result. The number of factors that produces the ex-post best PCA forecasts vary across target variables, forecasting models and forecasting horizons. However, PLS1 does not involve choice of k, because k=1 is used for all time series. Considering this difference, results of Table B.39 to B.44 are very encouraging. PLS1 often shows significantly lower RMSE than the ex-post best PCA forecasts.

Next, I compare PLS1 with the PCA forecasts obtained by information criteria in Table B.45 to B.50. These tables show that PLS1 still tend to give better forecasting performance than the best PCA forecasts chosen across different information criteria. Table B.51 and B.52 also support this finding, in more detailed and intensive comparison. Table B.51 show PLS 1 and PCA k, $k = 1, 2, ..., 12$ for the whole 148 target variables in 6-, 12- and 24-month ahead forecasting.[16] Table B.52 compares PLS 1 and

---

[16] 144 target variables for 6- and 12-month ahead forecasting, and 148 for 24-month ahead forecasting.

PCA with information criteria, from PCA : BIC ($\hat{k}_{BIC}$) to PCA : ABC-S ($\hat{k}_{ABC-S}$) in all the three forecasting horizons. The frequency counts how often the given method achieved the minimum RMSE in 436 variable-horizon combinations. The percentage is the frequency divided by 436 and multiplied by 100. Table B.51 demonstrates that in this intensive comparison, PLS1 outperforms other PCA-augment forecasts with given $k = 1, 2, ..., 12$, around 200 times out of 436 combinations, which accounts for around 50%. It implies that PLS1 outperformed PCA1, PCA2, ..., PCA12 in this intensive comparison, around 50%. Table B.52 also shows strong predictive power of PLS1. PLS1 dominates PCA-augmented forecasts with 9 information criteria, around 180 times out of 436 combinations. These results support stable and significant predictive improvement of PLS, since PLS1 alone can dominate 12 PCA with given k and 9 PCA with information criteria, about 50%. Even when some PCA combinations outperform PLS1, the improvement is often small, as Table B.39 to B.50 show.

2.4.5 Other Factor Estimation Methods

First, the targeted predictor forecasts are similar across different λ_2, given that forecasts in 'Mean' column are similar across λ_2, in Table B.1 to B.12. However, the dominant information criteria that gives the lowest RMSE tend to vary across λ_2, as Table B.5 to B.8 show.

Second, Rule SWa and SWb generally outperform Rule B in Weighted PC forecasts. Table B.1 to B.12 supports this finding: even for forecasting the price-related variables, Rule B is often dominated by either SWa or SWb. Even when Rule B outperforms the two, the predictive improvement tend to be often marginal.

Finally, Table B.5 to B.8 indicate that the mean of static One-sided forecasts by information criteria, denoted by 'mean' columns, often has higher RMSE than the dynamic forecasts. It implies that dynamic One-sided forecasts tend to have better predictive power than the average static One-sided forecasts by information criteria. However, the best performance of static One-sided forecasts by information criteria, RMSE in bold in Table B.5 to B.8, may outperform the dynamic One-sided forecasts. This result shows that if one chooses the ex-post best performing information criteria, then a simple static One-sided factor estimation can yield a better forecasting performance than dynamic One-sided forecasts. This tendency is also obvious in Table B.9 to B.12.

2.5 Conclusion

This paper investigates the empirical predictive performance of factor-augmented forecasts whose factors are estimated using various, widely used factor estimation methods. Using 7 factor estimation methods, 148 target variables are forecasted, according to 11 information criteria, under three forecasting horizons and three forecasting equations. This intensive analysis contributes to the existing literature by providing several new findings.

First, the number of contemporaneous factors used in forecasting is important for predictive power. Incorporating more factors in the forecasting equation may not always produce better forecasting results. Rather, the use of more than a certain number of factors tend to deteriorate forecasting performance. Consistency of certain estimated factors to the true factor space has been proven by many studies. This paper contributes to the existing literature by providing an empirical evidence that forecasting performance often deteriorate when more factors are incorporated in the forecasting equation.

Second, I find that consistently estimated number of factors in the predictors, may not necessarily lead to the best predictive performance in the actual forecasting practice. More specifically, 11 information criteria that estimate the number of factors, are tested. Usually, many factor estimation methods give good predictive results when information criteria are used, except for PLS. PLS1, which is not consistent to the true factor space relevant for a target variable, often offers a better predictive power. PLS-augmented forecasts with other information criteria often tend to show worse forecasting performance.

Third, the best forecasting performance of each factor estimation method across the 11 different information criteria is very similar. However, it is often hard to find a dominant information criterion that gives the best forecasting results for all the target variables, except for PLS1 for PLS. Even when the same factor estimation method is used, predictive power tend to vary significantly across information criteria. Therefore, the choice of factor estimation method, as well as information criterion is important for forecasting performance.

Finally, PLS1 very often produces the predictive power that is close to the best forecasting performance of the total combinations of all the factor estimation methods with all information criteria. PLS1 tend to show robust predictive power and often outperforms other factor estimation methods that involve more discretion about the number of factors and parameter values. When we consider the best forecasts of the whole combinations of 7 factor estimation methods with 11 information criteria, PLS1 generally provides the forecasts very close to the best results in many target variables.

I finish this paper by addressing possible future research. The possible usage of PLS-augmented forecasting in other economic and financial contexts can be considered. First, it can be used to construct macroeconomic uncertainty. Uncertainty is usually defined as the conditional volatility of a disturbance that was not predictable. Since PLS often yields lower prediction errors, one may eliminate predictable movement more efficiently with PLS-augmented forecasting when uncertainty measures are constructed. Second, forecasting with PLS can be used for the market return and cash flow prediction. More specifically, the forecasting performance of various factor-augmented forecasting, including PLS, along with the predictive variables proposed by literature, can be investigated.

REFERENCES

Ahn, S. C. and J. Bae, “Forecasting with partial least squares when a large number of predictors are available”, Working paper (2020).

Ahn, S. C. and A. R. Horenstein, “Eigenvalue ratio test for the number of factors”, Econometrica **81**, 3, 1203–1227 (2013).

Akaike, H., “Fitting autoregressive models for prediction”, Annals of the institute of Statistical Mathematics **21**, 1, 243–247 (1969).

Alessi, L., M. Barigozzi and M. Capasso, “Improved penalization for determining the number of factors in approximate factor models”, Statistics & Probability Letters **80**, 23-24, 1806–1813 (2010).

Anderson, T. W. *et al.*, “Asymptotic theory for principal component analysis”, Annals of Mathematical Statistics **34**, 1, 122–148 (1963).

Bai, J., “Inferential theory for factor models of large dimensions”, Econometrica **71**, 1, 135–171 (2003).

Bai, J. and S. Ng, “Determining the number of factors in approximate factor models”, Econometrica **70**, 1, 191–221 (2002).

Bai, J. and S. Ng, “Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions”, Econometrica **74**, 4, 1133–1150 (2006).

Bai, J. and S. Ng, “Determining the number of primitive shocks in factor models”, Journal of Business & Economic Statistics **25**, 1, 52–60 (2007).

Bai, J. and S. Ng, “Forecasting economic time series using targeted predictors”, Journal of Econometrics **146**, 2, 304–317 (2008a).

Bai, J. and S. Ng, *Large dimensional factor analysis* (Now Publishers Inc, 2008b).

Bernanke, B. S., J. Boivin and P. Eliasz, “Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach”, The Quarterly journal of economics **120**, 1, 387–422 (2005).

Boivin, J. and S. Ng, “Understanding and comparing factor-based forecasts”, Tech. rep., National Bureau of Economic Research (2005).

Boivin, J. and S. Ng, “Are more data always better for factor analysis?”, Journal of Econometrics **132**, 1, 169–194 (2006).

Butler, N. A. and M. C. Denham, “The peculiar shrinkage properties of partial least squares regression”, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **62**, 3, 585–593 (2000).

Carrasco, M. and B. Rossi, “In-sample inference and forecasting in misspecified factor models”, Journal of Business & Economic Statistics **34**, 3, 313–338 (2016).

Connor, G. and R. A. Korajczyk, “Performance measurement with the arbitrage pricing theory: A new framework for analysis”, Journal of Financial Economics (JFE) **15**, 3 (1986).

Dax, A., “The numerical rank of krylov matrices”, Linear Algebra and Its Applications **528**, 185–205 (2017).

De Mol, C., D. Giannone and L. Reichlin, “Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components?”, Journal of Econometrics **146**, 2, 318–328 (2008).

Doz, C., D. Giannone and L. Reichlin, “A two-step estimator for large approximate dynamic factor models based on kalman filtering”, Journal of Econometrics **164**, 1, 188–205 (2011).

Doz, C., D. Giannone and L. Reichlin, “A quasi–maximum likelihood approach for large, approximate dynamic factor models”, Review of economics and statistics **94**, 4, 1014–1024 (2012).

D’Agostino, A. and D. Giannone, “Comparing alternative predictors based on large-panel factor models”, Oxford bulletin of economics and statistics **74**, 2, 306–326 (2012).

Eaton, M. L. and D. E. Tyler, “On wielandt’s inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix”, The Annals of Statistics pp. 260–271 (1991).

Engle, R. and M. Watson, “A one-factor multivariate time series model of metropolitan wage rates”, Journal of the American Statistical Association **76**, 376, 774–781 (1981).

Fang, C. and P. Krishnaiah, “Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations”, Journal of Multivariate Analysis **12**, 1, 39–63 (1982).

Forni, M., M. Hallin, M. Lippi and L. Reichlin, “The generalized dynamic-factor model: Identification and estimation”, Review of Economics and statistics **82**, 4, 540–554 (2000).

Forni, M., M. Hallin, M. Lippi and L. Reichlin, “The generalized dynamic factor model: one-sided estimation and forecasting”, Journal of the American Statistical Association **100**, 471, 830–840 (2005).

Forni, M. and L. Reichlin, “Let’s get real: a factor analytical approach to disaggregated business cycle dynamics”, The Review of Economic Studies **65**, 3, 453–473 (1998).

Giannone, D., L. Reichlin and D. Small, “Nowcasting: The real-time informational content of macroeconomic data”, Journal of Monetary Economics **55**, 4, 665–676 (2008).

Giglio, S., B. Kelly and S. Pruitt, “Systemic risk and the macroeconomy: An empirical evaluation”, Journal of Financial Economics **119**, 3, 457–471 (2016).

Groen, J. J. and G. Kapetanios, “Revisiting useful approaches to data-rich macroeconomic forecasting”, FRB of New York Staff Report , 327 (2009).

Groen, J. J. and G. Kapetanios, “Revisiting useful approaches to data-rich macroeconomic forecasting”, Computational Statistics & Data Analysis **100**, 221–239 (2016).

Hallin, M. and R. Liška, “Determining the number of factors in the general dynamic factor model”, Journal of the American Statistical Association **102**, 478, 603–617 (2007).

Hastie, T., R. Tibshirani and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction* (Springer Science & Business Media, 2009).

Helland, I. S., “On the structure of partial least squares regression”, Communications in statistics-Simulation and Computation **17**, 2, 581–607 (1988).

Helland, I. S., “Partial least squares regression and statistical models”, Scandinavian Journal of Statistics pp. 97–114 (1990).

Helland, I. S. and T. Almøy, “Comparison of prediction methods when only a few components are relevant”, Journal of the American Statistical Association **89**, 426, 583–591 (1994).

Huang, D., F. Jiang, J. Tu and G. Zhou, “Investor sentiment aligned: A powerful predictor of stock returns”, The Review of Financial Studies **28**, 3, 791–837 (2015).

Huber, P. J. *et al.*, “Robust regression: asymptotics, conjectures and monte carlo”, Annals of statistics **1**, 5, 799–821 (1973).

Jones, C. S., “Extracting factors from heteroskedastic asset returns”, Journal of Financial economics **62**, 2, 293–325 (2001).

Jong, S. D., “Pls fits closer than pcr”, Journal of chemometrics **7**, 6, 551–557 (1993).

Kelly, B. and S. Pruitt, “Market expectations in the cross-section of present values”, The Journal of Finance **68**, 5, 1721–1756 (2013).

Kelly, B. and S. Pruitt, “The three-pass regression filter: A new approach to forecasting using many predictors”, Journal of Econometrics **186**, 2, 294–316 (2015).

Li, B., J. Morris and E. B. Martin, “Model selection for partial least squares regression”, Chemometrics and Intelligent Laboratory Systems **64**, 1, 79–89 (2002).

Light, N., D. Maslov and O. Rytchkov, “Aggregation of information about the cross section of stock returns: A latent variable approach”, The Review of Financial Studies **30**, 4, 1339–1381 (2017).

McCracken, M. W. and S. Ng, “Fred-md: A monthly database for macroeconomic research”, Journal of Business & Economic Statistics **34**, 4, 574–589 (2016).

Mehmood, T., K. H. Liland, L. Snipen and S. Sæbø, “A review of variable selection methods in partial least squares regression”, Chemometrics and Intelligent Laboratory Systems **118**, 62–69 (2012).

Moon, H. R. and M. Weidner, “Linear regression for panel with unknown number of factors as interactive fixed effects”, Econometrica **83**, 4, 1543–1579 (2015).

Naes, T. and I. S. Helland, “Relevant components in regression”, Scandinavian journal of statistics pp. 239–250 (1993).

Naes, T. and H. Martens, “Comparison of prediction methods for multicollinear data”, Communications in Statistics-Simulation and Computation **14**, 3, 545–576 (1985).

Onatski, A., “Determining the number of factors from empirical distribution of eigenvalues”, The Review of Economics and Statistics **92**, 4, 1004–1016 (2010).

Onatski, A., “Asymptotics of the principal components estimator of large factor models with weakly influential factors”, Journal of Econometrics **168**, 2, 244–258 (2012).

Phatak, A. and F. de Hoog, “Exploiting the connection between pls, lanczos methods and conjugate gradients: alternative proofs of some properties of pls”, Journal of Chemometrics: A Journal of the Chemometrics Society **16**, 7, 361–367 (2002).

Quah, D. and T. J. Sargent, “A dynamic index model for large cross sections”, in “Business cycles, indicators, and forecasting”, pp. 285–310 (University of Chicago Press, 1993).

Rorni, M. and L. Reichlin, “Dynamic common factors in large cross-sections”, Empirical economics **21**, 1, 27–42 (1996).

Rytchkov, O. and X. Zhong, “Information aggregation and p-hacking”, Management Science **66**, 4, 1605–1626 (2020).

Schwarz, G. *et al.*, “Estimating the dimension of a model”, Annals of statistics **6**, 2, 461–464 (1978).

Stein, C., “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution”, Tech. rep., Stanford University Stanford United States (1956).

Stewart, G. and J.-G. Sun, “Matrix perturbation theory academic press”, San Diego (1990).

Stock, J. H. and M. Watson, “Dynamic factor models”, Oxford Handbooks Online (2011).

Stock, J. H. and M. W. Watson, “New indexes of coincident and leading economic indicators”, NBER macroeconomics annual **4**, 351–394 (1989).

Stock, J. H. and M. W. Watson, “Diffusion indexes”, NBER working paper , w6702 (1998).

Stock, J. H. and M. W. Watson, “Forecasting inflation”, Journal of Monetary Economics **44**, 2, 293–335 (1999).

Stock, J. H. and M. W. Watson, “Forecasting using principal components from a large number of predictors”, Journal of the American statistical association **97**, 460, 1167–1179 (2002a).

Stock, J. H. and M. W. Watson, “Macroeconomic forecasting using diffusion indexes”, Journal of Business & Economic Statistics **20**, 2, 147–162 (2002b).

Stock, J. H. and M. W. Watson, “Forecasting with many predictors”, Handbook of economic forecasting **1**, 515–554 (2006).

Stock, J. H. and M. W. Watson, “Dynamic factor models”, (2010).

Stock, J. H. and M. W. Watson, “Generalized shrinkage methods for forecasting using many predictors”, Journal of Business & Economic Statistics **30**, 4, 481–493 (2012).

Stock, J. H. and M. W. Watson, “Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics”, in “Handbook of macroeconomics”, vol. 2, pp. 415–525 (Elsevier, 2016).

Taniguchi, M. and P. Krishnaiah, “Asymptotic distributions of functions of the eigenvalues of sample covariance matrix and canonical correlation matrix in multivariate time series”, Journal of multivariate analysis **22**, 1, 156–176 (1987).

Tu, Y. and T.-H. Lee, “Forecasting using supervised factor models”, Journal of Management Science and Engineering **4**, 1, 12–27 (2019).

Watson, M. W. and R. F. Engle, “Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models”, Journal of Econometrics **23**, 3, 385–400 (1983).

White, H., “Maximum likelihood estimation of misspecified models”, Econometrica: Journal of the Econometric Society pp. 1–25 (1982).

Wold, H., “A. 1985. partial least squares”, Kotz S, Johnson N L. Encyclopedia of Statistical Sciences. New York: Wiley pp. 581–591 (????).

Wold, H., “Estimation of principal components and related models by iterative least squares”, Multivariate analysis pp. 391–420 (1966).

Wold, H., “Nonlinear iterative partial least squares (nipals) modelling: some current developments”, in “Multivariate Analysis–III”, pp. 383–407 (Elsevier, 1973).

Wold, H., “Soft modelling by latent variables: the non-linear iterative partial least squares (nipals) approach”, Journal of Applied Probability **12**, S1, 117–142 (1975).

Wold, H., “Soft modeling: the basic design and some extensions”, Systems under indirect observation **2**, 343 (1982).

Yu, Y., T. Wang and R. J. Samworth, “A useful variant of the davis–kahan theorem for statisticians”, Biometrika **102**, 2, 315–323 (2015).

Zou, H. and T. Hastie, “Regularization and variable selection via the elastic net”, Journal of the royal statistical society: series B (statistical methodology) **67**, 2, 301–320 (2005).

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER 1

A.1 NIPLS algorithm

Let $\hat{\boldsymbol{P}}_{1:q} = (\hat{\boldsymbol{p}}_1, ..., \hat{\boldsymbol{p}}_q)$ be the $T \times q$ matrix of the first q PLS factors from the NIPLS algorithm. Set $\boldsymbol{X}^{(1)} = \boldsymbol{X}$, $\hat{\boldsymbol{\alpha}}_1 = \boldsymbol{X}^{(1)'}\boldsymbol{y}$, and $\hat{\boldsymbol{p}}_1 = \boldsymbol{X}^{(1)}\hat{\boldsymbol{\alpha}}_1$. For j= 2, ..., q, we iteratively create

$$\begin{aligned}
\hat{\boldsymbol{\psi}}^{(j-1)} &= \boldsymbol{X}^{(j-1)'}\hat{\boldsymbol{p}}_{j-1}(\hat{\boldsymbol{p}}'_{j-1}\hat{\boldsymbol{p}}_{j-1})^{-1};\\
\boldsymbol{X}^{(j)} &= \boldsymbol{X}^{(j-1)} - \hat{\boldsymbol{p}}_{j-1}\hat{\boldsymbol{\psi}}^{(j-1)'};\\
\hat{\boldsymbol{\alpha}}_j &= \boldsymbol{X}^{(j)'}\boldsymbol{y};\\
\hat{\boldsymbol{p}}_j &= \boldsymbol{X}^{(j)}\hat{\boldsymbol{\alpha}}_j
\end{aligned}$$

By construction, the PLS factor vectors $\hat{\boldsymbol{p}}_j$ are mutually orthogonal. For forecasting y_{T+2}, the values of the PLS factors at time $(T{+}1)$ needs to be predicted. Let $\hat{\boldsymbol{\delta}}_{1:q}$ be the OLS estimator from a regression of $\boldsymbol{y}$ on $\hat{\boldsymbol{P}}_{1:q}$; and let $\boldsymbol{x}^{(1)}_{\boldsymbol{\cdot}T+1} = \boldsymbol{x}_{\boldsymbol{\cdot}T+1}$, $\hat{p}_{1,T+1} = \boldsymbol{x}^{(1)}_{\boldsymbol{\cdot}T+1}{}' \hat{\alpha}_1$, and

$$\boldsymbol{x}^{(j)}_{\boldsymbol{\cdot}T+1} = \boldsymbol{x}^{(j-1)}_{\boldsymbol{\cdot}T+1} - \hat{\boldsymbol{\psi}}^{(j)}\hat{p}_{j-1,T+1}; \quad \hat{p}_{j,T+1} = \boldsymbol{x}^{(j)}_{\boldsymbol{\cdot}T+1}{}'\hat{\boldsymbol{\alpha}}_j$$

Then, the PLS forecast of y_{T+2} using the first q PLS factors is $\hat{y}^{PLS}_{T+2|q} = \hat{\boldsymbol{\delta}}'{}_{1:q}\hat{\boldsymbol{p}}_{1:q,T+1}$, where $\hat{\boldsymbol{p}}_{1:q,T+1} = (\hat{p}_{1,T+1}, ..., \hat{p}_{q,T+1})'$.

A.2 Notation and Preliminary Lemmas

All of the asymptotic results in this appendix are obtained as $N, T \to \infty$ jointly. We use some additional notation. First, the vector notation $\mathbf{1}_l$ denotes an $l \times 1$ vector of ones, while $\boldsymbol{I}_l$ denotes an $l \times l$ identity matrix. For the matrices, $\boldsymbol{A}_1$, ... , $\boldsymbol{A}_l$, that are any size,

$$\boldsymbol{Diag}(\boldsymbol{A}_1, \boldsymbol{A}_2, ..., \boldsymbol{A}_l) = \begin{pmatrix} \boldsymbol{A}_1 & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_2 & ... & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & ... & \boldsymbol{A}_l \end{pmatrix}$$

where the "$\mathbf{0}$" matrices are conformable zero matrices. Notice that $\boldsymbol{Diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_l)$ is not a square matrix because all of the matrices $\boldsymbol{A}_1$, ... , $\boldsymbol{A}_l$ are square matrices. We use the more common notation $\boldsymbol{diag}(\boldsymbol{A}_1, ..., \boldsymbol{A}_l)$ if all of $\boldsymbol{A}_1$, ... , $\boldsymbol{A}_l$ are square matrices or scalars. In Appendix A.2, n denotes some increasing integer functions of N and/or T.

The following lemmas are useful to prove the theorems in this paper.

Lemma B.1 (Theorem 2 of Yu, Wang and Samworth (2015)): Let $\boldsymbol{B}$ and $\boldsymbol{A} \in \mathbb{R}^{l\times l}$ be symmetric matrices. Choose two integers a and b such that $1 \leq a \leq b \leq l$. Assume that

$$\min\{\lambda_{a-1}(\boldsymbol{A}) - \lambda_a(\boldsymbol{A}), \lambda_b(\boldsymbol{A}) - \lambda_{b+1}(\boldsymbol{A})\} > 0$$

where we set $\lambda_0(\boldsymbol{A}) = \infty$ and $\lambda_{l+1}(\boldsymbol{A}) = -\infty$. Let $d = b - a + 1$. Then, there exists an orthonormal matrix $\boldsymbol{O}^{\boldsymbol{B}} \in \mathbb{R}^{d\times d}$ such that

$$\left\|\Xi(\boldsymbol{B}|a:b)\boldsymbol{O}^{\boldsymbol{B}} - \Xi(\boldsymbol{A}|a:b)\right\|_F \leq \frac{2^{3/2}\min\left\{d^{1/2}\|\boldsymbol{B}-\boldsymbol{A}\|_2, \|\boldsymbol{B}-\boldsymbol{A}\|_F\right\}}{\min\{\lambda_{a-1}(\boldsymbol{A}) - \lambda_a(\boldsymbol{A}), \lambda_b(\boldsymbol{A}) - \lambda_{b+1}(\boldsymbol{A})\}}$$

Remarks on Lemma B.1: (1) Let $\boldsymbol{B}$ and $\boldsymbol{A}$ be $l \times l$ symmetric random matrices, where l is a fixed positive integer or an increasing integer function of n. Suppose that $p\lim_{m\to\infty}\lambda_1(\boldsymbol{A}) = p\lim_{m\to\infty}\lambda_2(\boldsymbol{A}) > p\lim_{m\to\infty}\lambda_3(\boldsymbol{A})$, and that $\|\boldsymbol{B}-\boldsymbol{A}\|_2 = O_p(n^{-\varsigma})$. If we choose $a = 1$ and $b = 2$ for the above lemma, we can obtain the following result:

$$\begin{aligned}\left\|\Xi(\boldsymbol{B}|1:2)\boldsymbol{O}^{2\times 2} - \Xi(\boldsymbol{A}|1:2)\right\|_F &\leq \frac{2^2\|\boldsymbol{B}-\boldsymbol{A}\|_2}{\min\{\lambda_0(\boldsymbol{A}) - \lambda_1(\boldsymbol{A}), \lambda_2(\boldsymbol{A}) - \lambda_3(\boldsymbol{A})\}} \\ &= \frac{4\|\boldsymbol{B}-\boldsymbol{A}\|_2}{\lambda_2(\boldsymbol{A}) - \lambda_3(\boldsymbol{A})} = O_p(n^{-\varsigma})\end{aligned}$$

for some orthonormal matrix $\boldsymbol{O}_{2\times 2} \in \mathbb{R}^{2\times 2}$.

(2) An important implication of the lemma is that when some eigenvalues of a random matrix have the same probability limits, the eigenvectors corresponding to the eigenvalues are asymptotically unique up to an orthonormal transformation. The lemma explains why the PLS method cannot identify what individual factors in $\boldsymbol{f}_{(j)t}$ are correlated or uncorrelated with the target variable.

Lemma B.2: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $l \times l$ invertible matrices. Then,

$$\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1} = \boldsymbol{B}^{-1}\left(\boldsymbol{A} - \boldsymbol{B}\right)\boldsymbol{A}^{-1}$$

Lemma B.3: Let $\boldsymbol{B}$ and $\boldsymbol{A}$ be $l \times l$ symmetric matrices, where l is a fixed positive integer or an increasing integer function of n. Suppose that $\|\boldsymbol{B}-\boldsymbol{A}\|_2 = O_p(n^{-\varsigma})$. Then, for all $h = 1, ..., l$, $\lambda_h(\boldsymbol{B}) = \lambda_h(\boldsymbol{A}) + O_p(n^{-\varsigma})$.

Proof: Using Corollary 4.10 of Stewart and Sun (1990, p. 203), we have

$$|\lambda_h(\boldsymbol{B}) - \lambda_h(\boldsymbol{A})| \leq \max\left\{|\lambda_1(\boldsymbol{B}-\boldsymbol{A})|, |\lambda_l(\boldsymbol{B}-\boldsymbol{A})|\right\} = \|\boldsymbol{B}-\boldsymbol{A}\|_2 \qquad Q.E.D.$$

Lemma B.4: Let $\boldsymbol{B}$ and $\boldsymbol{A}$ be $l \times l$ symmetric random matrices, where l is a fixed positive integer or an increasing integer function of n. Define fixed integers K and $k(j)$ $(j = 0, 1, ..., J)$ such that $k(0) = 0$ and $\Sigma_{j=1}^{J}k(j) = K$. Let $ks(j) = \Sigma_{h=1}^{j}k(h)$. Assume that $\lambda_h(\boldsymbol{A}) = \sigma_j^2 + O_p(n^{-\varsigma})$ for $h = ks(j-1)+1, \ldots, k(j)$ and $\sigma_1^2 > \sigma_2^2 > ... > \sigma_J^2$. Let $\Xi_{(j)}^{\boldsymbol{A}} = \Xi(\boldsymbol{A}|ks(j-1)+1 : ks(j))$; and define $\Xi_{(j)}^{\boldsymbol{B}}$ similarly for $\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{C}$. Suppose that $\|\boldsymbol{C}\|_2 = O_p(n^{-\varsigma})$. Then, for each $j = 1, ..., J$, there exists a $k(j) \times k(j)$ matrix $\boldsymbol{O}_{jj}^{\boldsymbol{B}}$ such that $\left\|\Xi_{(j)}^{\boldsymbol{B}}\boldsymbol{O}_{jj}^{\boldsymbol{B}} - \Xi_{(j)}^{\boldsymbol{A}}\right\|_F = O_p(n^{-\varsigma})$.

Proof: Let $a = ks(j-1)+1$ and $b = ks(j)$, such that $b - a + 1 = k(j)$. Let $a' = ks(j-1)$ and $b' = ks(j) + 1$. Then, by Lemma B.1,

$$\left\|\Xi^{B}_{(j)}O^{B}_{jj} - \Xi^{A}_{(j)}\right\|_F \le \frac{2^{3/2}\min\{(k(j))^{1/2}\|C\|_2, \|C\|_F\}}{\min\{\lambda_{a'}(A) - \lambda_a(A), \lambda_b(A) - \lambda_{b'}(A)\}}$$
$$\le \frac{2^{3/2}(k(j))^{1/2}\|C\|_2}{\min\{\lambda_{a'}(A) - \lambda_a(A), \lambda_b(A) - \lambda_{b'}(A)\}}$$
$$= \frac{2^{3/2}(k(j))^{1/2}\|C\|_2}{\min\{\sigma^2_{j-1} - \sigma^2_j + O_p(n^{-\varsigma}), \sigma^2_j - \sigma^2_{j+1} + O_p(n^{-\varsigma})\}} = O_p(n^{-\varsigma})$$

which completes the proof. *Q.E.D*

Lemma B.5: Let B and A be $l_1 \times l_2$ random matrices, where l_2 is a fixed positive integer and and l_1 is a fixed positive integers or an increasing integer function of n. Assume that $\|B - A\|_F = O_p(n^{-\varsigma})$, and that $p\lim_{m\to\infty}A'A$ is finite and invertible. Then,

$$\|\mathcal{P}(B) - \mathcal{P}(A)\|_F = O_p(n^{-\varsigma}); \quad \|\mathcal{Q}(B) - \mathcal{Q}(A)\|_F = O_p(n^{-\varsigma})$$

Proof: Let $C = (B - A)'A + A'(B - A) + (B - A)'(B - A)$ so that $B'B = A'A + C$. Observe that $\|C\|_F = O_p(n^{-\varsigma})$. Thus, $p\lim_{m\to\infty}B'B$ is also finite and invertible. Therefore, by Lemma B.2,

$$\left\|(B'B)^{-1} - (A'A)^{-1}\right\| \le \left\|(B'B)^{-1}\right\| \left\|(A'A)^{-1}\right\| \|C\| = O_p(n^{-\varsigma})$$

Now, observe that

$$\begin{aligned}\mathcal{P}(B) - \mathcal{P}(A) &= A[(B'B)^{-1} - (A'A)^{-1}]A' + (B - A)(B'B)^{-1}A' \\ &\quad + A(B'B)^{-1}(B - A)' + (B - A)(B'B)^{-1}(B - A)' \\ &\equiv I + II + III + IV.\end{aligned}$$

Here, $\|I\| \le \|A\| \left\|(B'B)^{-1} - (A'A)^{-1}\right\| \|A'\| = O_p(n^{-\varsigma})$. Similarly, we can show $\|II\| = O_p(n^{-\varsigma})$; $\|III\| = O_p(n^{-\varsigma})$; and $\|IV\| = O_p(n^{-2\varsigma})$. Thus, $\|\mathcal{P}(B) - \mathcal{P}(A)\| \le \|I\| + \|II\| + \|III\| + \|IV\| = O_p(n^{-\varsigma})$. In addition, $\|\mathcal{Q}(B) - \mathcal{Q}(A)\| = O_p(n^{-\varsigma})$, because $\mathcal{Q}(B) - \mathcal{Q}(A) = \mathcal{P}(A) - \mathcal{P}(B)$. *Q.E.D.*

A.3 Proofs of Theorems

Lemma C.1: Define the following orthonormal matrix:

$$O^{\Omega_F} = \left(O^{\Omega_F}_{(1)}, ..., O^{\Omega_F}_{(J)}\right) = \begin{pmatrix} O^{\Omega_F}_{11} & 0_{k(1)\times k(2)} & \cdots & 0_{k(1)\times k(J)} \\ 0_{k(2)\times k(1)} & O^{\Omega_F}_{22} & \cdots & 0_{k(2)\times k(J)} \\ \vdots & \vdots & & \vdots \\ 0_{k(J)\times k(1)} & 0_{k(J)\times k(2)} & \cdots & O^{\Omega_F}_{JJ} \end{pmatrix}_{K\times K}$$
$$= Diag(O^{\Omega_F}_{11}, ..., O^{\Omega_F}_{JJ})$$

where $\boldsymbol{O}^{\boldsymbol{\Omega_F}}_{(j)}$ is a $K \times k(j)$ matrix and $\boldsymbol{O}^{\boldsymbol{\Omega_F}}_{jj}$ is a $k(j) \times k(j)$ orthonormal matrix for each $j = 1, ..., J$. Then, $\boldsymbol{O}^{\boldsymbol{\Omega_F}} = \Xi(\boldsymbol{\Omega_F}|1 : K)$.

Proof: The desired result holds because $\boldsymbol{O}^{\boldsymbol{\Omega_F}\prime}\boldsymbol{O}^{\boldsymbol{\Omega_F}} = \boldsymbol{I}_K$; $\boldsymbol{\Omega_F}\boldsymbol{O}^{\boldsymbol{\Omega_F}} = \boldsymbol{O}^{\boldsymbol{\Omega_F}}\boldsymbol{\Omega_F}$. *Q.E.D.*

Remark on Lemma C.1: The matrix $\boldsymbol{O}^{\boldsymbol{\Omega_F}}$is not unique because the $\boldsymbol{O}^{\boldsymbol{\Omega_F}}_{jj}$matrices could be any orthonormal matrices.

Lemma C.2: Under (A.2) – (A.4),

(i) $\left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}} - \boldsymbol{\Omega_F}\right\|_F = O_p(T^{-\gamma})$; $\quad \left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}} - \boldsymbol{I}_K\right\|_F = O_p(N^{-\gamma})$;

(ii) $\lambda_h(\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}) = \sigma_j^2 + O_p(T^{-\gamma})$; $\quad \lambda_q(\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}) = 1 + O_p(N^{-\gamma})$,

where $j = 1, \ ... \ , J$, $h = ks(j-1)+1, ks(j-1)+2, \ ... \ , ks(j)$, and $q = 1, \ 2, \ ... \ , \ K$.

Proof: Part (i) holds by (A.4). Observe that with (i),

$$\left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}} - \boldsymbol{\Omega_F}\right\|_2 \leq \left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}} - \boldsymbol{\Omega_F}\right\|_F = O_p(T^{-\gamma}); \quad \left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}} - \boldsymbol{I}_K\right\|_2 \leq \left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}} - \boldsymbol{I}_K\right\|_F = O_p(N^{-\gamma})$$

Thus, (ii) holds by Lemma B.3. *Q.E.D.*

Lemma C.3: Under (A.1) and (A.6),

(i) $\left\|(NT)^{-1/2}\boldsymbol{F'E}\right\|_F = O_p(1)$; $\left\|(NT)^{-1/2}\boldsymbol{\Phi' E'}\right\|_F = O_p(1)$; $\left\|(NT)^{-1/2}\boldsymbol{\Phi' E' F}\right\|_F = O_p(1)$;

(ii) $\left\|(NT)^{-1/2}\boldsymbol{E}\right\|_F = O_p(1) > 0$; $m^{1/2}\left\|(NT)^{-1}\boldsymbol{E'E}\right\|_F = O_p(1) > 0$.

Proof: We can show the first part of (i) by (A.6) because

$$\begin{aligned}
&\mathrm{E}\left((NT)^{-1}\left\|\boldsymbol{F'E}\right\|_F^2\right) \\
&\quad = \mathrm{E}\left((NT)^{-1}\mathrm{trace}(\boldsymbol{F'EE'F})\right) = \mathrm{E}\left((NT)^{-1}\mathrm{trace}\left[\Sigma_{i=1}^N\left(\Sigma_{t=1}^T\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right)\left(\Sigma_{t=1}^T\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right)'\right]\right) \\
&\quad = \mathrm{E}\left((NT)^{-1}\Sigma_{i=1}^N\mathrm{trace}\left[\left(\Sigma_{t=1}^T\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right)\left(\Sigma_{t=1}^T\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right)'\right]\right) \\
&\quad = \mathrm{E}\left(N^{-1}\Sigma_{i=1}^N\left\|T^{-1/2}\Sigma_{t=1}^T\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right\|_2^2\right) < c.
\end{aligned}$$

Similarly, the second part of (ii) can be shown because

$$\mathrm{E}\left((NT)^{-1}\left\|\boldsymbol{\Phi' E'}\right\|_F^2\right) = \mathrm{E}\left(T^{-1}\Sigma_{t=1}^T\left\|N^{-1/2}\Sigma_{i=1}^N\phi_{\boldsymbol{.}i}e_{it}\right\|_2^2\right) < c$$

The third part of (iii) holds because

$$\begin{aligned} \mathrm{E}\left((NT)^{-1}\left\|\boldsymbol{F'E\Phi}\right\|_F^2\right) &= N^{-1}\Sigma_{i=1}^{N}\mathrm{E}\left(\left\|T^{-1/2}\Sigma_{t=1}^{T}\boldsymbol{f}_{\boldsymbol{.}t}\phi'_{\boldsymbol{.}i}e_{it}\right\|_F^2\right) \\ &\leq N^{-1}\Sigma_{i=1}^{N}\mathrm{E}\left(\left\|T^{-1/2}\Sigma_{t=1}^{T}\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right\|_2^2\left\|\phi_{\boldsymbol{.}i}\right\|_2^2\right) \\ &= N^{-1}\Sigma_{i=1}^{N}\mathrm{E}\left(\left\|T^{-1/2}\Sigma_{t=1}^{T}\boldsymbol{f}_{\boldsymbol{.}t}e_{it}\right\|_2^2\right)\mathrm{E}\left(\left\|\phi_{\boldsymbol{.}i}\right\|_2^2\right) < c^2. \end{aligned}$$

Observe that because $\mathrm{rank}(\boldsymbol{E}) \leq \min\{N,T\} = m$, $\lambda_h(\boldsymbol{E'E}) = 0$ for all $h > m$. By this fact,

$$\begin{aligned} \left\|(NT)^{-1/2}\boldsymbol{E}\right\|_F^2 &= \mathrm{trace}\left((NT)^{-1}\boldsymbol{E'E}\right) = \Sigma_{h=1}^{m}\lambda_h\left((NT)^{-1}\boldsymbol{E'E}\right) \\ &= m^{-1}\Sigma_{h=1}^{m}\lambda_h\left(M^{-1}\boldsymbol{E'E}\right) \\ &\leq m^{-1}\times\left(m\times\lambda_1\left(M^{-1}\boldsymbol{E'E}\right)\right) = \lambda_1\left(M^{-1}\boldsymbol{E'E}\right); \end{aligned}$$

$$\begin{aligned} \left\|(NT)^{-1/2}\boldsymbol{E}\right\|_F^2 &= m^{-1}\Sigma_{h=1}^{m}\lambda_h\left(M^{-1/2}\boldsymbol{E'E}\right) \geq m^{-1}\left(m^c\times\lambda_{m^c}\left(M^{-1/2}\boldsymbol{E'E}\right)\right) \\ &= (m^c/m)\times\lambda_{m^c}\left(M^{-1/2}\boldsymbol{E'E}\right) = (m_c/m)(c+o_p(1)). \end{aligned}$$

These two results and (A.5) imply the first part of (ii). Finally, letting $\boldsymbol{A} = M^{-1}\boldsymbol{E'E}$, we can obtain the second part of (ii) because

$$\begin{aligned} m^{1/2}\left\|(NT)^{-1}\boldsymbol{E'E}\right\|_F &= m^{1/2}\left[m^{-2}\times\mathrm{trace}(\boldsymbol{AA})\right]^{1/2} = m^{1/2}\left[m^{-2}\Sigma_{h=1}^{m}(\lambda_h(\boldsymbol{A}))^2\right]^{1/2} \\ &\leq m^{1/2}\left[m^{-2}m\times(\lambda_1(\boldsymbol{A}))^2\right]^{1/2} = \lambda_1(\boldsymbol{A}); \end{aligned}$$

$$\begin{aligned} m^{1/2}\left\|(NT)^{-1}\boldsymbol{E'E}\right\|_F &= m^{1/2}\left[m^{-2}\times\mathrm{trace}(\boldsymbol{AA})\right]^{1/2} = \left[m^{-1}\Sigma_{h=1}^{m}(\lambda_h(\boldsymbol{AA}))^2\right]^{1/2} \\ &\geq \left[(m^c/m)\times(\lambda_{m^c}(\boldsymbol{A}))^2\right]^{1/2} \geq (m^c/m)^{1/2}(c+o_p(1))^{1/2}. \end{aligned}$$

This completes the proof. *Q.E.D.*

Lemma C.4: Let $\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} + \boldsymbol{E'F}(\boldsymbol{F'F})^{-1}$ and $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}} = N^{-1}\tilde{\boldsymbol{\Phi}}'\tilde{\boldsymbol{\Phi}}$. Then, under (A.1) – (A.6),

(i) $\left\|N^{-1/2}(\tilde{\boldsymbol{\Phi}} - \boldsymbol{\Phi})\right\|_F = O_p(T^{-1/2})$; $\left\|(\tilde{\Omega}_{\Phi} - \boldsymbol{I}_K)\right\|_F = O_p(m^{-\gamma})$;

(ii) $\left|\lambda_h(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}) - 1\right| = O_p(m^{-\gamma})$, for all $h = 1, 2, ..., K$;

(iii) $\left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2} - \boldsymbol{I}_K\right\|_F = O_p(m^{-\gamma})$.

Proof: The first part of (i) holds by Lemma C.3 because

$$\left\| N^{-1/2}(\tilde{\mathbf{\Phi}} - \mathbf{\Phi}) \right\|_F \leq T^{-1/2} \left\| (NT)^{-1/2} \boldsymbol{E'F} \right\|_F \left\| \left(T^{-1} \boldsymbol{F'F} \right)^{-1} \right\|_F = O_p(T^{-1/2})$$

For the second part of (ii), let

$$\boldsymbol{A} = \frac{1}{N^{1/2}T^{1/2}} \frac{\mathbf{\Phi}' \boldsymbol{E'F}}{N^{1/2}T^{1/2}} \left(\frac{\boldsymbol{F'F}}{T} \right)^{-1}; \quad \boldsymbol{B} = \frac{1}{T} \left(\frac{\boldsymbol{F'F}}{T} \right)^{-1} \frac{\boldsymbol{F'E}}{N^{1/2}T^{1/2}} \frac{\boldsymbol{E'F}}{N^{1/2}T^{1/2}} \left(\frac{\boldsymbol{F'F}}{T} \right)^{-1}$$

By Lemma C.3,

$$\|\boldsymbol{A}\|_F = O_p((TN)^{-1/2}) = O_p((mM)^{-1/2}); \quad \|\boldsymbol{B}\|_F = O_p(T^{-1})$$

Observe that $\tilde{\mathbf{\Omega}}_{\mathbf{\Phi}} - \boldsymbol{I}_K = \hat{\mathbf{\Omega}}_{\mathbf{\Phi}} - \boldsymbol{I}_K + \boldsymbol{A} + \boldsymbol{A'} + \boldsymbol{B}$, and that $\left\| \hat{\mathbf{\Omega}}_{\mathbf{\Phi}} - \boldsymbol{I}_K \right\|_F = O_p(N^{-\gamma})$ by (A.4). Thus, we have the second part of (ii) because

$$\left\| \tilde{\mathbf{\Omega}}_{\mathbf{\Phi}} - \boldsymbol{I}_K \right\|_F \leq \left\| \hat{\mathbf{\Omega}}_{\mathbf{\Phi}} - \boldsymbol{I}_K \right\|_F + 2\|\boldsymbol{A}\|_F + \|\boldsymbol{B}\|_F \leq O_p(m^{-\gamma})$$

Part (ii) holds by the second part of (i) and Lemma B.3.

Finally, let $\tilde{\mathbf{\Lambda}} = \boldsymbol{diag}(\tilde{\lambda}_1, ..., \tilde{\lambda}_K) = \mathbf{\Lambda}(\tilde{\mathbf{\Omega}}_{\mathbf{\Phi}}|1:K)$; and $\tilde{\mathbf{\Xi}} = \Xi(\tilde{\mathbf{\Omega}}_{\Phi})$. By (ii), $\tilde{\lambda}_h^{1/2} - 1 = (\tilde{\lambda}_k - 1)/(\tilde{\lambda}_k^{1/2} + 1) = O_p(m^{-\gamma})$, which implies $\left\| \tilde{\mathbf{\Lambda}}^{1/2} - \boldsymbol{I}_K \right\|_F = O_p(m^{-\gamma})$. Thus, (iii) holds because

$$\left\| \tilde{\mathbf{\Omega}}_{\mathbf{\Phi}}^{1/2} - \boldsymbol{I}_K \right\|_F = \left\| \tilde{\mathbf{\Xi}}(\tilde{\mathbf{\Lambda}}^{1/2} - \boldsymbol{I}_K)\tilde{\mathbf{\Xi}}' \right\|_F = \left\| \tilde{\mathbf{\Xi}} \right\|_F^2 \left\| \tilde{\mathbf{\Lambda}}^{1/2} - \boldsymbol{I}_K \right\|_F = O_p(m^{-\gamma})$$

This completes the proof. *Q.E.D.*

Some matrices are useful for the proofs of the following lemmas and theorem. Using the matrix $\tilde{\mathbf{\Phi}}$ defined in Lemma C.4, we can show that $\boldsymbol{X}$ and $\mathbf{S}_{NT}$ are of the following forms:

$$\frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} = \frac{\boldsymbol{F}}{T^{1/2}} \frac{\tilde{\mathbf{\Phi}}'}{N^{1/2}} + \frac{\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}}{N^{1/2}T^{1/2}}; \quad \mathbf{S}_{NT} = \frac{\boldsymbol{X'X}}{NT} = \boldsymbol{Z}_{NT} + \frac{\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}}{NT}$$

where $\boldsymbol{Z}_{NT} = (N^{-1/2}\tilde{\mathbf{\Phi}})\hat{\mathbf{\Omega}}_{\boldsymbol{F}}(N^{-1/2}\tilde{\mathbf{\Phi}}')$. We define the following matrices:

$$\boldsymbol{M}_{NT} = \tilde{\mathbf{\Omega}}_{\mathbf{\Phi}}^{1/2} \hat{\mathbf{\Omega}}_{\boldsymbol{F}} \tilde{\mathbf{\Omega}}_{\mathbf{\Phi}}^{1/2}; \quad \Xi_H^{\boldsymbol{Z}_{NT}} = N^{-1/2} \tilde{\mathbf{\Phi}} \tilde{\mathbf{\Omega}}_{\Phi}^{-1/2} \Xi(\boldsymbol{M}_{NT}|1:K)$$

where $\tilde{\mathbf{\Omega}}_{\mathbf{\Phi}} = N^{-1}\tilde{\mathbf{\Phi}}'\tilde{\mathbf{\Phi}}$.

Lemma C.5: $\mathbf{\Lambda}(\boldsymbol{M}_{NT}|1:K) = \mathbf{\Lambda}(\boldsymbol{Z}_{NT}|1:K)$ and $\Xi_H^{\boldsymbol{Z}_{NT}} = \Xi(\boldsymbol{Z}_{NT}|1:K)$.

Proof: We can easily show

$$\begin{aligned}\boldsymbol{Z}_{NT}\Xi_{H}^{\boldsymbol{Z}_{NT}} &= \left[N^{-1/2}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}N^{-1/2}\tilde{\boldsymbol{\Phi}}'\right]N^{-1/2}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2}\Xi(\boldsymbol{M}_{NT}|1:K)\\ &= N^{-1/2}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2}\boldsymbol{M}_{NT}\Xi(\boldsymbol{M}_{NT}|1:K)\\ &= N^{-1/2}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2}\Xi(\boldsymbol{M}_{NT}|1:K)\boldsymbol{\Lambda}(\boldsymbol{M}_{NT}|1:K)\\ &= \Xi_{H}^{\boldsymbol{Z}_{NT}}\boldsymbol{\Lambda}(\boldsymbol{M}_{NT}|1:K)\end{aligned}$$

This completes the proof. *Q.E.D.*

Lemma C.6: Under (A.1) – (A.6),

$$\|\boldsymbol{M}_{NT}-\boldsymbol{\Omega}_{\boldsymbol{F}}\|_F = O_p(m^{-\gamma});\quad \lambda_h\left(\boldsymbol{Z}_{NT}\right) = \lambda_h(\boldsymbol{M}_{NT}) = \sigma_j^2 + O_p(m^{-\gamma})$$

for $h = ks(j-1)+1,\ \ldots\ , ks(j)$ and $j = 1, \ldots, J$.

Proof: Observe that by Lemma C.4 and (A.5),

$$\begin{aligned}\left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}-\boldsymbol{\Omega}_{\boldsymbol{F}}\right\|_F &\leq \left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}-\boldsymbol{I}_K\right\|_F\|\boldsymbol{\Omega}_{\boldsymbol{F}}\|_F + \left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}-\boldsymbol{\Omega}_{\boldsymbol{F}}\right\|_F\\ &\qquad + \left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}-\boldsymbol{I}_K\right\|_F\left\|\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}-\boldsymbol{\Omega}_{\boldsymbol{F}}\right\|_F = O_p(m^{-\gamma})\end{aligned}$$

With this, we can obtain the first result:

$$\begin{aligned}\|\boldsymbol{M}_{NT}-\boldsymbol{\Omega}_{\boldsymbol{F}}\|_F &= \left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}-\boldsymbol{\Omega}_{\boldsymbol{F}}\right\|_F\\ &= \left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}-\boldsymbol{\Omega}_{\boldsymbol{F}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}+\boldsymbol{\Omega}_{\boldsymbol{F}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}-\boldsymbol{\Omega}_{\boldsymbol{F}}\right\|_F\\ &\leq \left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}}-\boldsymbol{\Omega}_{\boldsymbol{F}}\right\|_F\left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}\right\|_F+\|\boldsymbol{\Omega}_{\boldsymbol{F}}\|_F\left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{1/2}-\boldsymbol{I}_K\right\|_F = O_p(m^{-\gamma})\end{aligned}$$

Finally, because $\lambda_h\left(\boldsymbol{Z}_{NT}\right) = \lambda_h(\boldsymbol{M}_{NT})$ for all $h = 1, \ldots, K$, we can obtain the second result by the first result and Lemma B.3. *Q.E.D.*

Lemma C.7: Define $\Xi_{(j)}^{\boldsymbol{M}_{NT}} = \Xi(\boldsymbol{M}_{NT}|ks(j-1)+1:ks(j))$. Then, for each $j = 1, \ldots, D$, there exists a $k(j)\times k(j)$ orthonormal matrix $\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}}$ such that

$$\left\|\Xi_{(j)}^{\boldsymbol{M}_{NT}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}}-\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega}_{\boldsymbol{F}}}\right\|_F = O_p(m^{-\gamma})$$

Proof: Because $\|\boldsymbol{M}_{NT}-\boldsymbol{\Omega}_{\boldsymbol{F}}\|_2 \leq \|\boldsymbol{M}_{NT}-\boldsymbol{\Omega}_{\boldsymbol{F}}\|_F = O_p(m^{-\gamma})$, we can obtain the result by Lemma B.4 and Lemma C.6. *Q.E.D.*

Lemma C.8: Under (A.1) – (A.6),

$$\left\|M^{-1}\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}\right\|_2 = O_p(1);\quad \left\|(NT)^{-1}\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}\right\|_F = O_p(m^{-1/2})$$

Proof: Because $\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}$ and $= \boldsymbol{E'}\mathcal{P}(\boldsymbol{F})\boldsymbol{E}$ are positive semi-definite matrices,

$$\lambda_1\left(M^{-1}\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}\right) \leq \lambda_1\left(M^{-1}\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E} + M^{-1}\boldsymbol{E'}\mathcal{P}(\boldsymbol{F})\boldsymbol{E}\right) = \lambda_1\left(M^{-1}\boldsymbol{E'}\boldsymbol{E}\right) = O_p(1)$$

where the first inequality results from Lemma A.6 of Ahn and Horenstein (2013) and the last equality is due to (A.5). Thus, the first part holds. The second result holds because

$$\begin{aligned}\left\|(NT)^{-1}\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}\right\|_F &= m^{-1}\left\|M^{-1}\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}\right\|_F \\ &= m^{-1}\left[\Sigma_{h=1}^{m}\left(\lambda_h\left(M^{-1}\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{E}\right)\right)^2\right]^{1/2} \\ &\leq m^{-1}\left[m\left(\lambda_1\left(M^{-1}\boldsymbol{E'}\boldsymbol{E}\right)\right)^2\right]^{1/2} = O_p(m^{-1/2})\end{aligned}$$

This completes the proof. *Q.E.D.*

Lemma C.9: Under (A.1) – (A.6),

(i) $\|\mathbf{S}_{NT} - \boldsymbol{Z}_{NT}\|_2 = O_p(m^{-1})$; $\|\mathbf{S}_{NT} - \boldsymbol{Z}_{NT}\|_F = O_p(m^{-1/2})$;

(ii) $\lambda_h(\mathbf{S}_{NT}) = \sigma_j^2 + O_p(m^{-\gamma})$ for $h = ks(j-1)+1, ..., ks(j)$ and $j = 1, \ \ldots\ , J$;

(iii) $\lambda_q(\mathbf{S}_{NT}) = O_p(m^{-1})$, for $q = K+1, ..., m$.

Proof: The results in (i) immediately follow from Lemma C.8. By Lemma B.3 and (i), $\lambda_q(\mathbf{S}_{NT}) = \lambda_q(\boldsymbol{Z}_{NT}) + O_p(m^{-1})$ for all $q = 1, \ 2, \ \ldots\ , \ K$. In addition, by Lemma C.6, $\lambda_q(\boldsymbol{Z}_{NT}) = \lambda_q(\boldsymbol{M}_{NT}) = \lambda_q(\boldsymbol{\Omega_F}) + O_p(m^{-\gamma})$. Thus, (ii) holds because

$$\lambda_h(\mathbf{S}_{NT}) = \lambda_h(\boldsymbol{\Omega}_F) + O_p(m^{-\gamma}) + O_p(m^{-1}) = \sigma_j^2 + O_p(m^{-\gamma})$$

For $q \geq K+1$, (iii) holds because $\lambda_h(\boldsymbol{Z}_{NT}) = 0$. *Q.E.D.*

Lemma C.10: Under (A.1) – (A.6), for each $j = 1, ..., J$, there exists a $k(j) \times k(j)$ orthonormal matrix $\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}}$ such that $\left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}} - \boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}}\right\|_F = O_p(m^{-1})$ and $\left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime}\right\|_F = O_p(m^{-1})$.

Proof: The desired result is obtained by Lemma C.9 and Lemma B.4. *Q.E.D.*

Lemma C.11: Let $\boldsymbol{O}_{jj}^{*} = \boldsymbol{O}_{jj}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime}$, where $j = 1, ..., J$, and $\boldsymbol{O}_{jj}^{\boldsymbol{\Omega_F}}$, $\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}}$, and $\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}}$ are defined in Lemmas C.1, C.7, and C.10, respectively. Under (A.1) – (A.6), for $j = 1, ..., J$,

$$\left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - N^{-1/2}\boldsymbol{\Phi}_{(j)}\boldsymbol{O}_{jj}^{*}\right\|_F = O_p(m^{-\gamma})$$

Proof: Observe that

$$
\begin{aligned}
\boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}} &= N^{-1/2}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2}\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} \\
&= \left(N^{-1/2}\tilde{\boldsymbol{\Phi}} + N^{-1/2}\tilde{\boldsymbol{\Phi}}\left(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2} - \boldsymbol{I}_K\right)\right)\left(\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime} + \left(\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} - \boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right)\right) \\
&= N^{-1/2}\tilde{\boldsymbol{\Phi}}\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime} + N^{-1/2}\tilde{\boldsymbol{\Phi}}\left(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2} - \boldsymbol{I}_K\right)\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime} \\
&\quad + N^{-1/2}\tilde{\boldsymbol{\Phi}}\left(\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} - \boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right) \\
&\quad\; + N^{-1/2}\tilde{\boldsymbol{\Phi}}\left(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2} - \boldsymbol{I}_K\right)\left(\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} - \boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right) \\
&= N^{-1/2}\boldsymbol{\Phi}\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime} + N^{-1/2}\left(\tilde{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\right)\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime} \\
&\quad + N^{-1/2}\tilde{\boldsymbol{\Phi}}\left(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2} - \boldsymbol{I}_K\right)\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime} + N^{-1/2}\tilde{\boldsymbol{\Phi}}\left(\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} - \boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right) \\
&\quad\; + N^{-1/2}\tilde{\boldsymbol{\Phi}}\left(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2} - \boldsymbol{I}_K\right)\left(\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} - \boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right)
\end{aligned}
$$

Thus, by Lemma C.4 and C.7, we can have

$$
\begin{aligned}
&\left\|\boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}} - N^{-1/2}\boldsymbol{\Phi}\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right\|_F \\
&\quad \leq \left\|N^{-1/2}\left(\tilde{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\right)\right\|\left\|\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right\|_F + \left\|N^{-1/2}\tilde{\boldsymbol{\Phi}}\right\|_F\left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2} - \boldsymbol{I}_K\right\|_F\left\|\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right\|_F \\
&\qquad + \left\|N^{-1/2}\tilde{\boldsymbol{\Phi}}\right\|_F\left\|\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} - \boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right\|_F \\
&\qquad + \left\|N^{-1/2}\tilde{\boldsymbol{\Phi}}\right\|_F\left\|\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2} - \boldsymbol{I}_K\right\|_F\left\|\boldsymbol{\Xi}_{(j)}^{\boldsymbol{M}_{NT}} - \boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right\|_F \\
&\quad = O_p(m^{-\gamma})
\end{aligned}
$$

With this result and Lemma C.11, we can obtain the desired result because:

$$
\begin{aligned}
&\left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - N^{-1/2}\boldsymbol{\Phi}\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime}\right\|_F \\
&\quad = \left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime} + \boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime} - N^{-1/2}\boldsymbol{\Phi}\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime}\right\|_F \\
&\quad \leq \left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}}\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime}\right\|_F + \left\|\boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}} - N^{-1/2}\boldsymbol{\Phi}\boldsymbol{O}_{(j)}^{\boldsymbol{\Omega_F}}\boldsymbol{O}_{jj}^{\boldsymbol{M}_{NT}\prime}\right\|_F\left\|\boldsymbol{O}_{jj}^{\mathbf{S}_{NT}\prime}\right\|_F \\
&\quad = O_p(m^{-\gamma})
\end{aligned}
$$

This completes the proof. *Q.E.D.*

Lemma C.12: Under (A.1) – (A.6),

$$
\left\|(NT)^{-1/2}\boldsymbol{E}'\mathcal{Q}(\boldsymbol{F})\boldsymbol{u}\right\|_2 = O_p(1); \quad \left\|N^{-1}T^{-1/2}\tilde{\boldsymbol{\Phi}}'\boldsymbol{E}'\mathcal{Q}(\boldsymbol{F})\boldsymbol{u}\right\|_2 = O_p(m^{-1/2})
$$

Proof: By (A.7) and Lemma C.4,

$$\left\|\frac{\boldsymbol{E'}Q(\boldsymbol{F})\boldsymbol{u}}{N^{1/2}T^{1/2}}\right\|_F \leq \left\|\frac{\boldsymbol{E'u}}{N^{1/2}T^{1/2}}\right\|_F + \frac{1}{T^{1/2}}\left\|\frac{\boldsymbol{E'F}}{N^{1/2}T^{1/2}}\right\|_F \left\|\left(\frac{\boldsymbol{F'F}}{T}\right)^{-1}\right\|_F \left\|\frac{\boldsymbol{F'u}}{T^{1/2}}\right\|_F$$
$$= O_p(1) + O_p(T^{-1/2}) = O_p(1)$$

In addition,

$$\left\|\frac{\tilde{\boldsymbol{\Phi}}'\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{u}}{NT^{1/2}}\right\|_F$$
$$= \left\|\frac{\boldsymbol{\Phi}'\boldsymbol{E'u}}{NT^{1/2}}\right\|_F + \frac{1}{N^{1/2}T^{1/2}}\left\|\frac{\boldsymbol{\Phi}'\boldsymbol{E'F}}{N^{1/2}T^{1/2}}\right\|_F \left\|\left(\frac{\boldsymbol{F'F}}{T}\right)^{-1}\right\|_F \left\|\frac{\boldsymbol{F'u}}{T^{1/2}}\right\|_F$$
$$+ \frac{1}{T^{1/2}}\left\|\left(\frac{\boldsymbol{F'F}}{T}\right)^{-1}\right\|_F \left\|\frac{\boldsymbol{F'EE'u}}{NT^{3/2}}\right\|_F + \frac{1}{T}\left\|\left(\frac{\boldsymbol{F'F}}{T}\right)^{-1}\frac{\boldsymbol{F'EE'F}}{NT}\right\|_F \left\|\left(\frac{\boldsymbol{F'F}}{T}\right)^{-1}\frac{\boldsymbol{F'u}}{T^{1/2}}\right\|_F$$
$$= O_p(N^{-1/2}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1/2}) + O_p(T^{-1}) = O_p(m^{-1/2})$$

This completes the proof because Because at because the Frobenius norms of $\boldsymbol{E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{u}$ and $\tilde{\boldsymbol{\Phi}}\boldsymbol{'E'}\mathcal{Q}(\boldsymbol{F})\boldsymbol{u}$ respectively equal to their spectral norms. *Q.E.D.*

Lemma C.13: Under (A.1) – (A.7),

(i) $\left\|T^{-1}\boldsymbol{F'}_{(j)}\boldsymbol{y} - \sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_2 = O_p(T^{-\gamma})$, for $j = 1, ..., R$;

(ii) $\left\|T^{-1}\boldsymbol{F'}_{(j)}\boldsymbol{y}\right\|_2 = O_p(T^{-\gamma})$, for $j = R+1, ..., J$;

(iii) $\left\|N^{-1/2}T^{-1/2}\boldsymbol{E'y}\right\|_2 = O_p(1)$.

Proof: Part (i) holds because, for $j \leq R$,

$$\left\|\frac{\boldsymbol{F'}_{(j)}\boldsymbol{y}}{T} - \sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F = \left\|\Sigma_{j'=1}^R\frac{\boldsymbol{F'}_{(j)}\boldsymbol{F}_{(j')}}{T}\boldsymbol{\beta}_{(j')} + \frac{\boldsymbol{F'}_{(j)}\boldsymbol{u}}{T} - \sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F$$
$$\leq \left\|\frac{\boldsymbol{F'}_{(j)}\boldsymbol{F}_{(j')}}{T} - \sigma_j^2\boldsymbol{I}_{k(j)}\right\|_F \|\boldsymbol{\beta}_{(j')}\|_F$$
$$+ \Sigma_{j'=1,j'\neq j}^R\left\|\frac{\boldsymbol{F'}_{(j)}\boldsymbol{F}_{(j')}}{T}\boldsymbol{\beta}_{(j')}\right\|_F + \left\|\frac{\boldsymbol{F'}_{(j)}\boldsymbol{u}}{T}\right\|_F$$
$$= O_p(T^{-\gamma}) + O_p(T^{-\gamma}) + O_p(T^{-1/2}) = O_p(T^{-\gamma})$$

Similarly, (ii) holds because, for $j \geq R+1$,

$$\left\|\frac{\boldsymbol{F}'_{(j)}\boldsymbol{y}}{T}\right\|_F \leq \Sigma_{j'=1}^{R}\left\|\frac{\boldsymbol{F}'_{(j)}\boldsymbol{F}_{(j')}}{T}\right\|_F \left\|\boldsymbol{\beta}_{(j')}\right\|_F + \left\|\frac{\boldsymbol{F}'_{(j)}\boldsymbol{u}}{T}\right\|_F$$
$$= O_p(T^{-\gamma}) + O_p(T^{-1/2}) = O_p(T^{-\gamma})$$

Finally, (iii) holds because, by (A.7) and Lemma C.3,

$$\left\|\frac{\boldsymbol{E}'\boldsymbol{y}}{(NT)^{1/2}}\right\|_F = \Sigma_{j=1}^{R}\left\|\frac{\boldsymbol{E}'\boldsymbol{F}_{(j)}}{(NT)^{1/2}}\right\|_F \left\|\boldsymbol{\beta}_{(j)}\right\|_F + \left\|\frac{\boldsymbol{E}'\boldsymbol{u}}{(NT)^{1/2}}\right\|_F = O_p(1) \qquad \textit{Q.E.D.}$$

Lemma C.14: Under (A.1) – (A.7), $\left\|\boldsymbol{b}_{NT} - \Sigma_{j=1}^{R}\sigma_j^2 N^{-1/2}\boldsymbol{\Phi}_{(j)}\boldsymbol{\beta}_{(j)}\right\|_2 = O_p(T^{-\gamma})$.

Proof: Observe that

$$\begin{aligned}
&\boldsymbol{b}_{NT} - \Sigma_{j=1}^{R}\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\sigma_j^2\boldsymbol{\beta}_{(j)} \\
&\quad = \Sigma_{j=1}^{J}\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\frac{\boldsymbol{F}'_{(j)}\boldsymbol{y}}{T} + \frac{1}{T^{1/2}}\frac{\boldsymbol{E}'\boldsymbol{y}}{N^{1/2}T^{1/2}} - \Sigma_{j=1}^{R}\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\sigma_j^2\boldsymbol{\beta}_{(j)} \\
&\quad = \Sigma_{j=1}^{R}\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\left(\frac{\boldsymbol{F}'_{(j)}\boldsymbol{y}}{T} - \sigma_j^2\boldsymbol{\beta}_{(j)}\right) + \Sigma_{j=R+1}^{J}\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\frac{\boldsymbol{F}'_{(j)}\boldsymbol{y}}{T} + \frac{1}{T^{1/2}}\frac{\boldsymbol{E}'\boldsymbol{y}}{N^{1/2}T^{1/2}}
\end{aligned}$$

Thus, by Lemma C.13,

$$\begin{aligned}
&\left\|\boldsymbol{b}_{NT} - \Sigma_{j=1}^{R}\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F \\
&\qquad \leq \Sigma_{j=1}^{R}\left\|\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\right\|_F\left\|\frac{\boldsymbol{F}'_{(j)}\boldsymbol{y}}{T} - \sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F + \Sigma_{j=R+1}^{J}\left\|\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\right\|_F\left\|\frac{\boldsymbol{F}'_{(j)}\boldsymbol{y}}{T}\right\|_F \\
&\qquad\qquad + \frac{1}{T^{1/2}}\left\|\frac{\boldsymbol{E}'\boldsymbol{y}}{N^{1/2}T^{1/2}}\right\|_F \\
&\qquad = O_p(T^{-\gamma})
\end{aligned}$$

which completes the proof. *Q.E.D.*

Lemma C.15: Under (A.1) – (A.7), for $j = 1, \ldots, R$,

(i) $\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{O}^{*}_{jj}{}'\sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_2 = O_p(m^{-\gamma})$,

where $\boldsymbol{O}^{*}_{jj}$ is defined in Lemma C.11. For $j \geq R+1$,

(ii) $\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right\|_2 = O_p(m^{-\gamma})$.

Proof: For $j \leq R$,

$$\begin{aligned}
\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} &= \left(\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\boldsymbol{O}^*_{jj} + \left(\boldsymbol{\Xi}^{S_{NT}}_{(j)} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\boldsymbol{O}^*_{jj}\right)\right)' \left(\Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')} + \left(\boldsymbol{b}_{NT} - \Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')}\right)\right) \\
&= \boldsymbol{O}^*_{jj}{}'\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\left(\Sigma^R_{j'=1}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma^2_{j'}\boldsymbol{\beta}_{(j')}\right) + \boldsymbol{O}^*_{jj}{}'\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\left(\boldsymbol{b}_{NT} - \Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')}\right) \\
&\quad + \left(\boldsymbol{\Xi}^{S_{NT}}_{(j)} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\boldsymbol{O}^*_{jj}\right)'\Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')} + \left(\boldsymbol{\Xi}^{S_{NT}}_{(j)} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\boldsymbol{O}^*_{jj}\right)'\left(\boldsymbol{b}_{NT} - \Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')}\right)
\end{aligned}$$

Applying Lemmas C.11 and C.14 to this result, we can show

$$\begin{aligned}
&\left\|\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{O}^*_{jj}{}'\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\left(\Sigma^R_{j'=1}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma^2_{j'}\boldsymbol{\beta}_{(j')}\right)\right\|_F \\
&\leq \left\|\boldsymbol{O}^*_{jj}{}'\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\right\|_F\left\|\boldsymbol{b}_{NT} - \Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')}\right\|_F + \left\|\boldsymbol{\Xi}^{S_{NT}}_{(j)} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\boldsymbol{O}^*_{jj}\right\|_F\left\|\Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')}\right\|_F \\
&\quad + \left\|\boldsymbol{\Xi}^{S_{NT}}_{(j)} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\boldsymbol{O}^*_{jj}\right\|_F\left\|\boldsymbol{b}_{NT} - \Sigma^R_{j'=1}\sigma^2_{j'}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\boldsymbol{\beta}_{(j')}\right\|_F \\
&= O_p(T^{-\gamma}) + O_p(m^{-\gamma}) + O_p(T^{-\gamma}m^{-\gamma})
\end{aligned}$$

This implies

$$\text{(C.1)}\quad \left\|\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{O}^*_{jj}{}'\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\left(\Sigma^R_{j'=1}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma^2_{j'}\boldsymbol{\beta}_{(j')}\right)\right\|_F = O_p(m^{-\gamma})$$

In addition, the following holds:

$$\begin{aligned}
\text{(C.2)}\quad &\left\|\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\left(\Sigma^R_{j'=1}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma^2_{j'}\boldsymbol{\beta}_{(j')}\right) - \frac{\boldsymbol{\Phi}'_{(j)}\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F \leq \Sigma^R_{\substack{j'=1,\\ j'\neq j}}\left\|\frac{\boldsymbol{\Phi}'_{(j)}\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\right\|_F\left\|\sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F \\
&= O_p(m^{-\gamma});
\end{aligned}$$

$$\text{(C.3)}\quad \left\|\frac{\boldsymbol{\Phi}'_{(j)}\boldsymbol{\Phi}_{(j)}}{N}\sigma^2_j\boldsymbol{\beta}_{(j)} - \sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F \leq \left\|\frac{\boldsymbol{\Phi}'_{(j)}\boldsymbol{\Phi}_{(j)}}{N} - \boldsymbol{I}_{k(j)}\right\|_F\left\|\sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F = O_p(m^{-\gamma})$$

Thus, by (C.1) – (C.3), we can obtain (i) because

$$\begin{aligned}
\left\|\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{O}^*_{jj}\sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F &= \left\|\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{O}^*_{jj}\sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F \\
&\leq \left\|\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{O}^*_{jj}{}'\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\left(\Sigma^R_{j'=1}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma^2_{j'}\boldsymbol{\beta}_{(j')}\right)\right\|_F \\
&\quad + \left\|\boldsymbol{O}^*_{jj}\right\|_F\left\|\frac{\boldsymbol{\Phi}'_{(j)}}{N^{1/2}}\left(\Sigma^R_{j'=1}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma^2_{j'}\boldsymbol{\beta}_{(j')}\right) - \frac{\boldsymbol{\Phi}'_{(j)}\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F \\
&\quad + \left\|\boldsymbol{O}^*_{jj}\right\|_F\left\|\frac{\boldsymbol{\Phi}'_{(j)}\boldsymbol{\Phi}_{(j)}}{N}\sigma^2_j\boldsymbol{\beta}_{(j)} - \sigma^2_j\boldsymbol{\beta}_{(j)}\right\|_F
\end{aligned}$$

Part can be shown for $j \geq R+1$, similarly. *Q.E.D.*

Lemma C.16: Under (A.1) – (A.6), for $j = 1, ..., J$,

$$\left\| (NT)^{-1/2} \boldsymbol{X} \boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - T^{-1/2} \boldsymbol{F}_{(j)} \boldsymbol{O}_{jj}^{*} \right\|_F = O_p(m^{-\gamma})$$

where $\boldsymbol{O}_{jj}^{*}$ is defined in Lemma C.11.

Proof: By Lemma C.11, we can show

$$\text{(C.4)} \quad \left\| \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right\|_F \leq \left\| \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \right\|_F O_p(m^{-\gamma}) = O_p(m^{-\gamma}) \quad .$$

In addition, we can have

$$\text{(C.5)} \quad \left\| \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \right\|_F = O_p(m^{-\gamma})$$

because

$$\begin{aligned}
&\left\| \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \right\|_F \\
&= \left\| \frac{\boldsymbol{F}}{T^{1/2}} \frac{\boldsymbol{\Phi'}\boldsymbol{\Phi}_{(j)}}{N} + \frac{1}{N^{1/2}} \frac{\boldsymbol{E}\boldsymbol{\Phi}_{(j)}}{N^{1/2}T^{1/2}} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \right\|_F \\
&= \left\| -\frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \left(\boldsymbol{I}_{k(j)} - \frac{\boldsymbol{\Phi'}_{(j)}\boldsymbol{\Phi}_{(j)}}{N} \right) + \Sigma_{j'=1, j' \neq j}^{J} \frac{\boldsymbol{F}_{(j')}}{T^{1/2}} \frac{\boldsymbol{\Phi'}_{(j')}\boldsymbol{\Phi}_{(j)}}{N} + \frac{1}{N^{1/2}} \frac{\boldsymbol{E}\boldsymbol{\Phi}_{H1}}{N^{1/2}T^{1/2}} \right\|_F \\
&\leq \left\| \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \right\|_F \left\| \boldsymbol{I}_{k(j)} - \frac{\boldsymbol{\Phi'}_{(j)}\boldsymbol{\Phi}_{(j)}}{N} \right\|_F + \Sigma_{j'=1, j' \neq j}^{J} \left\| \frac{\boldsymbol{F}_{(j')}}{T^{1/2}} \right\|_F \left\| \frac{\boldsymbol{\Phi'}_{(j')}\boldsymbol{\Phi}_{(j)}}{N} \right\|_F + \frac{1}{N^{1/2}} \left\| \frac{\boldsymbol{E}\boldsymbol{\Phi}_{(j)}}{N^{1/2}T^{1/2}} \right\|_F \\
&= O_p(m^{-\gamma}) + O_p(m^{-\gamma}) + O_p(N^{-1/2})
\end{aligned}$$

Finally, we have

$$\begin{aligned}
&\left\| \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \boldsymbol{O}_{jj}^{*} \right\|_F \\
&\leq \left\| \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right\|_F + \left\| \frac{\boldsymbol{X}}{N^{1/2}T^{1/2}} \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \right\|_F \left\| \boldsymbol{O}_{jj}^{*} \right\|_F
\end{aligned}$$

which, with (C.4) and (C.5) imply the desired result. *Q.E.D.*

Lemma C.17: Under (A.1) – (A.8), the following holds. Let $\boldsymbol{\Xi}_{H}^{\mathbf{S}_{NT}} = (\boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}, ..., \boldsymbol{\Xi}_{(J)}^{\mathbf{S}_{NT}})$ $= \boldsymbol{\Xi}(\mathbf{S}_{NT}|1:K)$. Then,

(i) $\left\| \mathcal{Q}(\boldsymbol{\Xi}_{H}^{\mathbf{S}_{NT}}) - \mathcal{Q}(N^{-1/2}\tilde{\boldsymbol{\Phi}}) \right\|_F = O_p(m^{-1})$.

Let $\mathbf{S}_{NT}^{*} = \boldsymbol{X}\boldsymbol{X'}/(NT)$; $\boldsymbol{\Xi}_{H}^{*} = \boldsymbol{\Xi}(\mathbf{S}_{NT}^{*}|1:K)$; and $\tilde{\boldsymbol{F}} = \boldsymbol{F} + \boldsymbol{E}\boldsymbol{\Phi}(\boldsymbol{\Phi'}\boldsymbol{\Phi})^{-1}$. Then,

(ii) $\left\| \mathcal{Q}(\Xi_H^*) - \mathcal{Q}(T^{-1/2}\tilde{\boldsymbol{F}}) \right\|_F = O_p(m^{-1})$.

Proof: Let $\Xi_H^{\boldsymbol{Z}_{NT}} = \Xi(\boldsymbol{Z}_{NT}|1:K)$. Observe that $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2}\Xi(\boldsymbol{M}_{NT}|1:K)$ is an invertible matrix, and that $\Xi_H^{\boldsymbol{Z}_{NT}} = N^{-1/2}\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}^{-1/2}\Xi(\boldsymbol{M}_{NT}|1:K)$. Thus,

$$\text{(C.6)} \quad \mathcal{Q}(\Xi_H^{\boldsymbol{Z}_{NT}}) = \mathcal{Q}(N^{-1/2}\tilde{\boldsymbol{\Phi}}).$$

By Lemmas A.1 and C.9 and the fact that $\lambda_K(\boldsymbol{Z}_{NT}) > 0$ and $\lambda_{K+1}(\boldsymbol{Z}_{NT}) = 0$, there exists an orthonormal matrix $\boldsymbol{O}^{\mathbf{S}_{NT}}$ such that

$$\left\| \Xi_H^{\mathbf{S}_{NT}} \boldsymbol{O}^{\mathbf{S}_{NT}} - \Xi_H^{\boldsymbol{Z}_{NT}} \right\|_F \le \frac{2^{3/2}K^{1/2}\|\mathbf{S}_{NT} - \boldsymbol{Z}_{NT}\|_2}{\lambda_K(\boldsymbol{Z}_{NT})} = O_p(m^{-1})$$

where the last equality is due to Lemma C.9. Thus, by Lemma B.5, we have

$$\text{(C.7)} \quad \left\| \mathcal{Q}(\Xi_H^{\mathbf{S}_{NT}}) - \mathcal{Q}(\Xi_H^{\boldsymbol{Z}_{NT}}) \right\|_F = \left\| \mathcal{Q}(\Xi_H^{\mathbf{S}_{NT}}\boldsymbol{O}^{\mathbf{S}_{NT}}) - \mathcal{Q}(\Xi_H^{\boldsymbol{Z}_{NT}}) \right\|_F = O_p(m^{-1})$$

which, with (C.6) implies the result in (i). We can show (ii) similarly. It is straightforward to show that

$$\mathbf{S}_{NT}^* = \boldsymbol{Z}_{NT}^* + (NT)^{-1}\boldsymbol{E}\mathcal{Q}(N^{-1/2}\boldsymbol{\Phi})\boldsymbol{E'}$$

where $\boldsymbol{Z}_{NT}^* = (T^{-1/2}\tilde{\boldsymbol{F}})(N^{-1}\boldsymbol{\Phi'\Phi})^{-1}(T^{-1/2}\tilde{\boldsymbol{F}})$. Thus, by the same methods used to show (C.6) and (C.7), we can show

$$\mathcal{Q}(\Xi(\boldsymbol{Z}_{NT}^*|1:K)) = \mathcal{Q}(T^{-1/2}\tilde{\boldsymbol{F}})$$
$$\|\mathcal{Q}(\Xi_H^*\boldsymbol{O}^{**}) - \mathcal{Q}(\Xi(\boldsymbol{Z}_{NT}^*|1:K))\|_2 = \left\| \mathcal{Q}(\Xi_H^{\mathbf{S}_{NT}}\boldsymbol{O}^{\mathbf{S}_{NT}}) - \mathcal{Q}(\Xi_H^{\boldsymbol{Z}_{NT}}) \right\|_2 = O_p(m^{-1})$$

for some orthonormal matrix $\boldsymbol{O}^{**}$. These results imply (ii). *Q.E.D.*

Lemma C.18: Let $\boldsymbol{H}_{NT} = (NT)^{-1/2}\Xi_L^{\mathbf{S}_{NT}\prime}\mathcal{Q}(\boldsymbol{\Phi})\boldsymbol{E'}\mathcal{Q}(\tilde{\boldsymbol{F}})$, where $\tilde{\boldsymbol{F}}$ is defined in Lemma C.17. Assume that (A.1) – (A.7) hold. Then, the following holds.

(i) $\left\| \boldsymbol{c}_L^{\mathbf{S}_{NT}} - T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u} \right\|_2 = O_p(m^{-3/2})$;

(ii) $\left\| T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u} \right\|_2 = O_p(T^{-1/2})$.

Let $\boldsymbol{r}_{NT}$ be an $m \times 1$ random vector with $\|\boldsymbol{r}_{NT}\|_2 = O_p(1)$ which is independent of $\boldsymbol{u}$. Then,

(iii) $\left\| T^{-1/2}\boldsymbol{r'}_{NT}\boldsymbol{H}_{NT}\boldsymbol{u} \right\|_2 = O_p((Tm)^{-1/2})$.

Proof: Let $\boldsymbol{\Xi}_L^* = \boldsymbol{\Xi}(\mathbf{S}_{NT}^*|K+1:m)$, where $\mathbf{S}_{NT}^*$ is defined in Lemma C.17. Using the fact that $\boldsymbol{\Xi}_L^* \boldsymbol{\Lambda}_L^{\mathbf{S}_{NT}} = (NT)^{-1/2} \boldsymbol{X} \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}}$ and $\boldsymbol{X} = \tilde{\boldsymbol{F}} \boldsymbol{\Phi}' + \boldsymbol{E} \mathcal{Q}(\boldsymbol{\Phi})$, we can easily show that

$$
\begin{aligned}
&\boldsymbol{c}_L^{\mathbf{S}_{NT}} - T^{-1/2} \boldsymbol{H}_{NT} \boldsymbol{u} \\
&= \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \frac{\boldsymbol{X}'}{(NT)^{1/2}} \mathcal{Q}(\boldsymbol{\Xi}_H^*) \frac{\boldsymbol{y}}{T^{1/2}} + \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \frac{\boldsymbol{X}'}{(NT)^{1/2}} \mathcal{P}(\boldsymbol{\Xi}_H^*) \frac{\boldsymbol{y}}{T^{1/2}} - T^{-1/2} \boldsymbol{H}_{NT} \boldsymbol{u} \\
&= \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \frac{\boldsymbol{X}'}{(NT)^{1/2}} \mathcal{Q}(\boldsymbol{\Xi}_H^*) \frac{\boldsymbol{y}}{T^{1/2}} - T^{-1/2} \boldsymbol{H}_{NT} \boldsymbol{u} \\
&= \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \frac{\boldsymbol{X}'}{(NT)^{1/2}} \mathcal{Q}(\tilde{\boldsymbol{F}}) \frac{\boldsymbol{y}}{T^{1/2}} + \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}} \frac{\boldsymbol{X}'}{(NT)^{1/2}} \left(\mathcal{Q}(\boldsymbol{\Xi}_H^*) - \mathcal{Q}(\tilde{\boldsymbol{F}}) \right) \frac{\boldsymbol{y}}{T^{1/2}} - T^{-1/2} \boldsymbol{H}_{NT} \boldsymbol{u} \\
&= \frac{1}{T^{1/2}} \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \mathcal{Q}(\boldsymbol{\Phi}) \frac{\boldsymbol{E}'}{(NT)^{1/2}} \mathcal{Q}(\tilde{\boldsymbol{F}}) \boldsymbol{u} - T^{-1/2} \boldsymbol{H}_{NT} \boldsymbol{u} \\
&\qquad - \frac{1}{N^{1/2}} \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \mathcal{Q}(\boldsymbol{\Phi}) \frac{\boldsymbol{E}'}{(NT)^{1/2}} \mathcal{Q}(\tilde{\boldsymbol{F}}) \frac{\boldsymbol{E}\boldsymbol{\Phi}}{(NT)^{1/2}} \left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N} \right)^{-1} \boldsymbol{\beta} \\
&\qquad + \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \frac{\boldsymbol{X}'}{(NT)^{1/2}} \left(\mathcal{Q}(\boldsymbol{\Xi}_H^*) - \mathcal{Q}(\tilde{\boldsymbol{F}}) \right) \frac{\boldsymbol{y}}{T^{1/2}} \\
&= -\frac{1}{N^{1/2}} \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \mathcal{Q}(\boldsymbol{\Phi}) \frac{\boldsymbol{E}'}{(NT)^{1/2}} \mathcal{Q}(\tilde{\boldsymbol{F}}) \frac{\boldsymbol{E}\boldsymbol{\Phi}}{(NT)^{1/2}} \left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N} \right)^{-1} \boldsymbol{\beta} \\
&\qquad + \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime} \frac{\boldsymbol{X}'}{(NT)^{1/2}} \left(\mathcal{Q}(\boldsymbol{\Xi}_H^*) - \mathcal{Q}(\tilde{\boldsymbol{F}}) \right) \frac{\boldsymbol{y}}{T^{1/2}} \\
&\equiv -\boldsymbol{I} + \boldsymbol{II}
\end{aligned}
$$

For (i), it is sufficient to show that $\|\boldsymbol{I}\|_2 = O_p(m^{-3/2})$, and $\|\boldsymbol{II}\|_2 = O_p(m^{-3/2})$. By (A.5) and the fact that $\boldsymbol{E}'\boldsymbol{E} - \boldsymbol{E}'\mathcal{Q}(\tilde{\boldsymbol{F}})\boldsymbol{E}$ is positive semi-definite and $\mathcal{Q}(\boldsymbol{\Phi})$ is idempotent, we have

$$
\begin{aligned}
\|\boldsymbol{I}\|_2 &\leq \frac{1}{N^{1/2}m} \left\| \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}} \right\|_2 \|\mathcal{Q}(\boldsymbol{\Phi})\|_2 \left\| \frac{\boldsymbol{E}'}{M^{1/2}} \mathcal{Q}(\tilde{\boldsymbol{F}}) \frac{\boldsymbol{E}}{M^{1/2}} \right\|_2 \left\| \frac{\boldsymbol{\Phi}}{N^{1/2}} \left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N} \right)^{-1} \boldsymbol{\beta} \right\|_2 \\
&\leq \frac{1}{N^{1/2}m} \times 1 \times 1 \times \left\| \frac{\boldsymbol{E}'\boldsymbol{E}}{M} \right\|_2 \times O_p(1) \leq O_p(m^{-3/2}
\end{aligned}
$$

By Lemma C.7 and the fact that $\left\| (NT)^{-1/2} \boldsymbol{X} \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}} \right\|_2 = [\lambda_{K+1}(\mathbf{S}_{NT})]^{1/2} = O_p(m^{-1/2})$, we also have

$$
\|\boldsymbol{II}\|_2 \leq O_p(m^{-1/2}) \times O_p(m^{-1}) \times \left\| T^{-1} \boldsymbol{y} \right\|_2 = O_p(m^{-3/2})
$$

For (ii), observe that

$$\begin{aligned}
\left\|\frac{\boldsymbol{E}'\tilde{\boldsymbol{F}}}{N^{1/2}T}\right\|_F &= \frac{1}{T^{1/2}}\left\|\frac{\boldsymbol{E}'\boldsymbol{F}}{N^{1/2}T^{1/2}}\right\|_F + \frac{1}{m^{1/2}}m^{1/2}\left\|\frac{\boldsymbol{E}'\boldsymbol{E}}{NT}\right\|_F\left\|\frac{\boldsymbol{\Phi}}{N^{1/2}}\left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}\right)^{-1}\right\|_F \\
&= O_p(m^{-1/2}); \\
\left\|\frac{\tilde{\boldsymbol{F}}'\boldsymbol{u}}{T^{1/2}}\right\|_F &= \left\|\frac{\boldsymbol{F}'\boldsymbol{u}}{T^{1/2}}\right\|_F + \left\|\left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}\right)^{-1}\frac{\boldsymbol{\Phi}'}{N^{1/2}}\right\|_F\left\|\frac{\boldsymbol{E}'\boldsymbol{u}}{N^{1/2}T^{1/2}}\right\|_F = O_p(1);
\end{aligned}$$

$$\begin{aligned}
\left\|\frac{\tilde{\boldsymbol{F}}'\tilde{\boldsymbol{F}}}{T} - \frac{\boldsymbol{F}'\boldsymbol{F}}{T}\right\|_F &= 2\frac{1}{(NT)^{1/2}}\left\|\frac{\boldsymbol{F}'\boldsymbol{E}\boldsymbol{\Phi}}{(TN)^{1/2}}\right\|_F\left\|\left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}\right)^{-1}\right\|_F \\
&\quad + \left\|\left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}\right)^{-1}\frac{\boldsymbol{\Phi}'}{N^{1/2}}\right\|_F^2\left\|\frac{\boldsymbol{E}'\boldsymbol{E}}{NT}\right\|_F \\
&= O_p(m^{-1/2})
\end{aligned}$$

With these results, we can show that

$$\begin{aligned}
\left\|\frac{\boldsymbol{E}'\mathcal{Q}(\tilde{\boldsymbol{F}})\boldsymbol{u}}{N^{1/2}T}\right\|_2 &= \left\|\frac{\boldsymbol{E}'\mathcal{Q}(\tilde{\boldsymbol{F}})\boldsymbol{u}}{N^{1/2}T}\right\|_F \\
&\leq \frac{1}{T^{1/2}}\left\|\frac{\boldsymbol{E}'\boldsymbol{u}}{(NT)^{1/2}}\right\|_F + \frac{1}{T^{1/2}}\left\|\frac{\boldsymbol{E}'\tilde{\boldsymbol{F}}}{N^{1/2}T}\right\|_F\left\|\left(\frac{\tilde{\boldsymbol{F}}'\tilde{\boldsymbol{F}}}{T}\right)^{-1}\right\|_F\left\|\frac{\tilde{\boldsymbol{F}}'\boldsymbol{u}}{T^{1/2}}\right\|_F \\
&= O_p(T^{-1/2})
\end{aligned}$$

By this result and the facts that $\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}\prime}\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}} = \boldsymbol{I}_{m-K}$ and $\mathcal{Q}(N^{-1/2}\boldsymbol{\Phi})$ is idempotent, we can show that (ii) holds because

$$\begin{aligned}
\left\|T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u}\right\|_2 &= \left\|\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}}\right\|_2\left\|\mathcal{Q}\left(\frac{\boldsymbol{\Phi}}{N^{1/2}}\right)\right\|_2\left\|\frac{\boldsymbol{E}'\mathcal{Q}(T^{-1/2}\tilde{\boldsymbol{F}})\boldsymbol{u}}{N^{1/2}T}\right\|_2 \\
&= 1 \times 1 \times O_p(T^{-1/2}) = O_p(T^{-1/2})
\end{aligned}$$

For (iii), observe that $\boldsymbol{H}_{NT}$ is a function of $\boldsymbol{E}$, $\boldsymbol{F}$ and $\boldsymbol{\Phi}$, all of which are independent of $\boldsymbol{u}$. That is, $\boldsymbol{H}'_{NT}\boldsymbol{r}_{NT}$ and $\boldsymbol{u}$ are independent. Observe also that

$$\begin{aligned}
\|\boldsymbol{H}_{NT}\|_2 &\leq m^{-1/2}\left\|\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}}\right\|_2\|\mathcal{Q}(\boldsymbol{\Phi})\|_2\left\|M^{-1/2}\boldsymbol{E}\right\|_2\left\|\mathcal{Q}(\tilde{\boldsymbol{F}})\right\| \\
&= m^{-1/2} \times 1 \times 1 \times \left(\lambda_1(M^{-1}\boldsymbol{E}\boldsymbol{E}')\right)^{1/2} \times 1 = O_p(m^{-1/2})
\end{aligned}$$

By these results, we have

$$\begin{aligned}
&\mathrm{E}\left(\left\|T^{-1/2}\boldsymbol{r}'_{NT}\boldsymbol{H}_{NT}\boldsymbol{u}\right\|_2^2|\boldsymbol{H}_{NT}, \boldsymbol{r}_{NT}\right) \\
&= T^{-1/2}\boldsymbol{r}'_{NT}\boldsymbol{H}_{NT}\mathrm{E}(\boldsymbol{u}\boldsymbol{u}')\boldsymbol{H}'_{NT}\boldsymbol{r}_{NT} \leq T^{-1}\|\boldsymbol{r}_{NT}\|_2^2\|\boldsymbol{H}_{NT}\|_2^2\lambda_1\left(\mathrm{E}(\boldsymbol{u}\boldsymbol{u}')\right) = O_p((mT)^{-1})
\end{aligned}$$

which implies (iii). *Q.E.D.*

Proof of Lemma 2.4.1: The results (i) and (ii) hold by Lemma C.9. The result (iii) holds by Lemma C.14. The result (iv) holds by Lemma C.11. Finally, the results (v) and (vi) hold by Lemmas C.15. Finally, the parts (vii) – (ix) hold by Lemma C.18. *Q.E.D.*

Lemma C.19: Under (A.1) – (A.8), for $q \geq 1$,

$$\begin{aligned}
&\|\boldsymbol{v}_{H1}(q)\|_2 = O_p(m^{-\gamma});\\
&\|\boldsymbol{v}_{H2}(q)\|_2 = O_p(m^{-\gamma});\\
&\|\boldsymbol{v}_{L}(q)\|_2 = O_p(m^{-(q-1)}(T^{-1/2} + m^{-3/2}))
\end{aligned}$$

Proof: By Lemma C.9 , $\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}} = \sigma_j^2 + O_p(m^{-\gamma})$, for $j = 1, 2, ..., J$ and $h = 1, 2, ..., k(j)$. Thus, $\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}} - \lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}} = O_p(m^{-\gamma})$. Observe that

$$\begin{aligned}
&m^{\gamma}\left(\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}} - \lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}}\right) = O_p(1);\\
&(\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}})^{q-2} + (\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}})^{q-3}\lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}}\\
&\qquad\qquad + ... + \lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}}(\lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}})^{q-3} + (\lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}})^{q-2} = O_p(1)
\end{aligned}$$

Thus,

$$\begin{aligned}
&m^{\gamma}\left((\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}})^{q-1} - (\lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}})^{q-1}\right)\\
&\quad = m^{\gamma}(\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}} - \lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}})\Sigma_{j'=2}^{q-2}(\lambda_{ks(j-1)+h}^{\mathbf{S}_{NT}})^{q-j'}(\lambda_{ks(j-1)+1}^{\mathbf{S}_{NT}})^{j'}\\
&\quad = O_p(1) \times O_p(1) = O_p(1)
\end{aligned}$$

which implies that $\left\|(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\right\| = O_p(m^{-\gamma})$, where $\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}} = \mu_j^{\mathbf{S}_{NT}}\boldsymbol{I}_{k(j)}$. With this result, we can obtain the first part of the lemma because

$$\|\boldsymbol{v}_{H1}(q)\|_2 \leq \Sigma_{j=1}^{R}\left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}}\right\|_2\left\|(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\right\|_2\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right\|_2 = O_p(m^{-\gamma})$$

For $j \geq R+1$, $\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right\| = O_p(m^{-\gamma})$ by Lemma C.15. Thus, we have the second part of the lemma because

$$\|\boldsymbol{v}_{H2}(q)\|_2 \leq \Sigma_{j=R+1}^{J}\left\|\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}}\right\|_2\left\|(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\right\|_2\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right\|_2 = O_p(m^{-\gamma})$$

Finally, by Lemma C.18, we have

$$\begin{aligned}
\|v_L(q)\|_2 &= \left\|\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}}(\Lambda_L^{S_{NT}})^{q-1}\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2\\
&\leq \left\|\boldsymbol{\Xi}_L^{\mathbf{S}_{NT}}\right\|_2\left\|(\Lambda_L^{S_{NT}})^{q-1}\right\|_2\left\|\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2 \leq O_p(m^{-(q-1)})O_p(T^{-1/2} + m^{-3/2})
\end{aligned}$$

which implies the last part of the lemma. *Q.E.D.*

Corollary C.19: Under (A.1) – (A.6), for $q \geq 1$,

$$\|\boldsymbol{V}_{H1}(q)\|_F = O_p(m^{-\gamma}); \quad \|\boldsymbol{V}_{H2}(q)\|_F = O_p(m^{-\gamma}); \quad \|\boldsymbol{V}_L(q)\|_F = O_p(T^{-1/2} + m^{-3/2})$$

Proof: The results are obtained by Lemma C.19 because

$$\begin{aligned}
&\|\boldsymbol{V}_{H1}(q)\|_F \leq \Sigma_{j=1}^{R} \|\boldsymbol{v}_{H1}(j)\|_2;\\
&\|\boldsymbol{V}_{H2}(q)\|_F \leq \Sigma_{j=1}^{R} \|\boldsymbol{v}_{H2}(j)\|_2;\\
&\|\boldsymbol{V}_L(q)\|_F \leq q \times \|\boldsymbol{v}_L(1)\|_2
\end{aligned} \qquad Q.E.D.$$

Proof of Lemma 2.4.2: The parts (ii) – (iii) hold by Lemma C.19. For (i), observe that for each $j = 1, ..., R$,

$$\begin{aligned}
\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} \boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} &= \left(\frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} + \left(\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right) \right) \left(\boldsymbol{O}_{jj}^{*\prime} \sigma_j^2 \boldsymbol{\beta}_{(j)} + \left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{O}_{jj}^{*\prime} \sigma_j^2 \boldsymbol{\beta}_{(j)} \right) \right)\\
&= \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \sigma_j^2 \boldsymbol{\beta}_{(j)} + \left(\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right) \boldsymbol{O}_{jj}^{*\prime} \sigma_j^2 \boldsymbol{\beta}_{(j)}\\
&\quad + \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{O}_{jj}^{*\prime} \sigma_j^2 \boldsymbol{\beta}_{(j)} \right) + \left(\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right) \left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{O}_{jj}^{*\prime} \sigma_j^2 \boldsymbol{\beta}_{(j)} \right)
\end{aligned}$$

with Lemmas C.11 and C.15, it implies $\left\| \boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} \boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \sigma_j^2 N^{-1/2} \boldsymbol{\Phi}_{(j)} \boldsymbol{\beta}_{(j)} \right\|_F = O_p(m^{-\gamma})$. Thus, we have the desired result because

$$\|\boldsymbol{V}_0 - \boldsymbol{\Pi}_{NT} \boldsymbol{\Sigma}_R\|_F \leq \Sigma_{j=1}^{R} \left\| \boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}} \boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \sigma_j^2 N^{-1/2} \boldsymbol{\Phi}_{(j)} \boldsymbol{\beta}_{(j)} \right\|_F. \qquad Q.E.D.$$

Proof of Corollary 2.4.2: By Lemma 2.4.2 and Corollary C.19.

Proof of Theorem 1: By Lemma 2.4.1 and Corollary 2.4.2, $\|\boldsymbol{V}_L(q)\|_F = O_p(m^{-1/2})$. This result and Corollary 2.4.2 implies (i) because

$$\begin{aligned}
&\left\| \tilde{\boldsymbol{A}}_{1:q}^{PLS} - \boldsymbol{\Pi}_{NT} \boldsymbol{\Sigma}_R \boldsymbol{D}_0(q) \right\|_F\\
&\qquad \leq \|\boldsymbol{V}_0 \boldsymbol{D}_0(q) - \boldsymbol{\Pi}_{NT} \boldsymbol{\Sigma}_R \boldsymbol{D}_0(q)\|_F + \|\boldsymbol{V}_{H1}(q)\|_F + \|\boldsymbol{V}_{H2}(q)\|_F + \|\boldsymbol{V}_L(q)\|_F\\
&\qquad = O_p(m^{-\gamma}) + O_p(m^{-1/2}) = O_p(m^{-1/2})
\end{aligned}$$

For (ii), observe that

$$
\begin{aligned}
&N^{-1/2}\boldsymbol{\Pi}'_{NT}\boldsymbol{x}_{\boldsymbol{\cdot}T+1}\\
&\qquad = N^{-1}\begin{pmatrix}\boldsymbol{\beta}'_{(1)}\boldsymbol{\Phi}'_{(1)}\\ \vdots\\ \boldsymbol{\beta}'_{(R)}\boldsymbol{\Phi}'_{(R)}\end{pmatrix}\left(\Sigma_{j=1}^{R}\boldsymbol{\Phi}_{(j)}\boldsymbol{f}_{(j)T+1}+\Sigma_{j=R+1}^{J}\boldsymbol{\Phi}_{(j)}\boldsymbol{f}_{(j)T+1}+\boldsymbol{e}_{\boldsymbol{\cdot}T+1}\right)\\
&\qquad = \begin{pmatrix}\boldsymbol{\beta}'_{(1)}\boldsymbol{f}_{(1)T+1}\\ \vdots\\ \boldsymbol{\beta}'_{(R)}\boldsymbol{f}_{(R)T+1}\end{pmatrix}+\begin{pmatrix}O_p(N^{-\gamma})\\ \vdots\\ O_p(N^{-\gamma})\end{pmatrix}+\begin{pmatrix}\boldsymbol{\beta}'_{(1)}N^{-1}\boldsymbol{\Phi}'_{(1)}e_{\boldsymbol{\cdot}T+1}\\ \vdots\\ \boldsymbol{\beta}'_{(R)}N^{-1/2}\boldsymbol{\Phi}'_{(R)}e_{\boldsymbol{\cdot}T+1}\end{pmatrix}\\
&\qquad = \begin{pmatrix}\boldsymbol{\beta}'_{(1)}\boldsymbol{f}_{(1)T+1}\\ \vdots\\ \boldsymbol{\beta}'_{(R)}\boldsymbol{f}_{(R)T+1}\end{pmatrix}+\begin{pmatrix}O_p(N^{-\gamma})\\ \vdots\\ O_p(N^{-\gamma})\end{pmatrix}
\end{aligned}
$$

With this result and part (i), we can show

$$
\begin{aligned}
&\left\|N^{-1/2}\tilde{\boldsymbol{A}}_{1:q}^{PLS\prime}\boldsymbol{x}_{\boldsymbol{\cdot}T+1}-\boldsymbol{D}_0(q)'\boldsymbol{\Sigma}_R\boldsymbol{g}_{T+1}\right\|_F\\
&\qquad \le O_p(N^{-\gamma})+\left\|\tilde{\boldsymbol{A}}_{1:q}^{PLS}-\boldsymbol{\Pi}_{NT}\boldsymbol{\Sigma}_R\boldsymbol{D}_0(q)\right\|_F\left\|N^{-1/2}\boldsymbol{x}_{\boldsymbol{\cdot}T+1}\right\|_F=O_p(m^{-\gamma})
\end{aligned}
$$

Lemma C.20: The following results hold:

(i) $T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_0=(\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}},...,\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}})$;

(ii) $T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_{H1}(q)=\Sigma_{j=1}^{R}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}[(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}-(\mu_j^{\mathbf{S}_{NT}})^{q-1}\boldsymbol{I}_{k(j)}]\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}$;

(iii) $T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_{H2}(q)=\Sigma_{j=R+1}^{J}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}$;

(iv) $\boldsymbol{G}'_0\boldsymbol{G}_0=\boldsymbol{diag}\left(\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}},...,\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(R)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}\right)$;

(v) $\boldsymbol{G}'_0\boldsymbol{g}_{H1}(q)=\begin{pmatrix}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(1)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(1)}^{\mathbf{S}_{NT}})^{q-1}-(\bar{\boldsymbol{\Lambda}}_{(1)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}\\ \vdots\\ \boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(R)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(R)}^{\mathbf{S}_{NT}})^{q-1}-(\bar{\boldsymbol{\Lambda}}_{(R)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}\end{pmatrix}$;

(vi) $T^{-1/2}\boldsymbol{y}'\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q)=\tau_q$,

where $\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}}=\mu_j^{\mathbf{S}_{NT}}\boldsymbol{I}_{k(j)}$ and

$$
\begin{aligned}
\tau_q &= \Sigma_{j=1}^{R}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}-(\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\\
&\qquad -\Sigma_{j=1}^{R}\frac{\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}{\boldsymbol{c}_{(j}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}\left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}-(\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right).
\end{aligned}
$$

Proof: We can easily show (i) – (iii) using the fact that $\boldsymbol{\Xi}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{X}'\boldsymbol{y}/(N^{1/2}T) = \boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}$. The parts (iv) – (v) hold because

$$\begin{aligned}\boldsymbol{G}_0'\boldsymbol{G}_0 &= \left(\boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}, ..., \boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}\right)'\mathbf{S}_{NT}\left(\boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}, ..., \boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}\right)\\ &= \boldsymbol{diag}\left(\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}, ..., \boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(R)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}\right);\end{aligned}$$

$$\begin{aligned}\boldsymbol{G'}_0\boldsymbol{g}_{H1}(q) &= \left(\boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}, ..., \boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}\right)'\Sigma_{j=1}^{R}\boldsymbol{\Xi}_{(j)}^{S_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{S_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\\ &= \begin{pmatrix}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(1)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(1)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(1)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}\\ \vdots\\ \boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(R)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(R)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(R)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}\end{pmatrix}\end{aligned}$$

Finally, we can show

$$\begin{aligned}&\frac{\boldsymbol{y'G}_0}{T^{1/2}}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}\boldsymbol{G}_0'\boldsymbol{g}_{H1}(q)\\ &\qquad = \Sigma_{j=1}^{R}\frac{\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}{\boldsymbol{c}_{(j}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}\left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right);\end{aligned}$$

$$\begin{aligned}&T^{-1/2}\boldsymbol{y'}\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q)\\ &= T^{-1/2}\boldsymbol{y'g}_{H1}(q) - T^{-1/2}\boldsymbol{y'G}_0(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}\boldsymbol{G}_0'\boldsymbol{g}_{H1}(q)\\ &= \Sigma_{j=1}^{R}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\\ &- \Sigma_{j=1}^{R}\frac{\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}{\boldsymbol{c}_{(j}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}\left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right)\end{aligned}$$

which imply (vi). *Q.E.D.*

Lemma C.21: Under (A.1) – (A.8),

1. $\left\|\boldsymbol{G}_0 - T^{-1/2}(\boldsymbol{F}_{(1)}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{F}_{(R)}\boldsymbol{\beta}_{(R)})\boldsymbol{\Sigma}_R\right\|_F = O_p(m^{-\gamma})$;
2. $\left\|T^{-1/2}\boldsymbol{y'G}_0 - (\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{\beta}_{(R)}'\boldsymbol{\beta}_{(R)})\boldsymbol{\Sigma}_R^2\right\|_2 = O_p(m^{-\gamma})$.

Proof: Observe that for $j = 1, ..., R$,

$$
\begin{aligned}
&\frac{\boldsymbol{X}}{N^{1/2}T^{1/2}}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \sigma_j^2\frac{\boldsymbol{F}_{(j)}\boldsymbol{\beta}_{(j)}}{T^{1/2}} \\
&= \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\boldsymbol{O}^*_{jj}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} + \left(\frac{\boldsymbol{X}}{N^{1/2}T^{1/2}}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\boldsymbol{O}^*_{jj}\right)\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \sigma_j^2\frac{\boldsymbol{F}_{(j)}\boldsymbol{\beta}_{(j)}}{T^{1/2}} \\
&= \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\boldsymbol{O}^*_{jj}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} + \left(\frac{\boldsymbol{X}}{N^{1/2}T^{1/2}}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\boldsymbol{O}^*_{jj}\right)\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \sigma_j^2\frac{\boldsymbol{F}_{(j)}\boldsymbol{\beta}_{(j)}}{T^{1/2}} \\
&= \sigma_j^2\frac{\boldsymbol{F}_{(j)}\boldsymbol{\beta}_{(j)}}{T^{1/2}} + \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\left(\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{O}^*_{jj}{}'\sigma_j^2\boldsymbol{\beta}_{(j)}\right) + \left(\frac{\boldsymbol{X}}{N^{1/2}T^{1/2}}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\boldsymbol{O}^*_{jj}\right)\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} \\
&\qquad - \sigma_j^2\frac{\boldsymbol{F}_{(j)}\boldsymbol{\beta}_{(j)}}{T^{1/2}} \\
&= \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\left(\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{O}^*_{jj}{}'\sigma_j^2\boldsymbol{\beta}_{(j)}\right) + \left(\frac{\boldsymbol{X}}{N^{1/2}T^{1/2}}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}}\boldsymbol{O}^*_{jj}\right)\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}.
\end{aligned}
$$

This implies (i) because

$$
\begin{aligned}
&\left\|\boldsymbol{G}_0 - \frac{1}{T^{1/2}}(\boldsymbol{F}_{(1)}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{F}_{(R)}\boldsymbol{\beta}_{(R)})\boldsymbol{\Sigma}_R\right\|_F \\
&\qquad \le \Sigma_{j=1}^R\left\|\frac{\boldsymbol{X}}{N^{1/2}T^{1/2}}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \sigma_j^2\frac{\boldsymbol{F}_{(j)}\boldsymbol{\beta}_{(j)}}{T^{1/2}}\right\|_2 = O_p(m^{-\gamma})
\end{aligned}
$$

where the last equality is due to Lemma 2.4.1. For (ii), observe that

$$
\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_0 - (\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}, ..., \boldsymbol{\beta}'_{(R)}\boldsymbol{\beta}_{(R)})\boldsymbol{\Sigma}_R^2\right\|_2 \le \Sigma_{j=1}^R\left\|\frac{\boldsymbol{y}'}{T^{1/2}}\frac{\boldsymbol{X}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}}{(NT)^{1/2}} - \sigma_j^4\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)}\right\|_2
$$

Thus, we can also obtain (ii) by Lemma 2.4.1. *Q.E.D.*

Lemma C.22: Under (A.1) – (A.8), for $q \ge 1$,

(i) $\|\boldsymbol{g}^c_H(q)\|_2 = O_p(m^{-\gamma})$;

(ii) $\left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0(q)\right\|_2 = O_p(m^{-\gamma})$;

(iii) $\|\boldsymbol{g}_L(q)\|_2 = O_p\left(m^{-(q-1/2)}(T^{-1/2} + m^{-3/2})\right)$

Proof: When $q = 1$, $\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(1) = 0_{T\times 1}$. For $q \ge 2$,

$$
\begin{aligned}
\|\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q)\|_2 &\le \|\mathcal{Q}(\boldsymbol{G}_0)\|_2\left\|\frac{\boldsymbol{X}}{(NT)^{1/2}}\right\|_2\|\boldsymbol{v}_{H1}(q)\|_2 \\
&= \left\|\frac{\boldsymbol{X}}{(NT)^{1/2}}\right\|_2\|\boldsymbol{v}_{H1}(q)\|_2 = O_p(1)\times O_p(m^{-\gamma})
\end{aligned}
$$

by Lemma 2.4.2. The same lemma also implies

$$\|\boldsymbol{g}_{H2}(q)\|_2 \leq \left\| \frac{\boldsymbol{X}}{(NT)^{1/2}} \right\|_2 \|\boldsymbol{v}_{H1}(q)\|_2 = O_p(m^{-\gamma})$$

These results imply (i). Part (ii) holds by Lemma 2.4.2 because

$$\left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0(q)\right\|_2 \leq \left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}\boldsymbol{G}_0\right\|_2 \|\boldsymbol{g}_{H1}(q)\|_2 = O_p(m^{-\gamma}) \tag{C.9}$$

Part (ii) holds by Lemma 2.4.1 because

$$\|\boldsymbol{g}_L(q)\|_2 \leq \left\| \frac{\boldsymbol{X}}{(NT)^{1/2}} \boldsymbol{\Xi}_L^{\mathbf{S}_{NT}} \right\|_2 \left\|\boldsymbol{\Lambda}_L^{\mathbf{S}_{NT}}\right\|_2^{q-1} \left\|\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2 = O_p(m^{-(q-1/2)}(T^{-1/2} + m^{-3/2}))$$

This completes the proof. *Q.E.D.*

Corollary C.22: Under (A.1) – (A.8), for $q \geq 1$,

(i) $\|\boldsymbol{G}_H^c(q)\|_F = O_p(m^{-\gamma})$;

(ii) $\left\|\hat{\boldsymbol{D}}_0(q) - \boldsymbol{D}_0(q)\right\|_F = O_p(m^{-\gamma})$;

(iii) $\|\boldsymbol{G}_L(q)\|_F = O_p\left(m^{-1/2}(T^{-1/2} + m^{-3/2})\right)$

Proof: Parts (i) and (ii) hold by Lemma C.22 because

$$\|\boldsymbol{G}_H^c(q)\|_F \leq \Sigma_{j=1}^q \|\boldsymbol{g}_H^c(j)\|_2; \quad \left\|\hat{\boldsymbol{D}}_0(q) - \boldsymbol{D}_0(q)\right\|_F \leq \Sigma_{j=1}^q \left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0(q)\right\|_2$$

Part (iii) also holds by Lemma C.22 because $\|\boldsymbol{G}_L(q)\|_F \leq \Sigma_{j=1}^q \|\boldsymbol{g}_L(j)\|_2 \leq q \times \|\boldsymbol{g}_L(1)\|_2$. *Q.E.D.*

Lemma C.23: Let

$$\rho_{j,q} = \left[\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} \left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right)\right] / \boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}$$

Under (A.1) – (A.8), for $j = 1, 2, ..., R$ and $q = 1, 2, ...,$

$$\rho_{q,j} = \left[\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right] / \mu_j^{\mathbf{S}_{NT}} + O_p(m^{-2\gamma})$$

Proof: Observe that by Lemma 2.4.1, $\left\|\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}} - \bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}}\right\|_F = O_p(m^{-\gamma})$ for $j = 1, ..., R$. With this result, we can show

$$\rho_{q,j} - \frac{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}{\mu_j^{\mathbf{S}_{NT}}}$$

$$= \boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} \left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right) \left(\frac{1}{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}} - \frac{1}{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}\right)$$

$$= \boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} \frac{\left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}(\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}} - \boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right)\left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right)}{\left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right)\left(\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right)}$$

$$= O_p(m^{-2\gamma})$$

which completes the proof. *Q.E.D.*

Lemma C.24: Under (A.1) – (A.8), for $q \geq 1$,

(i) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_H^c(q)\right\|_2 = O_p(m^{-2\gamma})$;

(ii) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(q)\right\|_2 = O_p\left(m^{-(q-1)}(T^{-1/2} + m^{-3/2})\right)$;

Proof: We can obtain (i) by showing that

(C.10) $\left\|T^{-1/2}\boldsymbol{y}'\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q)\right\|_2 = O_p(m^{-2\gamma})$;
(C.11) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_{H2}(q)\right\|_2 = O_p(m^{-2\gamma})$.

Consider τ_q defined in Lemma C.20. By Lemma C.23 and Lemma 2.4.1, for $q = 1, 2, ..., R$,

$$\begin{aligned}
\tau_q &= \Sigma_{j=1}^{R}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} - \Sigma_{j=1}^{R}\rho_j \\
&= \Sigma_{j=1}^{R}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} \\
&\quad - \Sigma_{j=1}^{R}\frac{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}{\mu_j^{\mathbf{S}_{NT}}} + O_p(m^{-2\gamma}) \\
&= \Sigma_{j=1}^{R}\frac{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}\prime}(\bar{\boldsymbol{\Lambda}}_{(j)}^{S_{NT}} - \boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})((\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\mathbf{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}{\mu_j^{\mathbf{S}_{NT}}} + O_p(m^{-2\gamma}) = O_p(m^{-2\gamma})
\end{aligned}$$

Thus, (C.10) holds because, by Lemma C.20, $T^{-1/2}\boldsymbol{y}'\mathcal{Q}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q) = \tau_q$. Finally, by Lemma C.20 and Lemma 2.4.1, (C.11) also holds because

$$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_{H2}(q)\right\|_2 = \Sigma_{j=R+1}^{J}\left\|(\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}})^{q-1}\right\|_2\left\|\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}\right\|_2^2 = O_p(m^{-2\gamma})$$

We can obtain (ii) because

$$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(q)\right\|_2 = \left\|\boldsymbol{c}_L^{\mathbf{S}_{NT}\prime}(\boldsymbol{\Lambda}_L^{\mathbf{S}_{NT}})^{q-1}\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2 \leq O_p(m^{-(q-1)})\left\|\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2^2 \qquad Q.E.D.$$

Corollary C.24: Under (A.1) – (A.8), for $q \geq 1$,

(i) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}^c_H(q)\right\|_2 = O_p(m^{-2\gamma})$;

(ii) $\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_L(q)\right\|_2 = O_p\left((T^{-1/2}+m^{-3/2})^2\right)$

Proof: Observe that

$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}^c_H(q)\right\|_2 \leq \Sigma_{j=1}^{q}\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}^c_H(j)\right\|_2$;

$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_L(q)\right\|_2 \leq \Sigma_{j=1}^{R}\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(q)\right\|_2 \leq q \times \left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(1)\right\|_2$

Thus, (i) and (ii) hold by Lemma C.23. *Q.E.D.*

Proof of Lemma 2.4.3: Parts (i) and (ii) hold by Lemma C.21. Parts (iii) and (iv) hold by Lemma C.22. Part (vi) holds by Lemma C.23. *Q.E.D.*

Proof of Corollary 2.4.3: The results hold by Corollaries C.22 and C.24. *Q.E.D.*

Lemma C.25: Under (A.1) – (A.8), $\left\|[\hat{\boldsymbol{D}}_0(R)]^{-1} - [\boldsymbol{D}_0(R)]^{-1}\right\|_F = O_p(m^{-\gamma})$.

Proof: The result holds by Corollary 2.4.3 and Lemma A.2 because

$$\begin{aligned}\left\|[\hat{\boldsymbol{D}}_0(R)]^{-1} - [\boldsymbol{D}_0(R)]^{-1}\right\|_F &= \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1}\right\|_F\left\|\hat{\boldsymbol{D}}_0(R) - \boldsymbol{D}_0(R)\right\|_F\left\|[\boldsymbol{D}_0(R)]^{-1}\right\|_F \\ &= O_p(m^{-\gamma})\end{aligned}$$

Lemma C.26: Under (A.1) – (A.8), $\left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} - \boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2 = O_p(m^{-\gamma})$.

Proof: Let $\hat{\boldsymbol{\Sigma}}_R = \boldsymbol{diag}(\mu_1^{\mathbf{S}_{NT}}, ..., \mu_R^{\mathbf{S}_{NT}})$. Observe that for $j = 1, ..., R$,

$$\left|\frac{\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}}{\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{\Lambda}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(1)}} - \frac{\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}}{\mu_j^{\mathbf{S}_{NT}}\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}}\right| = O_p(m^{-\gamma}) \tag{C.12}$$

because

$$\begin{aligned}&\left|\frac{\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}}{\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{\Lambda}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(1)}} - \frac{\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}}{\mu_j^{\mathbf{S}_{NT}}\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}}\right| \\ &\qquad\leq \frac{\left|\mu_j^{\mathbf{S}_{NT}}\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{\Lambda}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(1)}\right|}{(\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{\Lambda}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(1)})\mu_j^{\mathbf{S}_{NT}}} \\ &\qquad\leq \left\|\frac{1}{(\boldsymbol{c}^{\mathbf{S}_{NT}\prime}_{(j)}\boldsymbol{\Lambda}^{\mathbf{S}_{NT}}_{(j)}\boldsymbol{c}^{\mathbf{S}_{NT}}_{(1)})\mu_j^{\mathbf{S}_{NT}}}\right\|_2\left\|\bar{\boldsymbol{\Lambda}}^{\mathbf{S}_{NT}}_{(j)} - \boldsymbol{\Lambda}^{\mathbf{S}_{NT}}_{(j)}\right\|_F\left\|\boldsymbol{c}^{\mathbf{S}_{NT}}_{(j)}\right\|_2^2 \\ &\qquad= O_p(1) \times O_p(m^{-\gamma}) \times O_p(1) = O_p(m^{-\gamma})\end{aligned}$$

With (C.12), we have

(C.13) $\left\|\left(\boldsymbol{G}_0'\boldsymbol{G}_0\right)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}-\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R\right\|_2=O_p(m^{-\gamma})$

because

$$\left\|\left(\boldsymbol{G}_0'\boldsymbol{G}_0\right)^{-1}\frac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}}-\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R\right\|_2$$

$$=\left\|\left(\frac{\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}{\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}{}'\boldsymbol{\Lambda}_{(1)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}\quad\cdots\quad\frac{\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}}{\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}{}'\boldsymbol{\Lambda}_{(R)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}}\right)'-\left(\frac{\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}{\mu_1^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}\quad\cdots\quad\frac{\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}}{\mu_1^{\mathbf{S}_{NT}}\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}}\right)'\right\|_2$$

$$\leq\Sigma_{j=1}^{R}\left|\frac{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}{}'\boldsymbol{\Lambda}_{(j)}^{\mathbf{S}_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}}-\frac{\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}{\mu_j^{\mathbf{S}_{NT}}\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}{}'\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}}}\right|$$

By Lemma 2.4.1, $1/\mu_j^{\mathbf{S}_{NT}}-1/\sigma_j^2=(\sigma_j^2-\mu_j^{\mathbf{S}_{NT}})/(\mu_j^{\mathbf{S}_{NT}}\sigma_j^2)=O_p(m^{-\gamma})$. Thus,

(C.14) $\left\|\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R-\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2=O_p(m^{-\gamma})$.

By (C.13) and (C.14), we can obtain the desired result because

$$\left\|\left(\boldsymbol{G}_0'\boldsymbol{G}_0\right)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}-\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2$$
$$\leq\left\|\left(\boldsymbol{G}_0'\boldsymbol{G}_0\right)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}-\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R\right\|_2+\left\|\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R-\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2\qquad Q.E.D.$$

Remark: Because $\boldsymbol{G}_0$, $\boldsymbol{G}_H^c(q)$, and $\boldsymbol{G}_L(q)$ are mutually orthogonal by construction, we can obtain the following results:

$$\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\tilde{\boldsymbol{P}}_{1:R}^{PLS}}{NT}-\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)=\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)+\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R);$$
$$\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\tilde{\boldsymbol{p}}_{R+1}^{PLS}}{NT}-\hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1)=\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1)+\boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1);$$
$$\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\boldsymbol{y}}{N^{1/2}T}-\hat{\boldsymbol{D}}_0(R)'T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}=T^{-1/2}\boldsymbol{G}_H^c(R)'\boldsymbol{y}+T^{-1/2}\boldsymbol{G}_L(R)'\boldsymbol{y};$$
$$\frac{\tilde{\boldsymbol{p}}_{R}^{PLS}{}'\boldsymbol{y}}{N^{1/2}T}-\hat{\boldsymbol{d}}_0(R)'\frac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}}=\frac{\boldsymbol{g}_H^c(R)'\boldsymbol{y}}{T^{1/2}}+\frac{\boldsymbol{g}_L(R)'\boldsymbol{y}}{T^{1/2}};$$
$$\frac{\tilde{\boldsymbol{p}}_{R+1}^{PLS}{}'\boldsymbol{y}}{N^{1/2}T}-\hat{\boldsymbol{d}}_0(R+1)'\frac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}}=\frac{\boldsymbol{g}_H^c(R+1)'\boldsymbol{y}}{T^{1/2}}+\frac{\boldsymbol{g}_L(R+1)'\boldsymbol{y}}{T^{1/2}}$$

Lemma C.27: Under (A.1) – (A.8),

$$\left\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)\right\|_F=O_p(m^{-2\gamma});$$
$$\left\|\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1)\right\|_2=O_p(m^{-2\gamma});\quad\left\|\boldsymbol{G}_H^c(R)'\boldsymbol{y}/T\right\|_2=O_p(m^{-2\gamma});$$
$$\left\|T^{-1/2}\boldsymbol{g}_H^c(R)'\boldsymbol{y}\right\|_2=O_p(m^{-2\gamma});\quad\left\|T^{-1/2}\boldsymbol{g}_H^c(R+1)'\boldsymbol{y}\right\|_2=O_p(m^{-2\gamma})$$

Proof: The results hold by Lemma 2.4.3 and Corollary 2.4.3. *Q.E.D.*

Lemma C.28: Under (A.1) – (A.8),

$$\begin{aligned}
&\left\|\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F = O_p\left(m^{-1}(T^{-1/2}+m^{-3/2})^2\right);\\
&\left\|\boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right\|_2 = O_p\left(m^{-R-1}(T^{-1/2}+m^{-3/2})^2\right);\\
&\left\|T^{-1/2}\boldsymbol{G}_L(R)'\boldsymbol{y}\right\|_2 = O_p\left((T^{-1/2}+m^{-3/2})^2\right);\\
&\left\|T^{-1/2}\boldsymbol{g}_L(R)'\boldsymbol{y}\right\|_2 = O_p\left(m^{-R-1}(T^{-1/2}+m^{-3/2})^2\right);\\
&\left\|T^{-1/2}\boldsymbol{g}_L(R+1)'\boldsymbol{y}\right\|_2 = O_p\left(m^{-R}(T^{-1/2}+m^{-3/2})^2\right);\\
&\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(q)\right\| = O_p\left(m^{-(q-1)}(T^{-1/2}+m^{-3/2})^2\right)
\end{aligned}$$

Proof: The results hold by Lemma 2.4.3 and Corollary 2.4.3. *Q.E.D.*

Proof of Lemma 2.4.4: All the results hold by Lemmas C.27 and C.28. *Q.E.D.*

Lemma C.29: Let $\boldsymbol{d}_0^*(q) = \boldsymbol{\Sigma}_R^{q-1}\mathbf{1}_R$ and $\boldsymbol{D}_0^*(q) = (\boldsymbol{d}_0^*(1), ..., \boldsymbol{d}_0^*(q))$. Under (A.1) – (A.8), as $m \to \infty$,

(i) $\hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) \to_p \boldsymbol{\Psi}^* \equiv \boldsymbol{D}_0^*(R)'\boldsymbol{Diag}(\sigma_1^6\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)}, ..., \sigma_1^6\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)})\boldsymbol{D}_0^*(R)$;

(ii) $\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) \to_p \boldsymbol{\psi}^* \equiv \boldsymbol{D}_0^*(R)'\boldsymbol{Diag}(\sigma_1^6\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)}, ..., \sigma_1^6\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)})\boldsymbol{d}_0^*(R+1)$

(iii) $T^{-1/2}\hat{\boldsymbol{D}}_0(R)'\boldsymbol{G'}_0\boldsymbol{y} \to_p \boldsymbol{\pi}^* \equiv \boldsymbol{D}_0^*(R)'\left(\sigma_1^4\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)}, ..., \sigma_1^4\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)}\right)'$.

Proof: By Lemma 2.4.1, for $j = 1, ..., R$, we have

(C.15) $\hat{\boldsymbol{\Lambda}}_{(j)}^{S_{NT}} \to_p \sigma_j^2\boldsymbol{I}_{k(j)}$; $\boldsymbol{c}_{(j)}^{\mathbf{S}_{NT}} \to_p \boldsymbol{O}_{jj}^*\sigma_j^2\boldsymbol{\beta}_{(j)}$

In addition, it is straightforward to show

(C.16) $\boldsymbol{G}_0'\boldsymbol{G}_0 = \boldsymbol{V'}_0\boldsymbol{\Lambda}_{H1}^{S_{NT}}\boldsymbol{V}_0 = \boldsymbol{Diag}(\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}{}'\boldsymbol{\Lambda}_{(1)}^{S_{NT}}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}, ..., \boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}{}'\boldsymbol{\Lambda}_{(R)}^{S_{NT}}\boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}})$;

(C.17) $\dfrac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}} = \begin{pmatrix}\boldsymbol{c}_{(1)}^{\mathbf{S}_{NT}}{}'\boldsymbol{\Xi}_{(1)}^{\mathbf{S}_{NT}}{}'\\ \vdots \\ \boldsymbol{c}_{(R)}^{\mathbf{S}_{NT}}{}'\boldsymbol{\Xi}_{(R)}^{\mathbf{S}_{NT}}{}'\end{pmatrix}$ $\dfrac{\boldsymbol{X}'\boldsymbol{y}}{N^{1/2}T} = \begin{pmatrix}\boldsymbol{c}_{(1)}^{S_{NT}}{}'\boldsymbol{c}_{(1)}^{S_{NT}}\\ \vdots \\ \boldsymbol{c}_{(R)}^{S_{NT}}{}'\boldsymbol{c}_{(R)}^{S_{NT}}\end{pmatrix}$

By Lemmas 2.4.1 and 2.4.3,

(C.18) $\left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0^*(q)\right\| \le \left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0(q)\right\| + \|\boldsymbol{d}_0(q) - \boldsymbol{d}_0^*(q)\| = O_p(m^{-\gamma})$

Then, the desired results can be obtained by (C.15) – (C.18) and Lemma 2.4.1. *Q.E.D.*

Proof of Theorem 2: We begin by considering the cases in which $R < K$. Observe that

$$[\hat{\boldsymbol{D}}_0(R)]^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G'}_0\boldsymbol{y} = \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\hat{\boldsymbol{D}}_0(R)'T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} \tag{C.19}$$

With this result and Lemma 2.4.4, we can show

$$\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\hat{\boldsymbol{D}}_0(R)]^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 = O_p(m^{-2\gamma}) \tag{C.20}$$

Lemma C.26 and (C.19) imply

$$\begin{aligned}
&\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\hat{\boldsymbol{D}}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\qquad \leq \left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\hat{\boldsymbol{D}}_0(R)]^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 \\
&\qquad + \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1}\right\|_F\left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} - \boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\qquad = O_p(m^{-2\gamma}) + O_p(m^{-\gamma}) = O_p(m^{-\gamma})
\end{aligned}$$

By this result and Lemma C.25, we can obtain (i) because

$$\begin{aligned}
&\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\quad = \left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\hat{\boldsymbol{D}}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R + [\hat{\boldsymbol{D}}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\| \\
&\quad \leq \left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\hat{\boldsymbol{D}}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 + \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1} - [\boldsymbol{D}_0(R)]^{-1}\right\|_F\left\|\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\quad = O_p(m^{-\gamma}) + O_p(m^{-\gamma}) = O_p(m^{-\gamma})
\end{aligned}$$

When $R = K$, we have $\hat{\boldsymbol{D}}_0(R) = \boldsymbol{D}_0(R)$. With this result, (C.19) and Lemma 2.4.4, we can obtain

$$\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\boldsymbol{D}_0(R)]^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 = O_p\left(m^{-R-1}(T^{-1/2} + m^{-3/2})^2\right)$$

By this result and Lemma C.25, we can show that (i) holds even when $R = K$, because

$$\begin{aligned}
&\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\qquad \leq \left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - [\boldsymbol{D}_0(R)]^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 \\
&\qquad + \left\|[\boldsymbol{D}_0(R)]^{-1}\right\|_F\left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} - \boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\qquad = O_p\left(m^{-R-1}(T^{-1/2} + m^{-3/2})^2\right) + O_p(m^{-\gamma}) = O_p(m^{-\gamma})
\end{aligned}$$

The part (ii) is obtained by Theorem 1 and the part (i) because

$$\begin{aligned}\left\|\tilde{y}^{PLS}_{T+2|R} - \hat{y}^{o}_{T+2}\right\|_2 &= \left\|\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R} N^{1/2}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R} - \hat{y}^{o}_{T+2}\right\|_2 \\ &= \left\|\boldsymbol{g}'_{T+1}\boldsymbol{\Sigma}_R\boldsymbol{D}_0(R)[\boldsymbol{D}_0(R)]^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R - \hat{y}^{o}_{T+2}\right\|_2 + O_p(m^{-\gamma}) \\ &= 0 + O_p(m^{-\gamma})\end{aligned}$$

Finally, for (iii), observe that by Lemmas 2.4.4 and C.29, we have

$$\begin{aligned}\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}}{N^{1/2}T} &= T^{-1/2}(\hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{y} + \boldsymbol{G}^c_H(R)'\boldsymbol{y} + \boldsymbol{G}_L(R)'\boldsymbol{y}) \\ &\to_p \boldsymbol{D}^*_0(R)\big(\sigma_1^4\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}, ..., \sigma_1^4\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}\big)';\end{aligned}$$

$$\begin{aligned}\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} &= \hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) + \boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R) \\ &\to_p \boldsymbol{D}^*_0(R)'\boldsymbol{Diag}(\sigma_1^6\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}, ..., \sigma_1^6\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)})\boldsymbol{D}^*_0(R)\end{aligned}$$

By these two results, we can obtain

$$\frac{\boldsymbol{y}'\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}}{T} = \frac{\boldsymbol{y}'\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{N^{1/2}T}\left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}}{N^{1/2}T}\to_p\Sigma_{j=1}^{R}\sigma_j^2\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)}$$

which implies (iii). *Q.E.D.*

Lemma C.30: Let $\tilde{\boldsymbol{\theta}} \equiv \left(\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}\right)^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1}$. Then,

(i) $\left\|\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 = O_p(m^{-\gamma})$, if $R < K$;

(ii) $\left\|\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 = O_p\left(m^{-1}(T^{-1/2} + m^{-3/2})^2\right)$, if $R = K$.

Proof: Using Lemma B.2 and some algebra, we can show

$$\begin{aligned}&\left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \\ &\quad= [\hat{\boldsymbol{D}}_0(R)]^{-1}\hat{\boldsymbol{d}}_0(R+1) + \boldsymbol{A}_1\left(\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\ &\qquad+ \boldsymbol{A}_1\left(\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)\boldsymbol{a}_2\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{A}_1 &= \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}; \\ \boldsymbol{a}_2 &= \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\boldsymbol{D}_0(R) + \boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)^{-1} \\ &\quad\times\left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) + \boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right)\end{aligned}$$

By Lemmas C.29 and 2.4.3, $\boldsymbol{A}_1 \to_p (\boldsymbol{\Psi}^*)^{-1}$ and $\boldsymbol{a}_2 \to_p (\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\psi}^*$. Thus, we can have

$$\text{(C.21)}\quad \left\|\tilde{\boldsymbol{\theta}} - [\hat{\boldsymbol{D}}_0(R)]^{-1}\hat{\boldsymbol{d}}_0(R+1)\right\|_2 = O_p\left(\left\|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F\right)$$

We begin by proving (ii). When $R = K$, we have $\hat{\boldsymbol{D}}_0(R) = \boldsymbol{D}_0(R)$, $\hat{\boldsymbol{d}}_0(R+1) = \boldsymbol{d}_0(R+1)$, and $\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R) = \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)$. By substituting these results into (C.21) and applying Lemma C.28, we can obtain (ii).

For (i), observe that $\left\|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F = O_p(m^{-2\gamma})$ because $\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R)$ asymptotically dominates $\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)$. Thus, from (C.20), we have

$$\text{(C.22)}\quad \left\|\tilde{\boldsymbol{\theta}} - [\hat{\boldsymbol{D}}_0(R)]^{-1}\hat{\boldsymbol{d}}_0(R+1)\right\|_2 = O_p(m^{-2\gamma})$$

Now, by Lemma C.25 and Corollary 2.4.3, we can show

$$\begin{aligned}
&\left\|[\hat{\boldsymbol{D}}_0(R)]^{-1}\hat{\boldsymbol{d}}_0(R+1) - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 \\
&\leq \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1}\hat{\boldsymbol{d}}_0(R+1) - [\hat{\boldsymbol{D}}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 \\
&\qquad + \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1}\boldsymbol{d}_0(R+1) - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 \\
&\leq \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1}\right\|_F\left\|\hat{\boldsymbol{d}}_0(R+1) - \boldsymbol{d}_0(R+1)\right\|_2 + \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1} - [\boldsymbol{D}_0(R)]^{-1}\right\|_F\|\boldsymbol{d}_0(R+1)\|_2 \\
&= O_p(m^{-\gamma})
\end{aligned}$$

which, with (C.26), implies

$$\begin{aligned}
&\left\|\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 \\
&\qquad \leq \left\|\tilde{\boldsymbol{\theta}} - [\hat{\boldsymbol{D}}_0(R)]^{-1}\hat{\boldsymbol{d}}_0(R+1)\right\|_2 + \left\|[\hat{\boldsymbol{D}}_0(R)]^{-1}\hat{\boldsymbol{d}}_0(R+1) - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 \\
&\qquad = O_p(m^{-2\gamma}) + O_p(m^{-\gamma}) = O_p(m^{-\gamma})
\end{aligned}$$

This completes the proof. *Q.E.D.*

Lemma C.31: Define $\mathcal{Y}_{1,NT} = \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1}/(NT)$. Under (A.1) – (A.8),

$$\begin{aligned}
\mathcal{Y}_{1,NT} &= \boldsymbol{g}^c_H(R+1)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1) \\
&\quad - \left(\boldsymbol{g}^c_H(R+1)'\boldsymbol{G}^c_H(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right)\boldsymbol{a}_3 \\
&\quad - \boldsymbol{a}_4'\left(\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\
&\quad + \boldsymbol{a}_4'\left(\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)\boldsymbol{a}_2
\end{aligned}$$

where $\boldsymbol{a}_2$ is defined in Lemma C.30, $\boldsymbol{a}_3 \to_p (\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\psi}^*$, and $\boldsymbol{a}_4 \to_p (\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\psi}^*$.

Proof: Using Lemma B.2, we can show

$$\begin{aligned}
\mathcal{Y}_{1,NT} &= \frac{\tilde{\boldsymbol{p}}_{R+1}^{PLS\prime}\tilde{\boldsymbol{p}}_{R+1}^{PLS}}{NT} - \frac{\tilde{\boldsymbol{p}}_{R+1}^{PLS\prime}\tilde{\boldsymbol{P}}_{1:R}^{PLS}}{NT}\left(\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS\prime}\tilde{\boldsymbol{P}}_{1:R}^{PLS}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS\prime}\tilde{\boldsymbol{p}}_{R+1}^{PLS}}{NT} \\
&= \boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1) \\
&\quad - \left(\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right)\left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1} \\
&\quad \times \hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) \\
&\quad - \left(\hat{\boldsymbol{d}}_0(R+1)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) + \boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right) \\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\left(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\
&\quad + \left(\hat{\boldsymbol{d}}_0(R+1)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) + \boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right) \\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\left(\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right) \\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\boldsymbol{D}_0(R) + \boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)^{-1} \\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) + \boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\
&= \boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1) \\
&\quad - \left(\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right)\boldsymbol{a}_3 \\
&\quad - \boldsymbol{a}_4'\left(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\
&\quad + \boldsymbol{a}_4'\left(\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)\boldsymbol{a}_2
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{a}_3 &= \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) \to_p (\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\psi}^*; \\
\boldsymbol{a}'_4 &= \left(\hat{\boldsymbol{d}}_0(R+1)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) + \boldsymbol{g}_H^*(R+1)'\boldsymbol{G}_H^*(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right) \\
&\qquad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1} \to_p \boldsymbol{\psi}^{*\prime}(\boldsymbol{\Psi}^*)^{-1}
\end{aligned}$$

by Lemmas C.29 and 2.4.3. *Q.E.D.*

Corollary C.31: Assume that (A.1) – (A.8) hold. When $R < K$,

(i) $\mathcal{Y}_{1,NT} = O_p(m^{-2\gamma})$ if $R < K$;

(ii) $\mathcal{Y}_{1,NT} = O_p\left(m^{-1}(T^{-1/2} + m^{-3/2})^2\right)$, if $R = K$.

Proof: When $R < K$,

$$\begin{aligned}&\left\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1)+\boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\right\|_2\\&\qquad\le\left\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1)\right\|_2+\left\|\boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\right\|_2\\&\qquad=O_p(m^{-2\gamma})+O_p(m^{-2(R+1/2)})O_p\left((T^{-1/2}+m^{-3/2})^2\right)=O_p(m^{-2\gamma});\end{aligned}$$

$$\begin{aligned}&\left\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R)+\boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right\|_2\\&\qquad\le\left\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R)\right\|_2+\left\|\boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right\|_2\\&\qquad=O_p(m^{-2\gamma})+O_p(m^{-R-1})O_p\left((T^{-1/2}+m^{-3/2})^2\right)=O_p(m^{-2\gamma});\end{aligned}$$

$$\begin{aligned}&\left\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)+\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F\\&\qquad\le\left\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)\right\|_F+\left\|\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F\\&\qquad=O_p(m^{-2\gamma})+O_p\left(m^{-1}(T^{-1/2}+m^{-3/2})^2\right)=O_p(m^{-2\gamma})\end{aligned}$$

which imply (i). When $R = K$,

$$\begin{aligned}&\left\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1)+\boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\right\|_2\\&\qquad=\left\|\boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\right\|_2=O_p(m^{-2(R+1/2)})O_p\left((T^{-1/2}+m^{-3/2})^2\right);\end{aligned}$$

$$\begin{aligned}&\left\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R)+\boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right\|_2\\&\qquad\le\left\|\boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right\|_2=O_p(m^{-R-1})O_p\left((T^{-1/2}+m^{-3/2})^2\right);\end{aligned}$$

$$\begin{aligned}&\left\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)+\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_2\\&\qquad=\left\|\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_2=O_p(m^{-1})O_p\left((T^{-1/2}+m^{-3/2})^2\right)\end{aligned}$$

which imply (ii). *Q.E.D.*

Lemma C.32: Define $\mathcal{Y}_{2,NT}=\boldsymbol{y}'\mathcal{Q}(\tilde{\boldsymbol{P}}_{1:R}^{PLS})\tilde{\boldsymbol{p}}_{R+1}^{PLS}/(N^{1/2}T)$. Under (A.1) – (A.8),

$$\begin{aligned}\mathcal{Y}_{2,NT}=T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_H^c(R+1)+T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(R+1)-T^{-1/2}\left(\boldsymbol{y}'\boldsymbol{G}_H^c(R)+\boldsymbol{y}'\boldsymbol{G}_L(R)\right)\boldsymbol{a}_3\\-\boldsymbol{a}_5'\left(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1)+\boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right)\\+\boldsymbol{a}_5'(\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)+\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R))\boldsymbol{a}_2\end{aligned}$$

where $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ are defined in Lemmas C.30 and C.31, and $\boldsymbol{a}_5\to_p(\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\pi}^*$.

Proof: Using Lemma B.2, we can show

$$
\begin{aligned}
\mathcal{Y}_{2,NT} &= \frac{\boldsymbol{y}'\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{N^{1/2}T} - \frac{\boldsymbol{y}'\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{N^{1/2}T}\left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}}{NT^{1/2}} \\
&= \frac{\boldsymbol{y}'\boldsymbol{g}^*_H(R+1)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}} \\
&\quad - \left(\frac{\boldsymbol{y}'\boldsymbol{G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}}\right)\left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}'_0\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}'_0\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1)\right) \\
&\quad - \left(\frac{\boldsymbol{y}'\boldsymbol{G}_0}{T^{1/2}}\hat{\boldsymbol{D}}_0(R) + \frac{\boldsymbol{y}'\boldsymbol{G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}}\right)\left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}'_0\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1} \\
&\quad \times \left(\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\
&\quad + \left(\frac{\boldsymbol{y}'\boldsymbol{G}_0}{T^{1/2}}\hat{\boldsymbol{D}}_0(R) + \frac{\boldsymbol{y}'\boldsymbol{G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}}\right) \\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}'{}_0\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}(\boldsymbol{G}^*_H(R)'\boldsymbol{G}^*_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)) \\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}'_0\boldsymbol{G}_0\boldsymbol{D}_0(R) + \boldsymbol{G}^*_H(R)'\boldsymbol{G}^*_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)^{-1} \\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}'_0\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) + \boldsymbol{G}^*_H(R)'\boldsymbol{g}^*_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\
&= \frac{\boldsymbol{y}'\boldsymbol{g}^c_H(R+1)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}} - \left(\frac{\boldsymbol{y}'\boldsymbol{G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}}\right)\boldsymbol{a}_3 \\
&\quad - \boldsymbol{a}'_5\left(\boldsymbol{G}^*_H(R)'\boldsymbol{g}^*_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right) \\
&\quad + \boldsymbol{a}'_5(\boldsymbol{G}^*_H(R)'\boldsymbol{G}^*_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R))\boldsymbol{a}_2
\end{aligned}
$$

where

$$\boldsymbol{a}_5 = \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}'_0\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\left(\hat{\boldsymbol{D}}_0(R)\frac{\boldsymbol{G}'_0\boldsymbol{y}}{T^{1/2}} + \frac{\boldsymbol{G}^*_H(R)'\boldsymbol{y}}{T^{1/2}} + \frac{\boldsymbol{G}_L(R)'\boldsymbol{y}}{T^{1/2}}\right) \to_p (\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\pi}^*$$

by Lemmas C.29 and 2.4.3. *Q.E.D.*

Corollary C.32: Under (A.1) – (A.8), When $R < K$,

(i) $\mathcal{Y}_{2,NT} = O_p(m^{-2\gamma})$, if $R < K$;

(ii) $\mathcal{Y}_{2,NT} = O_p\left((T^{-1/2} + m^{-3/2})^2\right)$, if $R = K$.

Proof: When $R < K$,

$$
\begin{aligned}
&\left\|\frac{\boldsymbol{y}'\boldsymbol{g}^c_H(R+1)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}}\right\|_2 \\
&\qquad \le \left\|\frac{\boldsymbol{y}'\boldsymbol{g}^c_H(R+1)}{T^{1/2}}\right\|_2 + \left\|\frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}}\right\|_2 \\
&\qquad = O_p(m^{-2\gamma}) + O_p(m^{-R})O_p((T^{-1} + m^{-3/2})^2) = O_p(m^{-2\gamma});
\end{aligned}
$$

$$\left\|\frac{\boldsymbol{y'G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y'G}_L(R)}{T^{1/2}}\right\|_2$$
$$\leq \left\|\frac{\boldsymbol{y'G}^c_H(R)}{T^{1/2}}\right\|_2 + \left\|\frac{\boldsymbol{y'G}_L(R)}{T^{1/2}}\right\|_2 = O_p(m^{-2\gamma}) + O_p((T^{-1} + m^{-3/2})^2) = O_p(m^{-2\gamma});$$

$$\left\|\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right\|_2$$
$$\leq \left\|\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1)\right\|_2 + \left\|\boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right\|_2$$
$$= O_p(m^{-2\gamma}) + O_p(m^{-R-1})O_p((T^{-1} + m^{-3/2})^2) = O_p(m^{-2\gamma});$$

$$\left\|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F$$
$$\leq \left\|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R)\right\|_F + \left\|\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F$$
$$= O_p(m^{-2\gamma}) + O_p(m^{-1})O_p((T^{-1} + m^{-3/2})^2) = O_p(m^{-2\gamma})$$

which imply (i). When $R = K$,

$$\left\|\frac{\boldsymbol{y'g}^c_H(R+1)}{T^{1/2}} + \frac{\boldsymbol{y'g}_L(R+1)}{T^{1/2}}\right\|_2 = \left\|\frac{\boldsymbol{y'g}_L(R+1)}{T^{1/2}}\right\|_2 = O_p(m^{-R})O_p\left((T^{-1/2} + m^{-3/2})^2\right)$$
$$\left\|\frac{\boldsymbol{y'G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y'G}_L(R)}{T^{1/2}}\right\|_2 = \left\|\frac{\boldsymbol{y'G}_L(R)}{T^{1/2}}\right\|_2 = O_p\left((T^{-1/2} + m^{-3/2})^2\right);$$
$$\left\|\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right\|_2$$
$$= \left\|\boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right\|_2 = O_p(m^{-R-1})O_p\left((T^{-1/2} + m^{-3/2})^2\right);$$
$$\left\|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right\|_F = O_p(m^{-1})O_p\left((T^{-1/2} + m^{-3/2})^2\right)$$

which imply (ii). *Q.E.D.*

Proof of Lemma 2.4.5: The parts (i) and (iii) hold by Lemma C.30. The parts (ii) and (vi) hold by Corollaries C.31 and C.32.

Proof of Theorem 3: Observe that

$$\frac{\boldsymbol{y'}\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R+1})\boldsymbol{y}}{T} = \frac{\boldsymbol{y'}\mathcal{P}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}}{T}$$
$$+ \frac{y'\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{N^{1/2}T}\left(\frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}}{N^{1/2}T}$$
$$= \frac{\boldsymbol{y'}\mathcal{P}(\tilde{P}^{PLS}_{1:R})\boldsymbol{y}}{T} + \frac{(\mathcal{Y}_{1,NT})^2}{\mathcal{Y}_{2,NT}} = \Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)} + \frac{(\mathcal{Y}_{1,NT})^2}{\mathcal{Y}_{2,NT}} + o_p(1)$$

where the last equality is due to Theorem 2. When $R = K$,

$$(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = O_p(((m/T)^{1/2} + m^{-1})^2) = O_p(m/T)$$

by Corollaries C.31 and C.32. If $m/T \to \infty$, $(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = o_p(1)$. If $m/T = O(1) > 0$, $(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = |O_p(1)| > 0$. Finally, when $R < K$, $(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = O_p(m^{-2\gamma}) = o_p(1)$. This completes the proof of the theorem. *Q.E.D.*

Lemma C.33: Let $\mathcal{Y}_{NT} = \mathcal{Y}_{2,NT}/\mathcal{Y}_{1,NT}$. Then,

$$N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R+1} = \begin{pmatrix} N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} \\ 0 \end{pmatrix} - \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix} \mathcal{Y}_{NT}$$

Proof: Observe that

$$\text{(C.23)} \quad N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R+1} = \left(\frac{\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}{}' \tilde{\boldsymbol{P}}^{PLS}_{1:R+1}}{NT} \right)^{-1} \frac{\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}{}' \boldsymbol{y}}{N^{1/2}T}$$

$$= \begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R} \tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R} \tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R} \boldsymbol{y}}{N^{1/2}T} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \boldsymbol{y}}{N^{1/2}T} \end{pmatrix}.$$

By the inversion rule for partitioned matrix,

$$\text{(C.24)} \quad \begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R} \tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R} \tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \end{pmatrix}^{-1} = \begin{pmatrix} \left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R} \tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} \right)^{-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix}$$

$$+ \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix} \left(\frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R}) \tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \right)^{-1} \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}'$$

where $\tilde{\boldsymbol{\theta}}$ is defined in Lemma 2.4.5. In addition,

$$\text{(C.25)} \quad \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}' \begin{pmatrix} \tilde{\boldsymbol{P}}^{PLS\prime}_{1:R} \boldsymbol{y}/(N^{1/2}T) \\ \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \boldsymbol{y}/(N^{1/2}T) \end{pmatrix} = -\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1} \mathcal{Q}(\tilde{\boldsymbol{P}}^{PLS}_{1:R}) \boldsymbol{y}/(N^{1/2}T)$$

We can obtain the desired result by substituting (C.24) and (C.25) into (C.23). *Q.E.D.*

Proof of Theorem 4: The part (i) holds by Lemma C.34. For (ii), observe that

$$\begin{pmatrix} [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix}' \begin{pmatrix} \boldsymbol{D}_0(R)' \\ \boldsymbol{d}_0(R+1)' \end{pmatrix} = \mathbf{0}_{1\times R}$$

Using this fact and (i), we can easily show

$$
\begin{aligned}
&\tilde{\boldsymbol{y}}^{PLS}_{T+2|R+1} - \tilde{\boldsymbol{y}}^{o}_{T+2|R} \\
&= \boldsymbol{x}'_{\cdot T+1}\tilde{\boldsymbol{A}}^{PLS}_{1:R+1}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R+1} - \boldsymbol{x}'_{\cdot T+1}\tilde{\boldsymbol{A}}^{PLS}_{1:R}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R} \\
&= \frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}(\tilde{\boldsymbol{A}}^{PLS}_{1:R}, \tilde{\boldsymbol{\alpha}}^{PLS}_{R+1})\begin{pmatrix} N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} \\ 0 \end{pmatrix} + \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R+1}\begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix} - \frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R}N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} \\
&= \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R+1}\begin{pmatrix} [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix} \\
&+ \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R}(\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)) \\
&= \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}(\boldsymbol{V}_{H1}(R+1) + \boldsymbol{V}_{H2}(R+1) + \boldsymbol{V}_{L}(R+1))\begin{pmatrix} [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix} \\
&+ \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R}(\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)).
\end{aligned}
$$

When $R < K$,

$$
\begin{aligned}
&\mathcal{Y}_{NT} = O_p(1); \\
&\|\boldsymbol{V}_{H1}(R+1) + \boldsymbol{V}_{H2}(R+1) + \boldsymbol{V}_{L}(R+1)\|_F = O_p(m^{-\gamma}); \\
&\left\|\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 = O_p(m^{-\gamma}).
\end{aligned}
$$

Thus, we have $\left\|\tilde{y}^{PLS}_{T+2|R+1}\tilde{\boldsymbol{y}}^{o}_{R+1} - \tilde{y}^{PLS}_{T+2|R}\right\|_2 = O_p(m^{-\gamma})$, which implies (i) because

$$
\left\|\tilde{y}^{PLS}_{T+2|R+1} - \hat{y}^{o}_{T+2}\right\|_2 \leq \left\|\tilde{y}^{PLS}_{T+2|R+1} - \tilde{y}^{PLS}_{T+2|R}\right\|_2 + \left\|\tilde{y}^{PLS}_{T+2|R} - \hat{y}^{o}_{T+2}\right\|_2 = O_p(m^{-\gamma})
$$

When $R = K$,

$$
\begin{aligned}
&\mathcal{Y}_{NT} = O_p(m); \\
&\|\boldsymbol{V}_{H1}(R+1) + \boldsymbol{V}_{H2}(R+1) + \boldsymbol{V}_{L}(R+1)\|_F = \|\boldsymbol{V}_{L}(R+1)\|_F; \\
&\left\|\tilde{\boldsymbol{\theta}} - [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1)\right\|_2 = O_p(m^{-1}(T^{-1/2} + m^{-3/2})^2)
\end{aligned}
$$

With these results, we have

$$
\begin{aligned}
\text{(C.26)}\quad &\left\|\tilde{\boldsymbol{y}}^{PLS}_{T+2|R+1} - \tilde{\boldsymbol{y}}^{PLS}_{T+2|R} - \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\boldsymbol{V}_L(R+1)\begin{pmatrix} [\boldsymbol{D}_0(R)]^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix}\right\|_2 \\
&\qquad = O_p((T^{-1/2} + m^{-3/2})^2)
\end{aligned}
$$

Notice that by Lemma C.18,

$$
\text{(C.27)}\quad \left\|\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\boldsymbol{v}_L(1)\right\|_2 = \left\|\frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}}\boldsymbol{\Xi}^{\mathbf{S}_{NT}}_{L}\boldsymbol{c}^{\mathbf{S}_{NT}}_{L}\right\|_2 = O_p((mT)^{-1/2} + m^{-3/2})
$$

because

$$\begin{aligned}\left\|N^{-1/2}\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\Xi_L^{\mathbf{S}_{NT}}\right\|_2 &= \left(N^{-1}\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\Xi_L^{\mathbf{S}_{NT}}\Xi_L^{S_{NT}\prime}\boldsymbol{x}_{\boldsymbol{\cdot}T+1}\right)^{1/2}\\ &\leq \left(N^{-1}\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\boldsymbol{x}_{\boldsymbol{\cdot}T+1}\right)^{1/2} = O_p(1)\end{aligned}$$

For $q \geq 2$,

$$\begin{aligned}\left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{v}_L(q)\right\|_2 &= \left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\Xi_L^{\mathbf{S}_{NT}}\left(\Lambda_L^{\mathbf{S}_{NT}}\right)^{q-1}\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2\\ &\leq \left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\Xi_L^{\mathbf{S}_{NT}}\right\|_2\left\|\left(\Lambda_L^{\mathbf{S}_{NT}}\right)^{q-1}\right\|_2\left\|\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2\\ &\leq O_p(1)O_p(m^{-(q-1)})O_p((T^{-1/2}+m^{-3/2})) \leq O_p(m^{-3/2})\end{aligned}$$

Thus, we have

$$\begin{aligned}\left\|\mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{V}_L(R+1)\right\|_2 &\leq (R+1)\times\left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\Xi_L^{\mathbf{S}_{NT}}\boldsymbol{c}_L^{\mathbf{S}_{NT}}\right\|_2|\mathcal{Y}_{NT}|\\ &\leq O_p(m^{1/2}T^{-1/2}+m^{-1/2})\end{aligned} \tag{C.28}$$

By (C.26) and (C.28), we can have

$$\left\|\tilde{\boldsymbol{y}}^{PLS}_{T+2|R+1} - \tilde{\boldsymbol{y}}^{PLS}_{T+2|R}\right\|_2 = O_p(m^{1/2}T^{-1/2}+m^{-1/2})$$

which implies the parts (ii) and (iii). *Q.E.D.*

A.4 Tables and Figures

Number of PLS factors used (q)	Mean in-sample R^2	Standard error of in-sample R^2	Out-of-sample R^2
PLS1	62.55	4.58	50.64
PLS2	72.88	3.74	53.03
PLS3	78.68	3.24	49.32
PLS4	82.02	3.00	46.36
PLS5	84.35	2.78	41.10
PLS6	86.08	2.61	36.91
PLS7	87.40	2.50	33.10
PLS8	88.48	2.40	28.87
PLS9	89.39	2.33	24.45
PLS10	90.17	2.25	18.77

Table A.1: In-Sample and Out-of-Sample Percentage R^2 of PLS Regressions ($R = 2$, $K = 4$)

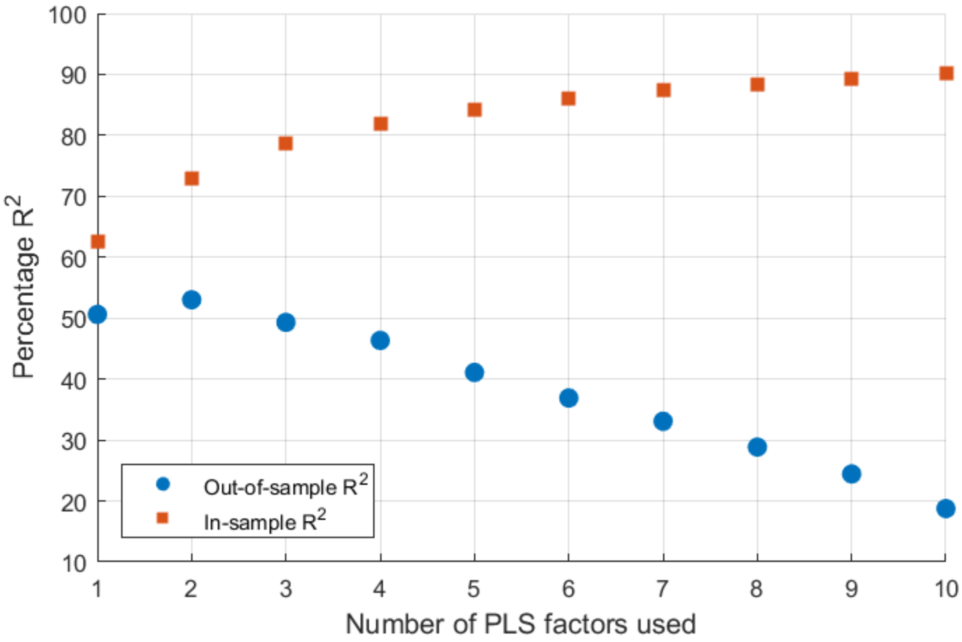

Figure A.1: Graphical Representation of Table A.1

Notes: Spurious correlation as PLS iteration proceeds. There are four total factors and two informative factors, so the forecasting power is maximized at the second iteration. After the second iteration, the out-of-sample forecasting power significantly decreases while the adjusted in-sample fit always increases as more PLS factors are used.

$T = 100, N = 80$										
a_x	a_y	ρ_c	ρ_e	ρ_f	PLS1	PC5	OLS	$\tilde{a}_x$	$\tilde{a}_y$	$\tilde{K}$
0.1	0.3	0	0	0	0.137	**0.162**	-2.851	0.102	0.116	2.015
0.1	0.3	0.3	0.3	0.3	**0.112**	0.08	-3.043	0.147	0.095	2.629
0.1	0.3	0.5	0.5	0.5	**0.112**	0.072	-2.871	0.207	0.098	3.191
0.1	0.5	0	0	0	0.241	**0.245**	-2.268	0.102	0.181	1.999
0.1	0.5	0.3	0.3	0.3	**0.292**	0.169	-1.671	0.15	0.142	2.679
0.1	0.5	0.5	0.5	0.5	**0.287**	0.151	-2.034	0.203	0.141	3.139
0.1	0.7	0	0	0	**0.35**	0.321	-1.79	0.102	0.247	2.001
0.1	0.7	0.3	0.3	0.3	**0.394**	0.25	-1.23	0.151	0.197	2.687
0.1	0.7	0.5	0.5	0.5	**0.399**	0.205	-0.815	0.196	0.179	3.012
0.2	0.3	0	0	0	0.21	**0.23**	-2.691	0.216	0.222	3.336
0.2	0.3	0.3	0.3	0.3	0.214	**0.222**	-2.925	0.185	0.167	2.531
0.2	0.3	0.5	0.5	0.5	**0.233**	0.216	-2.763	0.192	0.142	2.402
0.2	0.5	0	0	0	0.359	**0.377**	-1.835	0.216	0.342	3.33
0.2	0.5	0.3	0.3	0.3	**0.366**	0.359	-1.831	0.181	0.26	2.457
0.2	0.5	0.5	0.5	0.5	**0.354**	0.297	-1.719	0.188	0.22	2.325
0.2	0.7	0	0	0	0.518	**0.525**	-1.135	0.213	0.467	3.279
0.2	0.7	0.3	0.3	0.3	**0.522**	0.482	-0.883	0.18	0.356	2.451
0.2	0.7	0.5	0.5	0.5	**0.547**	0.479	-0.619	0.193	0.309	2.405
0.3	0.3	0	0	0	0.23	**0.239**	-2.706	0.364	0.295	4.748
0.3	0.3	0.3	0.3	0.3	0.205	**0.218**	-3.044	0.312	0.246	3.69
0.3	0.3	0.5	0.5	0.5	**0.254**	0.25	-3.158	0.276	0.208	2.921
0.3	0.5	0	0	0	0.404	**0.405**	-1.559	0.363	0.457	4.737
0.3	0.5	0.3	0.3	0.3	0.364	**0.368**	-1.892	0.311	0.393	3.665
0.3	0.5	0.5	0.5	0.5	**0.396**	0.389	-1.914	0.277	0.335	2.934
0.3	0.7	0	0	0	0.599	**0.611**	-0.889	0.363	0.624	4.743
0.3	0.7	0.3	0.3	0.3	**0.595**	0.589	-0.692	0.313	0.539	3.678
0.3	0.7	0.5	0.5	0.5	**0.586**	0.578	-0.551	0.275	0.45	2.895
0.4	0.3	0	0	0	0.244	**0.254**	-2.456	0.462	0.309	4.987
0.4	0.3	0.3	0.3	0.3	**0.278**	0.277	-2.504	0.456	0.303	4.735
0.4	0.3	0.5	0.5	0.5	**0.31**	0.302	-2.964	0.417	0.27	3.984
0.4	0.5	0	0	0	0.406	**0.428**	-1.599	0.462	0.482	4.987
0.4	0.5	0.3	0.3	0.3	0.434	**0.443**	-1.643	0.454	0.472	4.704
0.4	0.5	0.5	0.5	0.5	0.442	**0.443**	-1.558	0.417	0.43	3.992
0.4	0.7	0	0	0	0.588	**0.603**	-0.54	0.462	0.661	4.993
0.4	0.7	0.3	0.3	0.3	0.617	**0.623**	-0.661	0.459	0.651	4.776
0.4	0.7	0.5	0.5	0.5	0.626	**0.632**	-0.659	0.42	0.591	4.024
0.5	0.3	0	0	0	**0.282**	0.278	-2.553	0.551	0.315	5
0.5	0.3	0.3	0.3	0.3	0.266	**0.275**	-2.174	0.556	0.32	4.987
0.5	0.3	0.5	0.5	0.5	**0.245**	0.243	-3.096	0.554	0.312	4.781
0.5	0.5	0	0	0	0.454	**0.466**	-1.531	0.551	0.495	5
0.5	0.5	0.3	0.3	0.3	0.446	**0.447**	-1.646	0.556	0.5	4.973
0.5	0.5	0.5	0.5	0.5	0.489	**0.5**	-1.459	0.557	0.493	4.825
0.5	0.7	0	0	0	0.622	**0.639**	-0.575	0.551	0.678	4.999
0.5	0.7	0.3	0.3	0.3	0.627	**0.64**	-0.602	0.555	0.678	4.963
0.5	0.7	0.5	0.5	0.5	0.628	**0.645**	-0.864	0.555	0.668	4.806

0.6	0.3	0	0	0	**0.256**	0.255	-2.348	0.641	0.326	5
0.6	0.3	0.3	0.3	0.3	**0.198**	0.196	-2.848	0.645	0.324	4.998
0.6	0.3	0.5	0.5	0.5	**0.279**	0.275	-2.562	0.652	0.321	4.985
0.6	0.5	0	0	0	0.477	**0.495**	-1.417	0.641	0.506	5
0.6	0.5	0.3	0.3	0.3	0.486	**0.497**	-1.511	0.645	0.509	5
0.6	0.5	0.5	0.5	0.5	0.433	**0.45**	-1.728	0.652	0.508	4.981
0.6	0.7	0	0	0	0.641	**0.66**	-0.591	0.641	0.688	5
0.6	0.7	0.3	0.3	0.3	0.631	**0.646**	-0.636	0.645	0.692	4.999
0.6	0.7	0.5	0.5	0.5	0.65	**0.673**	-0.542	0.651	0.692	4.977

Table A.2: Forecasting Performances of PLS, PC, and OLS Regressions ($R = 1$, $K = 5$, $T = 100, N = 80$))

Notes: This table shows the forecasting performances of the PLS1, PC5 and OLS regressions under different data processes. For all cases, data are generated with $N = 80$ and $T = 100$. For each case, the best performance is marked by bold. The total number of factors in predictor variables is five and $\boldsymbol{\Omega}^*_{\boldsymbol{F}} = 5 \times \boldsymbol{I}_5$, so that the optimal number of PLS and PC factors are respectively one and five ($R = 1$ and $K = 5$). The term $\tilde{K}$ denotes the estimated number of the factors in predictor variables by Ahn and Horenstein (2013), $\tilde{a}_x$ is the average in-sample R^2 from regressions of individual predictors on the $\tilde{K}$ PC factors, and $\tilde{a}_y$ is the average in-sample R^2 from regressions of the target variable on the $\tilde{K}$ PC factors.

$T = 200, N = 160$										
a_x	a_y	ρ_c	ρ_e	ρ_f	PLS1	PC5	OLS	$\tilde{a}_x$	$\tilde{a}_y$	$\tilde{K}$
0.1	0.3	0	0	0	0.175	**0.207**	-2.989	0.132	0.209	4.168
0.1	0.3	0.3	0.3	0.3	**0.197**	0.181	-2.664	0.082	0.111	2.15
0.1	0.3	0.5	0.5	0.5	**0.185**	0.132	-2.608	0.095	0.1	2.203
0.1	0.5	0	0	0	0.354	**0.364**	-1.847	0.134	0.339	4.228
0.1	0.5	0.3	0.3	0.3	**0.369**	0.306	-1.676	0.084	0.186	2.181
0.1	0.5	0.5	0.5	0.5	**0.287**	0.206	-2.252	0.095	0.16	2.211
0.1	0.7	0	0	0	**0.524**	0.514	-1.212	0.136	0.468	4.291
0.1	0.7	0.3	0.3	0.3	**0.496**	0.406	-1.17	0.083	0.256	2.16
0.1	0.7	0.5	0.5	0.5	**0.485**	0.332	-0.752	0.094	0.22	2.192
0.2	0.3	0	0	0	0.207	**0.225**	-2.326	0.244	0.281	5
0.2	0.3	0.3	0.3	0.3	0.269	**0.271**	-2.538	0.246	0.272	4.894
0.2	0.3	0.5	0.5	0.5	**0.209**	0.208	-2.993	0.214	0.222	3.862
0.2	0.5	0	0	0	0.436	**0.441**	-1.849	0.244	0.452	5
0.2	0.5	0.3	0.3	0.3	**0.407**	0.403	-1.653	0.246	0.437	4.88
0.2	0.5	0.5	0.5	0.5	0.389	**0.393**	-2.01	0.208	0.352	3.741
0.2	0.7	0	0	0	0.605	**0.624**	-0.743	0.244	0.629	5
0.2	0.7	0.3	0.3	0.3	0.547	**0.556**	-0.7	0.246	0.598	4.873
0.2	0.7	0.5	0.5	0.5	**0.562**	0.553	-0.616	0.21	0.489	3.777
0.3	0.3	0	0	0	0.262	**0.264**	-2.221	0.337	0.294	5
0.3	0.3	0.3	0.3	0.3	0.282	**0.283**	-2.49	0.341	0.292	5
0.3	0.3	0.5	0.5	0.5	0.272	**0.279**	-2.941	0.349	0.292	4.988
0.3	0.5	0	0	0	**0.473**	0.472	-1.495	0.337	0.479	5
0.3	0.5	0.3	0.3	0.3	0.433	**0.456**	-1.68	0.341	0.472	5
0.3	0.5	0.5	0.5	0.5	0.441	**0.458**	-1.542	0.349	0.47	4.982
0.3	0.7	0	0	0	0.636	**0.656**	-0.817	0.337	0.661	5
0.3	0.7	0.3	0.3	0.3	0.626	**0.647**	-0.617	0.341	0.655	5
0.3	0.7	0.5	0.5	0.5	0.605	**0.62**	-0.7	0.347	0.643	4.959
0.4	0.3	0	0	0	**0.306**	0.305	-2.691	0.431	0.302	5
0.4	0.3	0.3	0.3	0.3	**0.256**	0.249	-2.986	0.434	0.303	5
0.4	0.3	0.5	0.5	0.5	0.291	**0.302**	-2.56	0.44	0.3	5
0.4	0.5	0	0	0	0.467	**0.496**	-1.485	0.431	0.491	5
0.4	0.5	0.3	0.3	0.3	0.438	**0.447**	-1.412	0.434	0.49	5
0.4	0.5	0.5	0.5	0.5	0.475	**0.496**	-1.75	0.44	0.485	5
0.4	0.7	0	0	0	0.641	**0.668**	-0.567	0.431	0.677	5
0.4	0.7	0.3	0.3	0.3	0.634	**0.655**	-0.722	0.434	0.673	5
0.4	0.7	0.5	0.5	0.5	0.627	**0.645**	-0.733	0.44	0.67	5
0.5	0.3	0	0	0	0.27	**0.278**	-2.359	0.526	0.308	5
0.5	0.3	0.3	0.3	0.3	**0.285**	0.282	-2.689	0.528	0.31	5
0.5	0.3	0.5	0.5	0.5	0.268	**0.279**	-2.978	0.532	0.305	5
0.5	0.5	0	0	0	0.429	**0.439**	-1.447	0.526	0.497	5
0.5	0.5	0.3	0.3	0.3	0.479	**0.496**	-1.525	0.527	0.497	5
0.5	0.5	0.5	0.5	0.5	0.501	**0.516**	-1.756	0.532	0.496	5
0.5	0.7	0	0	0	0.641	**0.669**	-0.738	0.526	0.686	5
0.5	0.7	0.3	0.3	0.3	0.644	**0.665**	-0.632	0.528	0.685	5
0.5	0.7	0.5	0.5	0.5	0.65	**0.666**	-0.788	0.533	0.685	5

0.6	0.3	0	0	0	0.264	**0.278**	-2.522	0.621	0.31	5
0.6	0.3	0.3	0.3	0.3	0.314	**0.324**	-2.55	0.622	0.311	5
0.6	0.3	0.5	0.5	0.5	0.311	**0.315**	-3.382	0.626	0.31	5
0.6	0.5	0	0	0	0.493	**0.503**	-1.552	0.62	0.504	5
0.6	0.5	0.3	0.3	0.3	0.464	**0.485**	-1.78	0.622	0.499	5
0.6	0.5	0.5	0.5	0.5	0.465	**0.472**	-1.994	0.626	0.5	5
0.6	0.7	0	0	0	0.649	**0.672**	-0.542	0.621	0.694	5
0.6	0.7	0.3	0.3	0.3	0.668	**0.688**	-0.582	0.622	0.693	5
0.6	0.7	0.5	0.5	0.5	0.671	**0.693**	-0.768	0.626	0.691	5

Table A.3: Forecasting Performances of PLS, PC, and OLS Regressions ($R = 1$, $K = 5$, $T = 200, N = 160$))

Notes: This table shows the forecasting performances of the PLS1, PC5 and OLS regressions under different data processes. For all cases, data are generated with $N = 160$ and $T = 200$. For each case, the best performance is marked by bold. The total number of factors in predictor variables is five and $\boldsymbol{\Omega}^*_{\boldsymbol{F}} = 5 \times \boldsymbol{I}_5$, so that the optimal number of PLS and PC factors are respectively one and five ($R = 1$ and $K = 5$). The term $\tilde{K}$ denotes the estimated number of the factors in predictor variables by Ahn and Horenstein (2013), $\tilde{a}_x$ is the average in-sample R^2 from regressions of individual predictors on the $\tilde{K}$ PC factors, and $\tilde{a}_y$ is the average in-sample R^2 from regressions of the target variable on the $\tilde{K}$ PC factors.

Data generating parameters				$T = N = 100$			$T = N = 200$		
a_x	ρ_d	ρ_e	ρ_f	PLS1	PLS2	PLS3	PLS1	PLS2	PLS3
0.1	0	0	0	**0.462**	0.397	0.29	**0.612**	0.549	0.5
0.1	0	0	0.3	**0.445**	0.374	0.243	**0.6**	0.511	0.457
0.1	0	0	0.5	**0.484**	0.421	0.3	**0.593**	0.517	0.455
0.1	0.3	0.3	0	**0.431**	0.421	0.359	**0.612**	0.569	0.519
0.1	0.3	0.3	0.3	**0.458**	0.45	0.376	**0.598**	0.561	0.52
0.1	0.3	0.3	0.5	**0.424**	0.395	0.333	**0.603**	0.56	0.51
0.1	0.5	0.5	0	0.426	**0.455**	0.446	**0.604**	0.59	0.517
0.1	0.5	0.5	0.3	0.448	**0.462**	0.423	**0.604**	0.594	0.536
0.1	0.5	0.5	0.5	0.416	**0.427**	0.397	**0.581**	0.565	0.524
0.2	0	0	0	0.567	**0.569**	0.448	**0.643**	0.624	0.556
0.2	0	0	0.3	0.573	**0.574**	0.485	**0.664**	0.639	0.577
0.2	0	0	0.5	**0.562**	0.561	0.471	**0.646**	0.612	0.568
0.2	0.3	0.3	0	0.561	**0.576**	0.524	**0.662**	0.649	0.595
0.2	0.3	0.3	0.3	0.585	**0.592**	0.528	**0.672**	0.654	0.624
0.2	0.3	0.3	0.5	0.565	**0.567**	0.5	**0.664**	0.644	0.599
0.2	0.5	0.5	0	**0.554**	0.55	0.506	**0.632**	0.595	0.573
0.2	0.5	0.5	0.3	0.55	**0.551**	0.513	**0.64**	0.613	0.582
0.2	0.5	0.5	0.5	0.559	**0.58**	0.53	**0.634**	0.623	0.582
0.3	0	0	0	**0.616**	0.612	0.525	**0.665**	0.638	0.584
0.3	0	0	0.3	0.632	**0.64**	0.56	**0.696**	0.668	0.609
0.3	0	0	0.5	0.629	**0.636**	0.54	**0.677**	0.671	0.611
0.3	0.3	0.3	0	0.598	**0.613**	0.545	**0.654**	0.627	0.585
0.3	0.3	0.3	0.3	0.574	**0.605**	0.54	**0.67**	0.653	0.603
0.3	0.3	0.3	0.5	0.621	**0.637**	0.581	**0.698**	0.682	0.641
0.3	0.5	0.5	0	0.596	**0.608**	0.566	**0.686**	0.666	0.626
0.3	0.5	0.5	0.3	0.568	**0.596**	0.538	**0.651**	0.643	0.605
0.3	0.5	0.5	0.5	0.61	**0.626**	0.6	**0.662**	0.643	0.594

Table A.4: Forecasting by PLS Regressions with Different Numbers of Factors ($R = K = 3$, $N = T = 100$ and $N = T = 200$)

Notes: This table reports the forecasting performances of three regressions with three different numbers of informative PLS factors: one (PLS1), two (PLS2), and three (PLS3). For each data specification, the highest out-of-sample R-square is in bold. The other parameters used to generate the data are set at $a_y = 0.7$ and $\boldsymbol{\Omega}^*_{\boldsymbol{F}} = \boldsymbol{diag}(3, 5, 7)$.

Data generating parameters				$T = N = 1000$			$T = N = 2000$		
a_x	ρ_d	ρ_e	ρ_f	PLS1	PLS2	PLS3	PLS1	PLS2	PLS3
0.1	0	0	0	0.651	**0.669**	0.614	0.703	**0.719**	0.692
0.1	0	0	0.3	0.667	**0.684**	0.648	0.657	**0.672**	0.624
0.1	0	0	0.5	0.67	**0.7**	0.664	0.696	**0.704**	0.666
0.1	0.3	0.3	0	0.631	**0.663**	0.612	0.641	**0.654**	0.63
0.1	0.3	0.3	0.3	**0.681**	0.68	0.637	0.692	**0.701**	0.669
0.1	0.3	0.3	0.5	0.627	**0.651**	0.611	0.667	**0.676**	0.644
0.1	0.5	0.5	0	0.635	**0.652**	0.618	0.68	**0.695**	0.67
0.1	0.5	0.5	0.3	0.681	**0.703**	0.672	0.716	**0.723**	0.697
0.1	0.5	0.5	0.5	0.646	**0.666**	0.628	0.684	**0.695**	0.668
0.2	0	0	0	0.674	**0.694**	0.678	0.667	**0.68**	0.665
0.2	0	0	0.3	0.701	**0.719**	0.7	0.683	**0.691**	0.676
0.2	0	0	0.5	0.655	**0.685**	0.666	0.657	**0.666**	0.649
0.2	0.3	0.3	0	0.7	**0.722**	0.708	0.714	**0.727**	0.721
0.2	0.3	0.3	0.3	0.673	**0.704**	0.686	0.687	**0.704**	0.685
0.2	0.3	0.3	0.5	0.685	**0.707**	0.681	0.698	**0.703**	0.693
0.2	0.5	0.5	0	0.651	**0.688**	0.674	0.688	**0.709**	0.702
0.2	0.5	0.5	0.3	0.657	**0.682**	0.657	0.683	**0.691**	0.681
0.2	0.5	0.5	0.5	0.664	**0.692**	0.666	0.719	**0.73**	0.724
0.3	0	0	0	0.626	**0.656**	0.648	0.686	**0.699**	0.696
0.3	0	0	0.3	0.651	**0.687**	0.677	0.701	**0.715**	0.714
0.3	0	0	0.5	0.65	**0.683**	0.678	0.681	**0.69**	0.689
0.3	0.3	0.3	0	0.664	**0.7**	0.686	0.695	**0.712**	0.71
0.3	0.3	0.3	0.3	0.643	0.674	**0.675**	0.675	**0.693**	0.685
0.3	0.3	0.3	0.5	0.666	**0.696**	0.686	0.665	**0.689**	0.681
0.3	0.5	0.5	0	0.67	**0.695**	0.687	0.689	0.71	**0.71**
0.3	0.5	0.5	0.3	0.653	**0.678**	0.666	0.69	**0.705**	0.697
0.3	0.5	0.5	0.5	0.666	**0.676**	0.667	0.676	**0.692**	0.689

Table A.5: Forecasting by PLS Regressions with Different Numbers of Factors ($R = K = 3$, $N = T = 1000$ and $N = T = 2000$)

Notes: This table reports the forecasting performances of three regressions with three different numbers of informative PLS factors: one (PLS1), two (PLS2), and three (PLS3). For each data specification, the highest out-of-sample R-square is in bold. The other parameters used to generate the data are set at $a_y = 0.7$ and $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 5, 7)$.

Data generating parameters				$T = N = 7000$		
a_x	ρ_d	ρ_e	ρ_f	PLS1	PLS2	PLS3
0.1	0	0	0	0.612	**0.617**	0.613
0.1	0	0	0.3	0.58	**0.59**	0.586
0.1	0	0	0.5	0.579	**0.591**	0.589
0.1	0.3	0.3	0	0.61	**0.616**	0.613
0.1	0.3	0.3	0.3	0.579	**0.589**	0.586
0.1	0.3	0.3	0.5	0.578	**0.591**	0.59
0.1	0.5	0.5	0	0.611	**0.617**	0.615
0.1	0.5	0.5	0.3	0.579	**0.591**	0.588
0.1	0.5	0.5	0.5	0.578	**0.593**	0.592
0.2	0	0	0	0.611	**0.615**	0.614
0.2	0	0	0.3	0.579	**0.588**	0.587
0.2	0	0	0.5	0.578	0.59	**0.59**
0.2	0.3	0.3	0	0.61	**0.615**	0.614
0.2	0.3	0.3	0.3	0.578	**0.588**	0.587
0.2	0.3	0.3	0.5	0.577	**0.59**	0.59
0.2	0.5	0.5	0	0.61	**0.616**	0.616
0.2	0.5	0.5	0.3	0.578	**0.589**	0.589
0.2	0.5	0.5	0.5	0.577	0.591	**0.592**
0.3	0	0	0	0.61	**0.615**	0.614
0.3	0	0	0.3	0.578	**0.587**	0.587
0.3	0	0	0.5	0.577	0.589	**0.59**
0.3	0.3	0.3	0	0.609	**0.615**	0.614
0.3	0.3	0.3	0.3	0.577	**0.587**	0.587
0.3	0.3	0.3	0.5	0.576	0.589	**0.59**
0.3	0.5	0.5	0	0.61	**0.615**	0.615
0.3	0.5	0.5	0.3	0.577	**0.588**	0.588
0.3	0.5	0.5	0.5	0.576	0.59	**0.591**

Table A.6: Forecasting by PLS Regressions with Different Numbers of Factors ($R = K = 3, N = T = 7000$)

Notes: This table reports the forecasting performances of three regressions with three different numbers of informative PLS factors: one (PLS1), two (PLS2), and three (PLS3). For each data specification, the highest out-of-sample R-square is in bold. The other parameters used to generate the data are set at $a_y = 0.7$ and $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 5, 7)$. To save computation time, only 100 different data sets are generated for each specification.

Data generating parameters				$T = N = 10,000$		
a_x	ρ_d	ρ_e	ρ_f	PLS1	PLS2	PLS3
0.1	0	0	0	0.731	0.734	**0.736**
0.1	0	0	0.3	0.731	0.735	**0.738**
0.1	0	0	0.5	0.739	0.744	**0.747**
0.1	0.3	0.3	0	0.729	0.733	**0.734**
0.1	0.3	0.3	0.3	0.729	0.734	**0.736**
0.1	0.3	0.3	0.5	0.738	0.742	**0.745**
0.1	0.5	0.5	0	0.729	0.733	**0.734**
0.1	0.5	0.5	0.3	0.729	0.734	**0.736**
0.1	0.5	0.5	0.5	0.737	0.743	**0.745**
0.2	0	0	0	0.732	0.736	**0.738**
0.2	0	0	0.3	0.732	0.737	**0.74**
0.2	0	0	0.5	0.741	0.746	**0.748**
0.2	0.3	0.3	0	0.731	0.735	**0.737**
0.2	0.3	0.3	0.3	0.731	0.736	**0.739**
0.2	0.3	0.3	0.5	0.74	0.745	**0.747**
0.2	0.5	0.5	0	0.731	0.735	**0.737**
0.2	0.5	0.5	0.3	0.731	0.736	**0.739**
0.2	0.5	0.5	0.5	0.739	0.745	**0.747**
0.3	0	0	0	0.733	0.737	**0.738**
0.3	0	0	0.3	0.733	0.738	**0.74**
0.3	0	0	0.5	0.741	0.746	**0.749**
0.3	0.3	0.3	0	0.732	0.736	**0.738**
0.3	0.3	0.3	0.3	0.732	0.737	**0.739**
0.3	0.3	0.3	0.5	0.74	0.746	**0.748**
0.3	0.5	0.5	0	0.732	0.736	**0.738**
0.3	0.5	0.5	0.3	0.732	0.737	**0.739**
0.3	0.5	0.5	0.5	0.74	0.746	**0.748**

Table A.7: Forecasting by PLS Regressions with Different Numbers of Factors ($R = K = 3, N = T = 10,000$)

Notes: This table reports the forecasting performances of three regressions with three different numbers of informative PLS factors: one (PLS1), two (PLS2), and three (PLS3). For each data specification, the highest out-of-sample R-square is in bold. The other parameters used to generate the data are set at $a_y = 0.7$ and $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 5, 7)$. To save computation time, only 100 different data sets are generated for each specification.

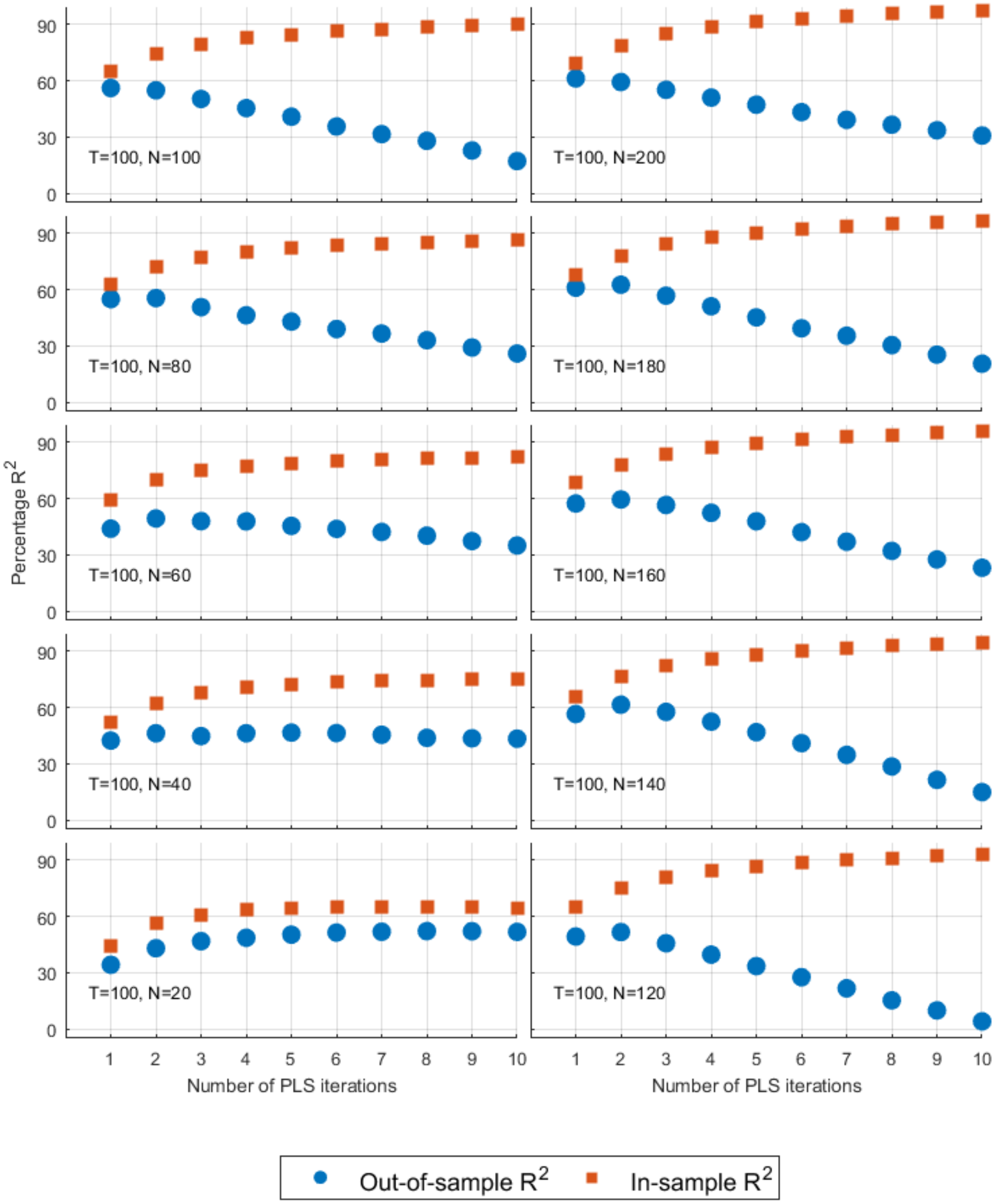

Figure A.2: Performances of PLS Regression and Spurious Correlation ($T = 100$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5)$, $R = 2, K = 4$, $a_x = 0.2, a_y = 0.7$, and $\rho_f = \rho_e = \rho_c = 0.5$.

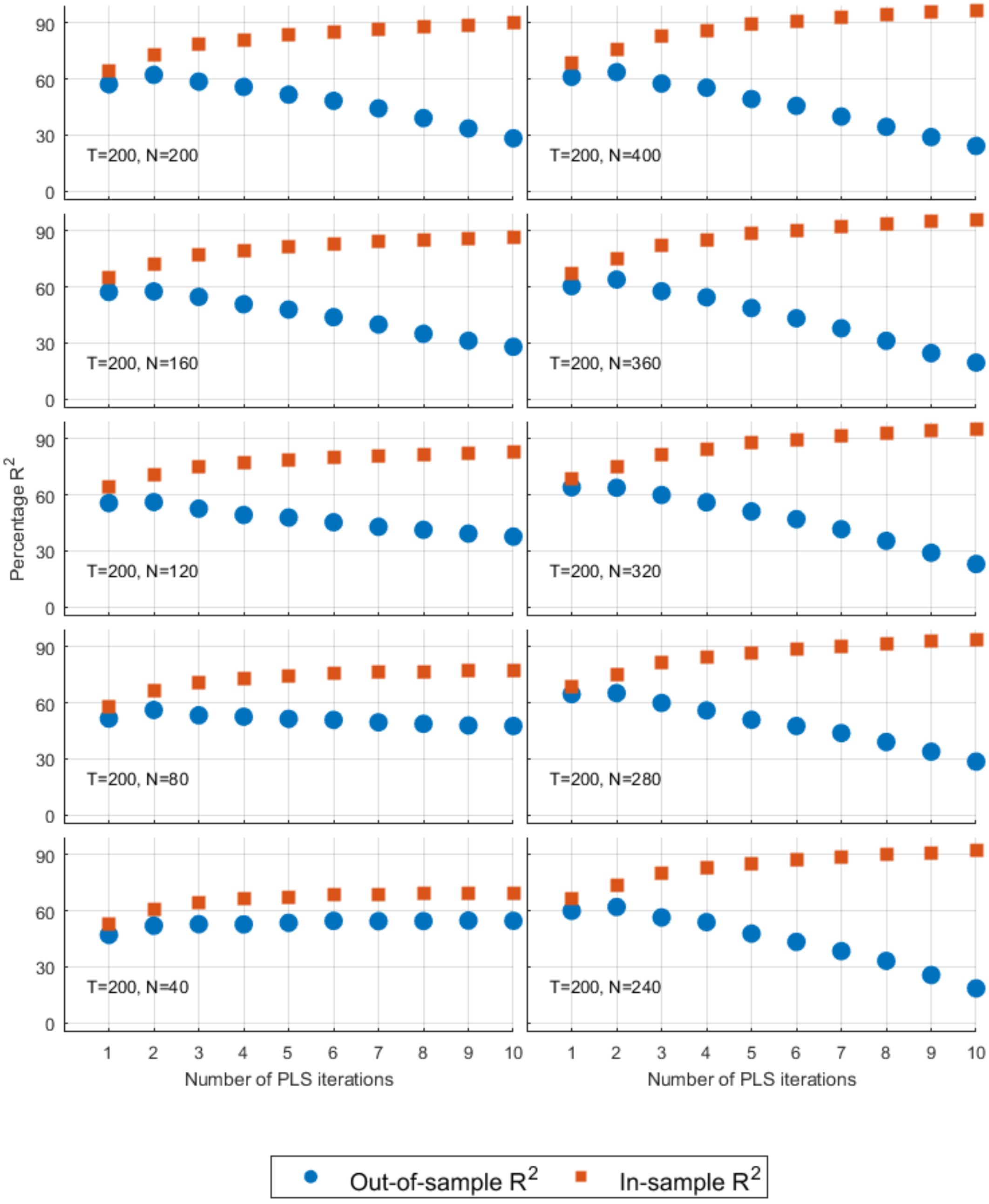

Figure A.3: Performances of PLS Regression and Spurious Correlation ($T = 200$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5)$, $R = 2, K = 4$, $a_x = 0.2, a_y = 0.7$, and $\rho_f = \rho_e = \rho_c = 0.5$.

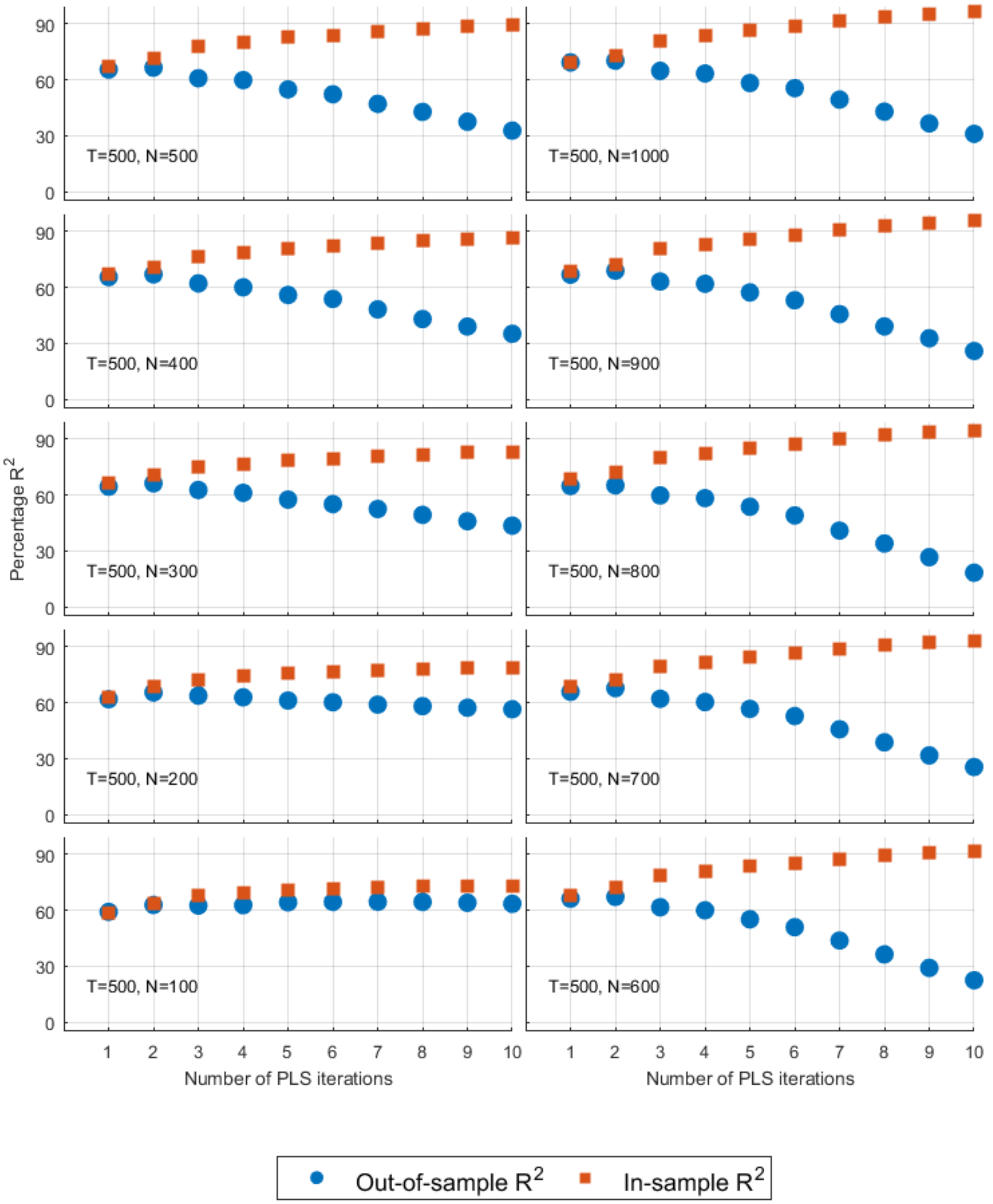

Figure A.4: Performances of PLS Regression and Spurious Correlation ($T = 500$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega_F^*} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4$, $a_x = 0.2, a_y = 0.7$, and $\rho_f = \rho_e = \rho_c = 0.5$.

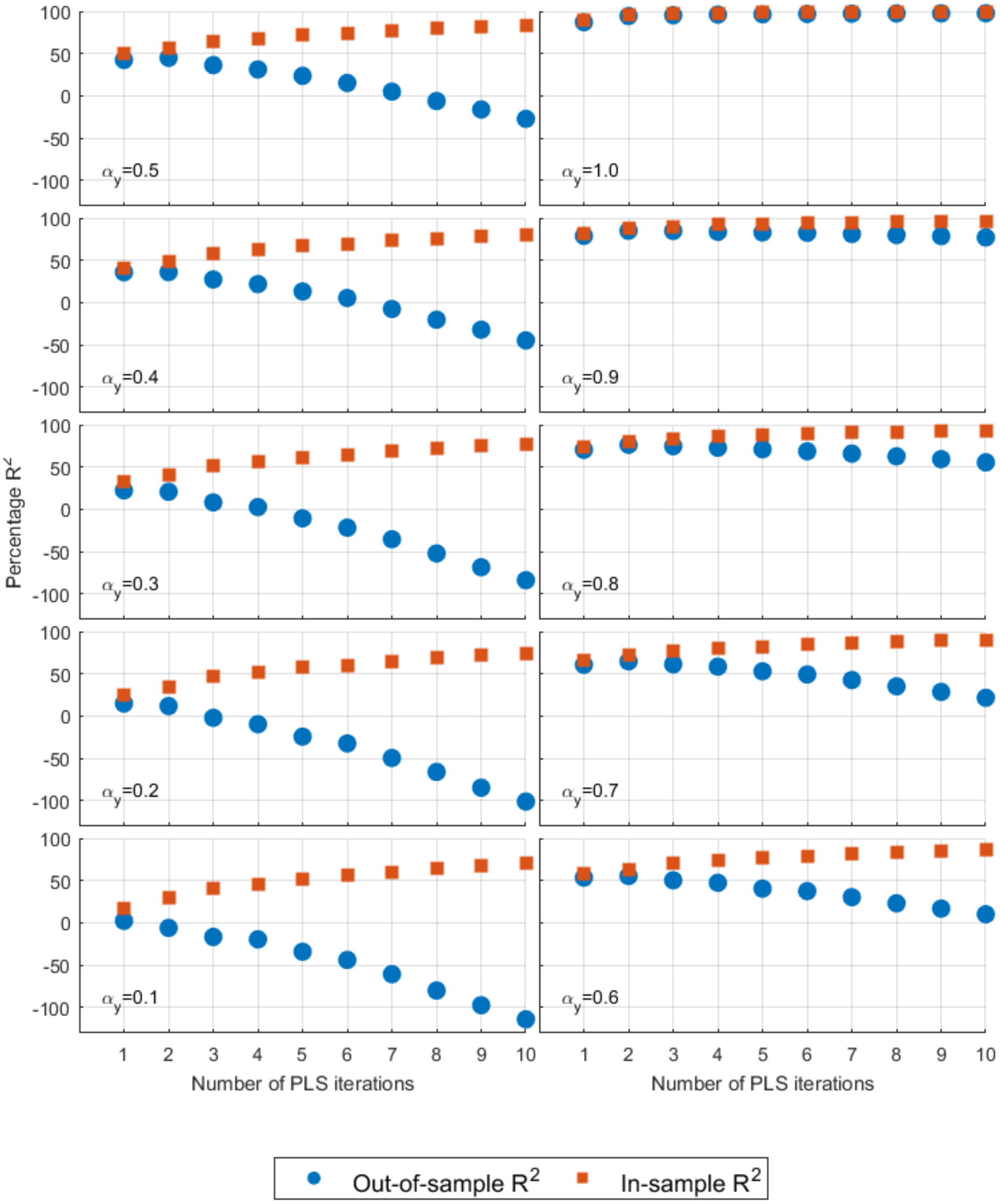

Figure A.5: Performance of PLS Regression and Spurious Correlation ($a_x = 0.2$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3,3,5,5), R = 2, K = 4, a_x = 0.2, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

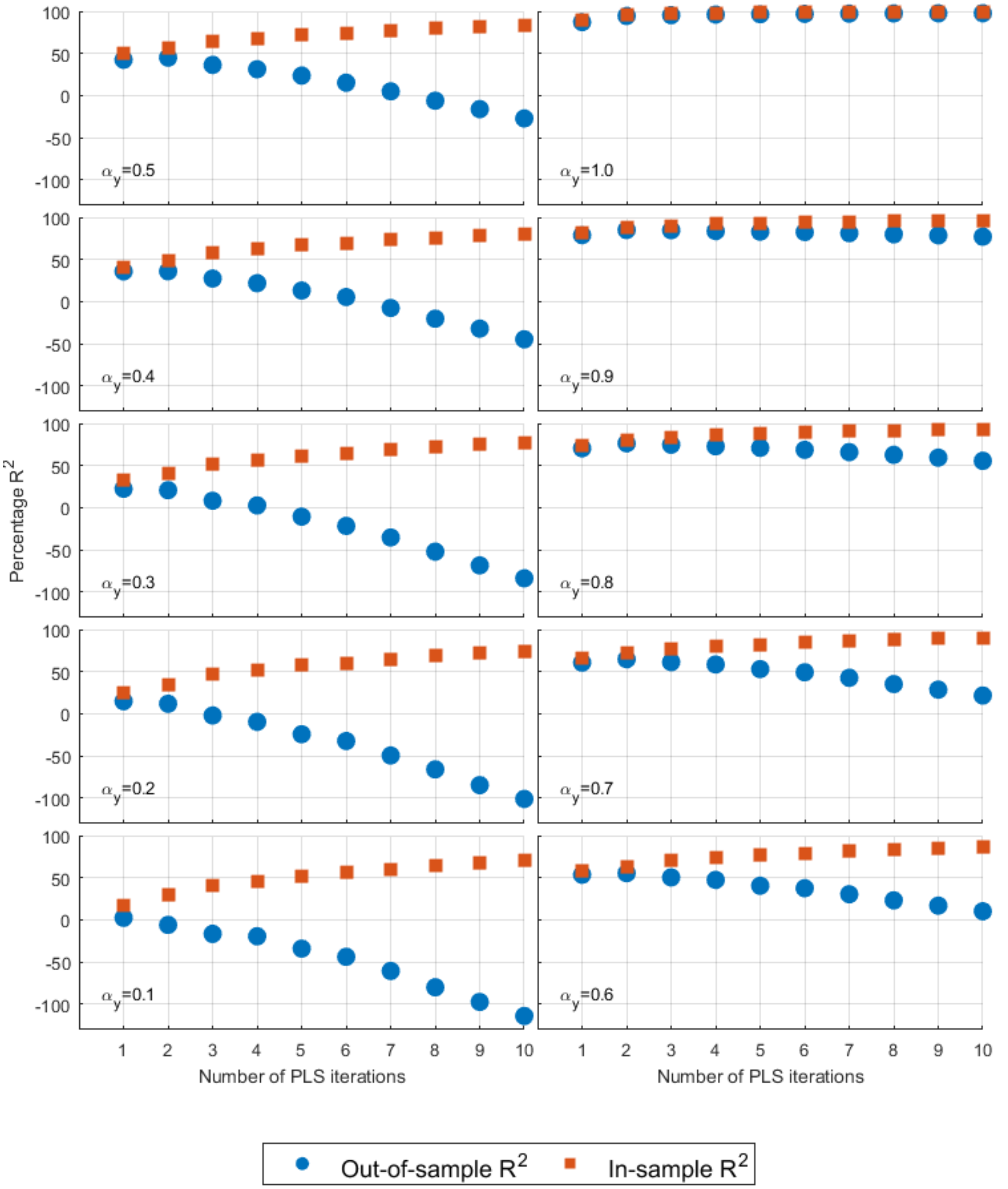

Figure A.6: Performance of PLS Regression and Spurious Correlation ($a_x = 0.5$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega_F^*} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4, a_x = 0.5, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

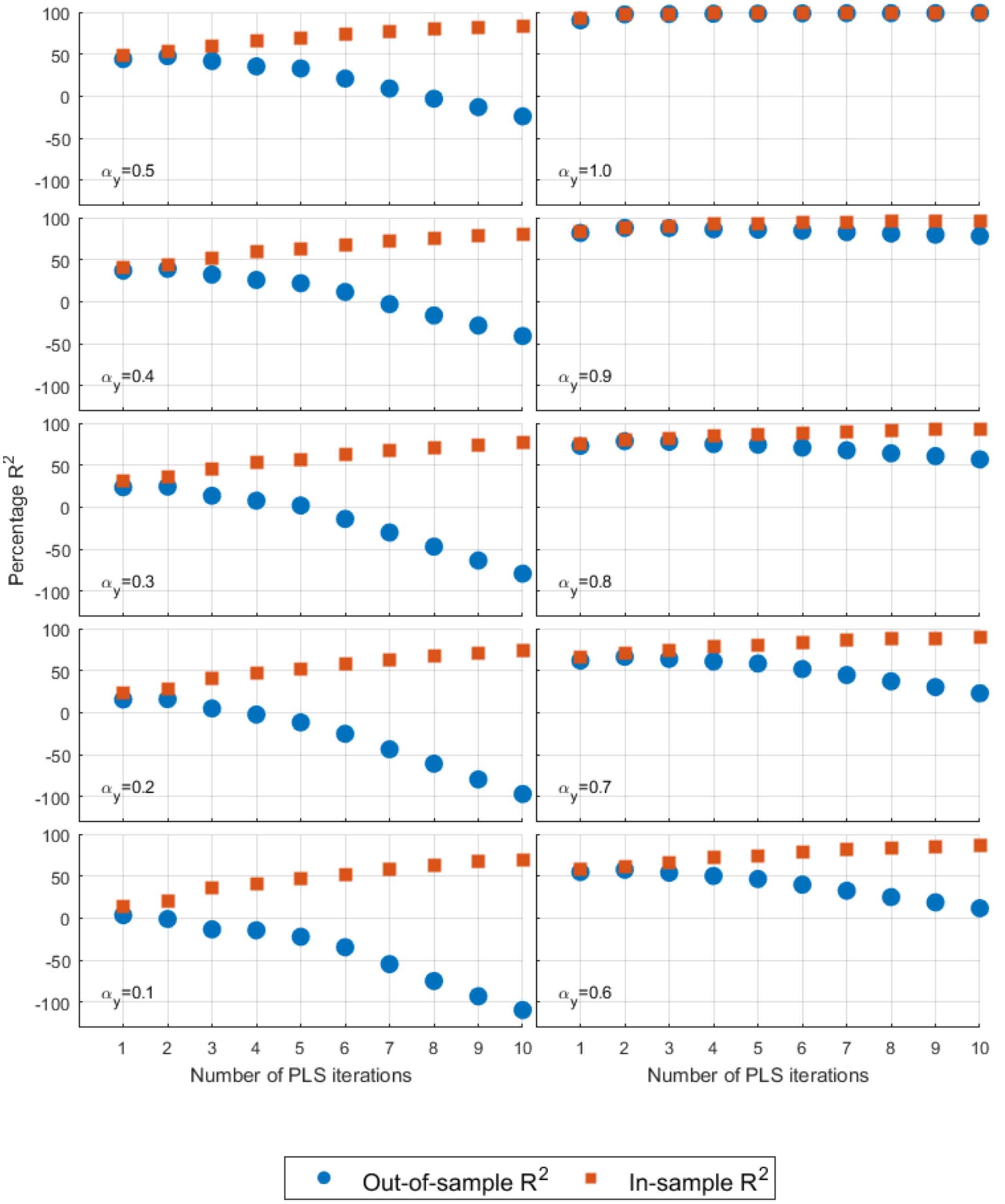

Figure A.7: Performance of PLS Regression and Spurious Correlation ($a_x = 0.7$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4, a_x = 0.7, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

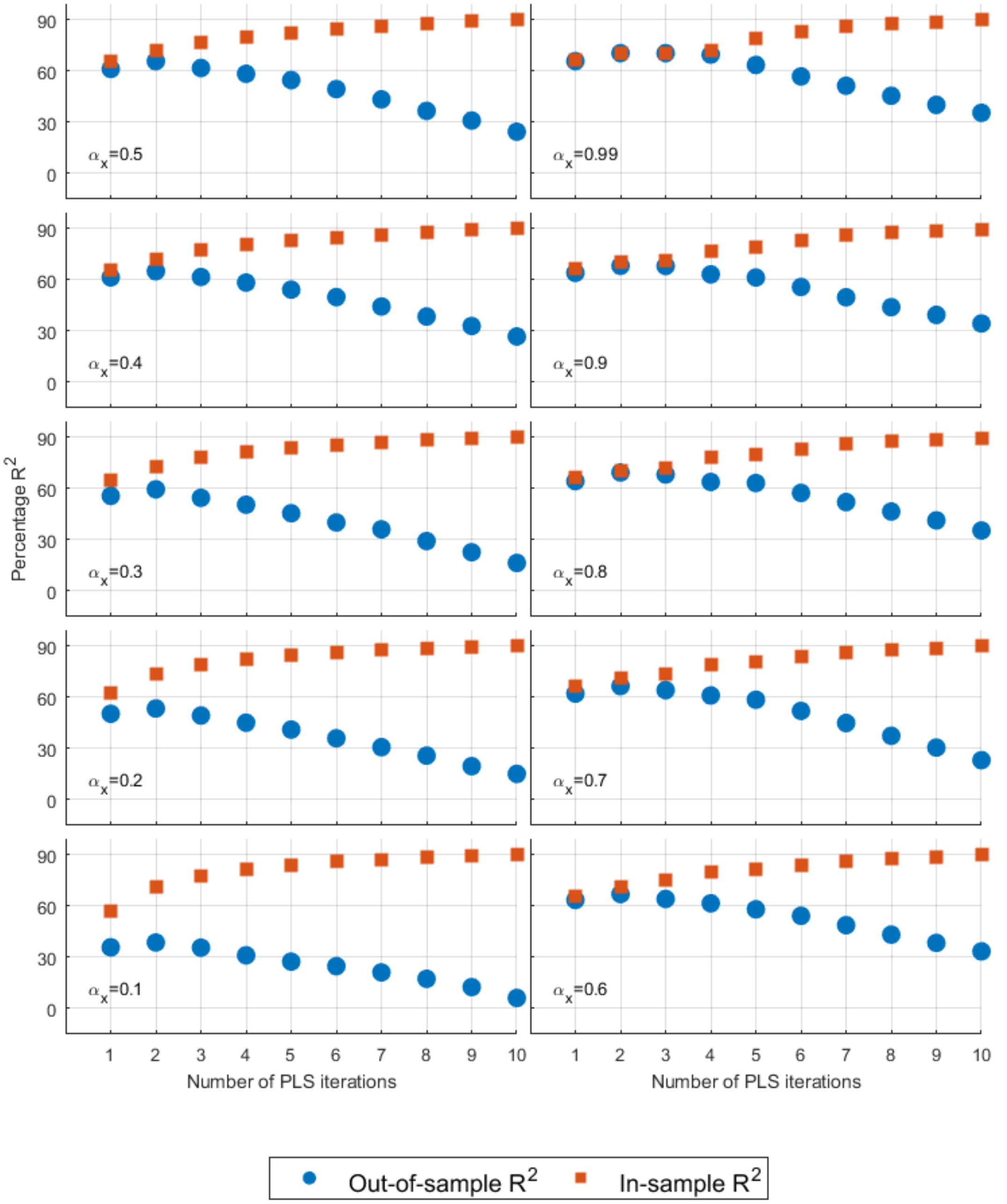

Figure A.8: Performance of PLS Regression and Spurious Correlation ($a_y = 0.7$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4, a_y = 0.7, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

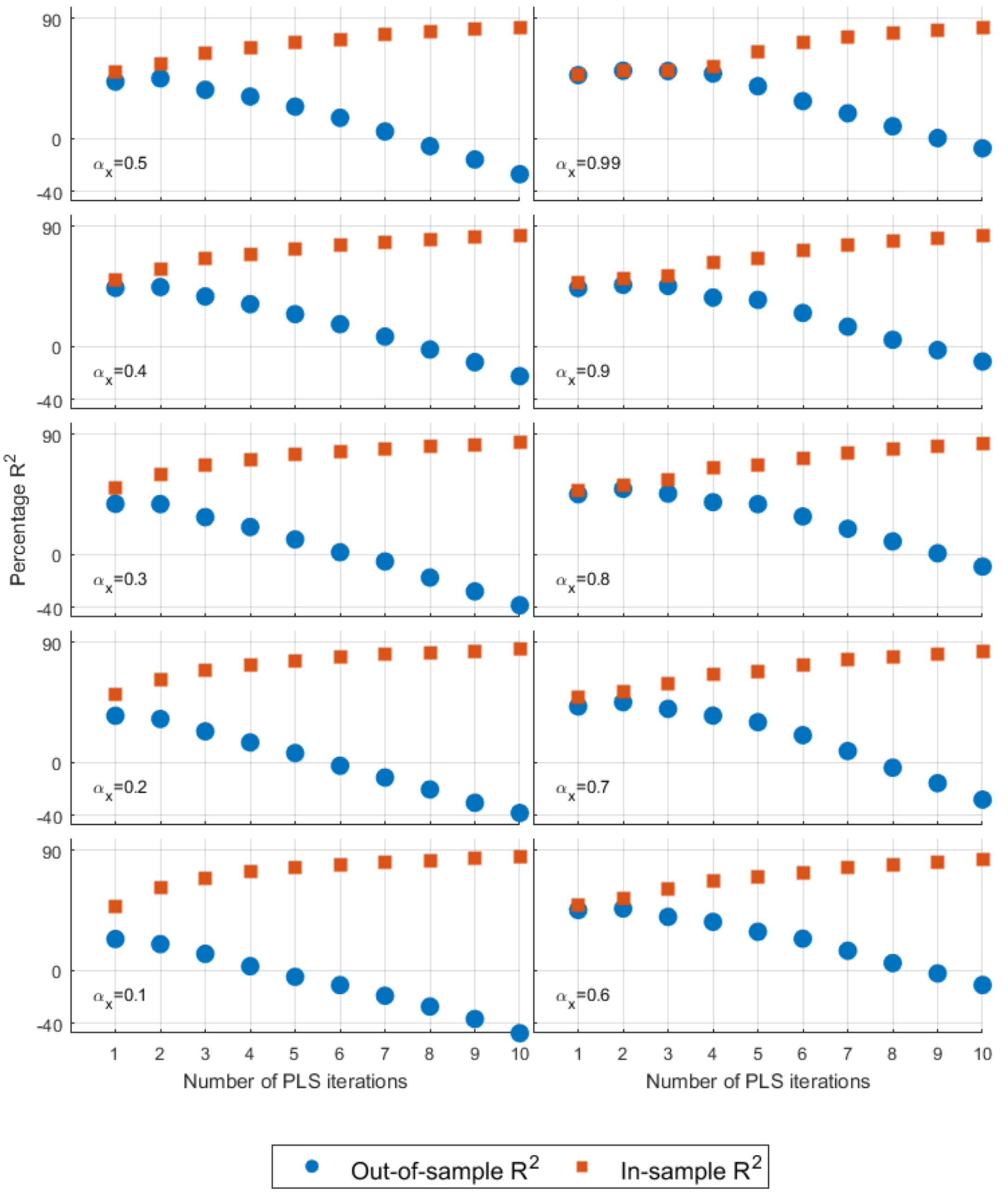

Figure A.9: Performance of PLS Regression and Spurious Correlation ($a_y = 0.5$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4, a_y = 0.5, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

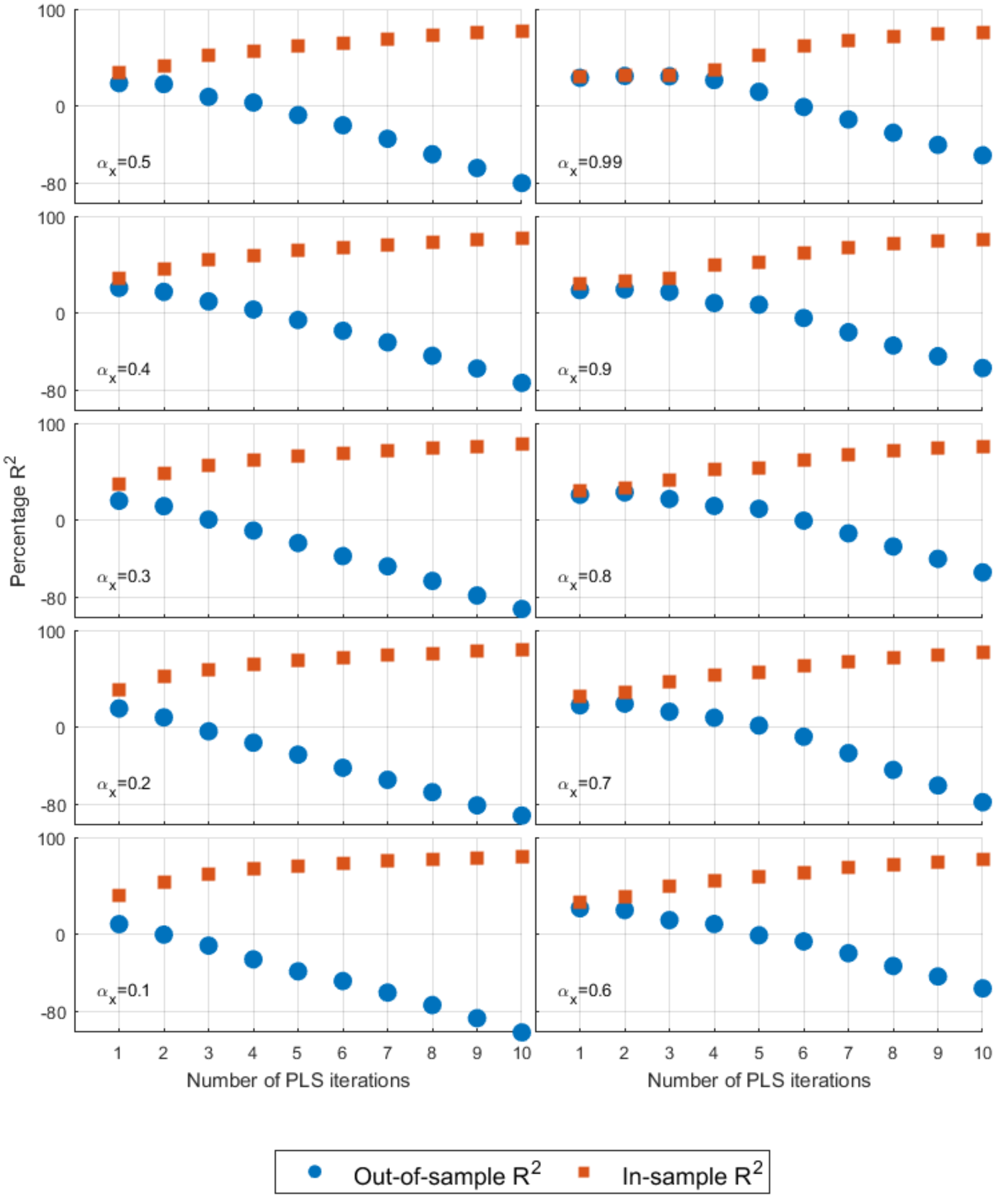

Figure A.10: Performance of PLS Regression and Spurious Correlation ($a_y = 0.3$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4, a_y = 0.3, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

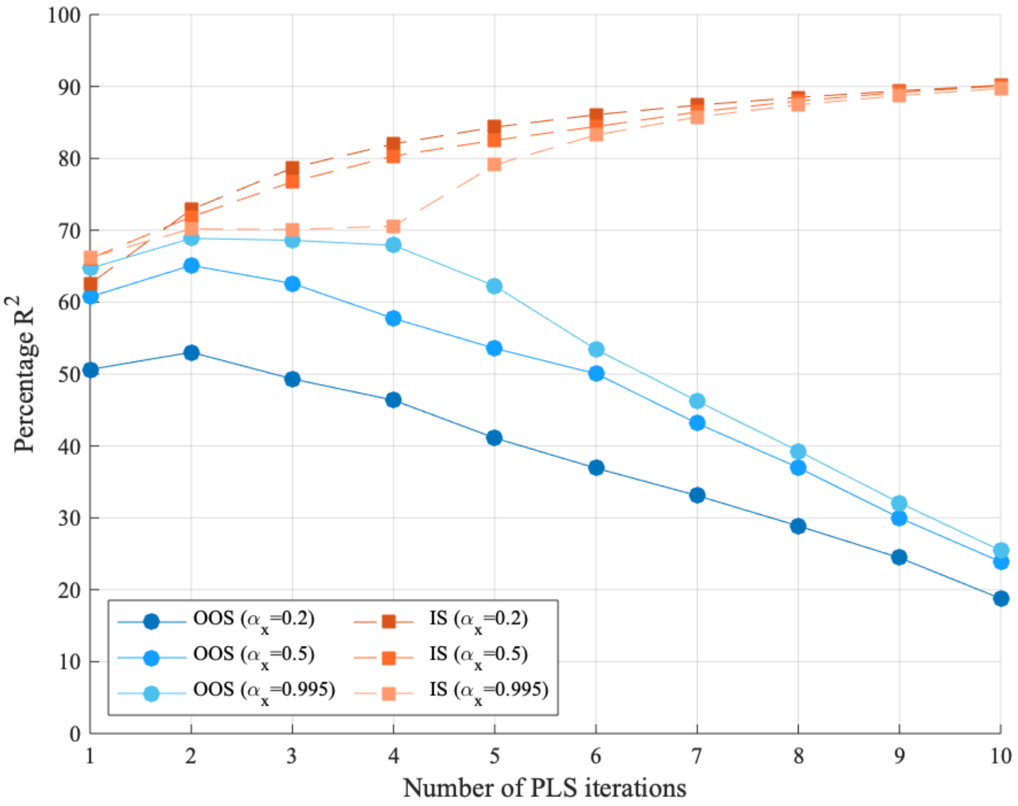

Figure A.11: Forecasting with Uninformative and Spurious Factors ($N = T = 100$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4, a_y = 0.7, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

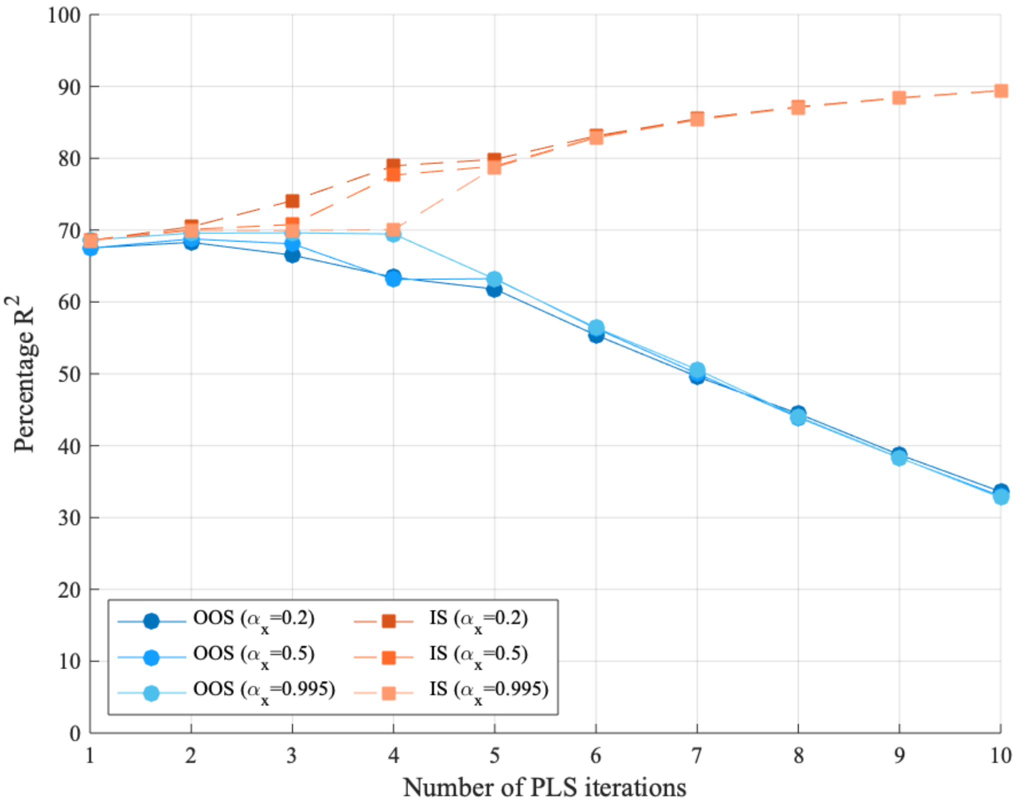

Figure A.12: Forecasting with Uninformative and Spurious Factors ($N = T = 2000$)

Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}^*_{\boldsymbol{F}} = \boldsymbol{diag}(3, 3, 5, 5), R = 2, K = 4, a_y = 0.7, \rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 2,000$.

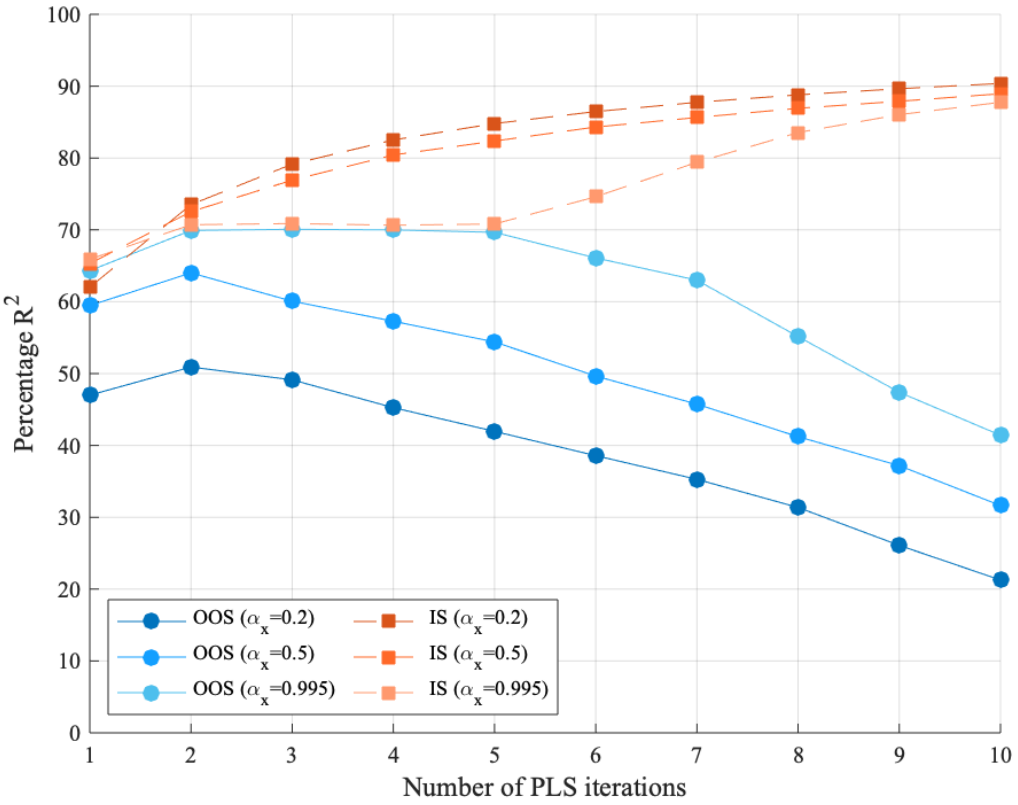

Figure A.13: Forecasting with Uninformative and Spurious Factors ($K = 6, R = 2, N = T = 100$)

Notes: The parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*}= \boldsymbol{diag}(3,3,3,5,5,3)$, $\boldsymbol{\beta}^{*} = (1,0,0,1,0,0)'$, $R=2, K=6, a_y=0.7, \rho_f=\rho_e=\rho_c=0.5$, and $N=T=100$.

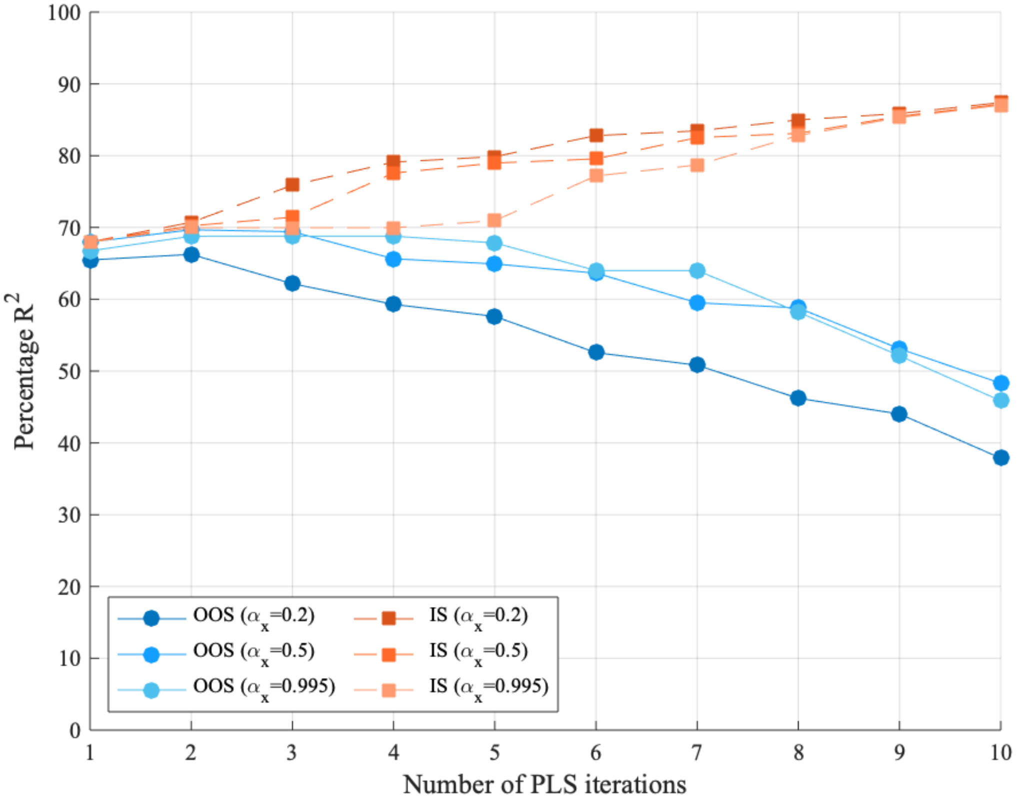

Figure A.14: Forecasting with Uninformative and Spurious Factors ($K = 6, R = 2, N = T = 2000$)

Notes: The parameters for data generating processes are set at $\boldsymbol{\Omega}^*_{\boldsymbol{F}} = \boldsymbol{diag}(3, 3, 3, 5, 5, 3)$, $\boldsymbol{\beta}^* = (1, 0, 0, 1, 0, 0)'$, $R = 2, K = 6$, $a_y = 0.7$, $\rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 2,000$.

Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	ratio	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
100	20	0.2	0.40	0.26	0.32	0.36	0.39	0.42	**0.43**	6.32	3.00
100	60	0.6	0.44	0.47	**0.48**	0.46	0.42	0.39	0.36	2.45	1.74
100	100	1	0.48	0.48	**0.52**	0.49	0.44	0.41	0.37	2.08	1.14
100	160	1.6	0.58	0.57	**0.59**	0.56	0.51	0.45	0.39	1.92	1.16
100	200	2	0.62	0.62	**0.64**	0.59	0.52	0.47	0.42	1.87	1.01
Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	ratio	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
200	40	0.2	0.42	0.32	0.38	0.43	0.45	0.45	**0.45**	6.15	2.76
200	120	0.6	0.56	0.53	**0.58**	0.56	0.55	0.52	0.50	2.11	1.14
200	200	1	0.64	0.62	**0.64**	0.61	0.55	0.51	0.46	1.86	0.67
200	320	1.6	0.64	0.63	**0.65**	0.60	0.55	0.49	0.45	1.67	0.58
200	400	2	0.65	0.62	**0.66**	0.62	0.59	0.55	0.51	1.85	0.56
Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	ratio	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
500	100	0.2	0.60	0.57	**0.61**	0.61	0.61	0.60	0.61	4.42	2.66
500	300	0.6	0.65	0.62	**0.66**	0.63	0.61	0.59	0.56	1.96	0.42
500	500	1	0.68	0.64	**0.69**	0.65	0.61	0.58	0.54	1.93	0.36
500	800	1.6	0.67	0.66	**0.68**	0.64	0.60	0.56	0.51	1.88	0.36
500	1000	2	0.68	0.65	**0.68**	0.65	0.63	0.58	0.54	1.84	0.41

Table A.8: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression Across Different Sample Sizes

Notes: This table reports the forecasting performances of the regressions with different numbers of the PLS factors and the estimated optimal number of PLS factors by the cross-validation method we consider. The data used are simulated using a five-factor model with $\Omega_F^* = \boldsymbol{diag}(3,3,5,5,7)$ and $\boldsymbol{\beta}^* = (1,0,1,0,1)'$. The other data-generating parameters are set at $a_x = 0.2$, $a_y = 0.7$, and $\rho_f = \rho_e = \rho_c = 0.5$.

Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	a_x	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
100	100	0.1	0.37	0.39	**0.40**	0.40	0.37	0.34	0.31	2.34	1.71
100	100	0.3	0.59	0.58	**0.61**	0.59	0.56	0.52	0.49	2.15	1.18
100	100	0.5	0.62	0.61	**0.64**	0.63	0.58	0.54	0.49	2.04	0.98
100	100	0.7	0.66	0.60	**0.67**	0.66	0.62	0.60	0.56	2.47	1.05
100	100	0.9	0.69	0.64	0.70	**0.70**	0.67	0.65	0.64	2.98	1.20
Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	a_x	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
200	200	0.1	0.49	0.47	**0.51**	0.48	0.45	0.42	0.38	1.95	0.78
200	200	0.3	0.66	0.65	**0.66**	0.63	0.61	0.57	0.54	1.81	0.64
200	200	0.5	0.69	0.68	**0.70**	0.67	0.62	0.60	0.55	1.94	0.68
200	200	0.7	0.69	0.66	**0.71**	0.70	0.65	0.63	0.60	2.26	0.80
200	200	0.9	0.69	0.64	0.69	**0.69**	0.67	0.63	0.62	2.87	0.98

Table A.9: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression Across Different a_x

Notes: This table reports the forecasting performances of the regressions with different numbers of the PLS factors and the estimated optimal number of PLS factors by the cross-validation method we consider. The data used are simulated using a five-factor model with $\Omega_F^* = \boldsymbol{diag}(3,3,5,5,7)$ and $\boldsymbol{\beta}^* = (1,0,1,0,1)'$. The other data-generating parameters are set at $a_y = 0.7$, and $\rho_f = \rho_e = \rho_c = 0.5$.

Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	a_y	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
100	100	0.1	0.01	**0.04**	-0.12	-0.27	-0.41	-0.55	-0.69	1.16	0.67
100	100	0.3	0.20	**0.23**	0.16	0.06	-0.05	-0.16	-0.26	1.25	0.63
100	100	0.5	0.32	**0.36**	0.31	0.22	0.12	0.03	-0.05	1.43	0.85
100	100	0.7	0.46	0.47	**0.50**	0.48	0.43	0.38	0.34	2.24	1.20
100	100	0.9	0.70	0.64	0.73	**0.73**	0.72	0.72	0.71	4.18	2.15
Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	a_y	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
200	200	0.1	0.02	**0.02**	-0.11	-0.25	-0.39	-0.50	-0.63	1.07	0.27
200	200	0.3	0.25	**0.26**	0.17	0.09	0.00	-0.10	-0.23	1.11	0.33
200	200	0.5	**0.44**	0.43	0.40	0.31	0.21	0.14	0.05	1.29	0.49
200	200	0.7	0.62	0.61	**0.63**	0.58	0.53	0.49	0.44	1.65	0.65
200	200	0.9	0.80	0.75	**0.81**	0.80	0.79	0.79	0.78	3.06	1.20

Table A.10: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression Across Different a_y

Notes: This table reports the forecasting performances of the regressions with different numbers of the PLS factors and the estimated optimal number of PLS factors by the cross-validation method we consider. The data used are simulated using a five-factor model with $\Omega_F^* = \boldsymbol{diag}(3, 3, 5, 5, 7)$ and $\boldsymbol{\beta}^* = (1, 0, 1, 0, 1)'$. The other data-generating parameters are set at $a_x = 0.2$, and $\rho_f = \rho_e = \rho_c = 0.5$.

Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	ρ_{eu}	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
100	100	0	0.51	0.51	**0.53**	0.50	0.45	0.43	0.39	2.18	1.23
100	100	0.1	0.57	0.58	**0.59**	0.56	0.51	0.48	0.43	1.99	1.09
100	100	0.2	0.55	0.55	**0.57**	0.53	0.48	0.43	0.38	2.33	1.28
100	100	0.3	0.55	0.53	**0.56**	0.54	0.51	0.47	0.42	2.25	1.33
100	100	0.4	0.61	0.57	**0.63**	0.60	0.57	0.53	0.49	2.54	1.42
100	100	0.5	0.57	0.54	**0.57**	0.57	0.57	0.56	0.55	2.72	1.68
100	100	0.6	0.58	0.55	**0.61**	0.59	0.57	0.55	0.53	2.96	1.71
100	100	0.7	0.55	0.52	0.55	0.55	**0.56**	0.54	0.54	3.70	2.37
100	100	0.8	0.60	0.56	0.60	0.62	0.61	0.62	**0.62**	4.74	2.54
100	100	0.9	0.60	0.53	0.58	0.58	0.60	0.60	**0.61**	5.71	3.25
100	100	1	**0.88**	0.60	0.68	0.71	0.75	0.78	0.81	9.83	0.73

Table A.11: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression When Some Predictor Has Direct Forecasting Power ($N = T = 100$)

Notes: This table reports the forecasting performances of the regressions with different numbers of the PLS factors and the estimated optimal number of PLS factors by the cross-validation method we consider. ($N = T = 100$) The first predictor's idiosyncratic component is correlated with the error term of the target variable: $e^*_{1t} = \rho_{eu}^{1/2} u^*_{t+1} + (1 - \rho_{eu})^{1/2} v^*_{1t}$, where the v_{1t} are random draws from $N(0, 1)$. When $\rho_{eu} = 1$, the idiosyncratic component of x_{1t}, e^*_{1t} is perfectly correlated with the error term of the target variable. Other than this, the data generating processes used for this table are identical to those which ae used for Table A.8.

Sample specification			Cross-validation	Forecasting with each PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	ρ_{eu}	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	$ave(\hat{R}_{CV})$	$st(\hat{R}_{CV})$
200	200	0	0.61	0.60	**0.61**	0.59	0.55	0.52	0.50	1.79	0.64
200	200	0.1	0.65	0.64	**0.66**	0.62	0.59	0.56	0.53	1.79	0.73
200	200	0.2	0.60	0.58	**0.62**	0.58	0.56	0.51	0.47	2.05	0.82
200	200	0.3	0.62	0.62	**0.64**	0.61	0.58	0.54	0.50	1.99	0.79
200	200	0.4	0.62	0.62	**0.64**	0.62	0.60	0.57	0.55	2.23	0.93
200	200	0.5	0.63	0.61	**0.64**	0.62	0.60	0.58	0.57	2.23	1.16
200	200	0.6	0.63	0.57	**0.64**	0.64	0.63	0.63	0.62	3.07	1.60
200	200	0.7	0.58	0.56	0.60	0.61	0.60	**0.61**	0.60	3.32	2.02
200	200	0.8	0.66	0.61	0.65	0.65	0.66	0.68	**0.68**	5.04	2.69
200	200	0.9	**0.72**	0.54	0.64	0.65	0.66	0.67	0.69	7.87	2.77
200	200	1	**0.85**	0.62	0.68	0.69	0.70	0.73	0.76	9.98	0.28

Table A.12: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression When Some Predictor Has Direct Forecasting Power ($N = T = 200$)

Notes: This table reports the forecasting performances of the regressions with different numbers of the PLS factors and the estimated optimal number of PLS factors by the cross-validation method we consider. ($N = T = 200$) The first predictor's idiosyncratic component is correlated with the error term of the target variable: $e^*_{1t} = \rho_{eu}^{1/2} u^*_{t+1} + (1-\rho_{eu})^{1/2} v^*_{1t}$, where the v_{1t} are random draws from $N(0,1)$. When $\rho_{eu} = 1$, the idiosyncratic component of x_{1t}, e^*_{1t} is perfectly correlated with the error term of the target variable. Other than this, the data generating processes used for this table are identical to those which ae used for Table A.8.

Variables	PLS1	PLS2	PLS3	PLS4	PC1	PC2	PC3	PC4	PLS BIC	PLS CV	PC BIC	PC AH
Industrial Production	**34.6**	30.6	7.5	-56.7	7.7	22	24.9	29.1	-520.2	33.8	32.3	27.9
Personal Income	**34.5**	21.9	-14.4	-96.5	11.2	13.8	9.6	10.4	-319.1	30.4	13.7	16.8
Mfg & Trade Sales	**30.8**	26	-4.6	-54.5	2.7	30.6	26.9	29.1	-559.1	29.6	26	23.7
Nonag. Employment	46.1	40.2	-0.5	-80.4	38.3	45.7	43.7	43.4	-403	49.9	**51.3**	46.2
CPI	**60.7**	60.1	58.3	60.4	59.2	58.6	56.3	54.9	48.8	54.6	55.4	58.4
Consumption Deflator	50.3	48.1	46	47.8	**51.4**	49.1	45.7	43.2	26.5	45.3	41.6	48.9
CPI exc. Food	**56.9**	54.2	52.9	54.5	54.6	52.6	49.6	48.9	42.7	49.8	48.3	51.9
Producer Price Index	65.9	66.2	63.7	**66.4**	65.3	65.9	65.4	64.9	59.7	65.1	65.1	65.4

Table A.13: Forecasting Results for Eight Major Macroeconomic Variables

Notes: The eight target variables are forecasted. The method that gives the highest out-of-sample R^2 is in bold.

Category	PLS1	PLS2	PLS3	PLS4	PC1	PC2	PC3	PC4	PLS BIC	PLS CV	PC BIC	PC AH
Overall	**35**	34.3	15.6	0.5	24.1	33.8	32.4	33	-79.1	30.3	34.2	28.2
Output and Income	**34.1**	32.7	9.5	-52.5	4.3	16.1	16.5	17.3	-433.6	30.6	29.9	19.4
Labor Market	39.2	41.8	20.4	15.4	28	41.5	39.9	38.7	-82.6	40.2	**43.5**	38.2
Housing	45.5	46.8	27.4	44	51.6	52	52.4	53.2	35.5	33.6	52	**54.6**
Consumption	**13.6**	2.8	-46	-126.8	-0.4	11.6	10.4	10.6	-598.9	9.5	11.9	6.8
Money and Credit	44.6	47.6	42.3	42.4	39.9	**49.2**	48.1	46.7	20.3	27	42	44.3
Interest and Exchange Rates	11	-1.8	-16.6	-27.5	11.6	11.1	11	8.9	-122.8	5.3	-0.9	**12.1**
Prices	**60**	57.2	55.6	57.4	59.2	58.9	56.6	54.8	44.2	53.9	55.3	58.4
Stock Market	6.8	-1.6	-23.4	-26.7	**8**	3.2	2.1	-0.8	-150.7	-2.6	1	1.5

Table A.14: Forecasting Results for 144 Macroeconomic Variables

Notes: The whole 144 target variables are forecasted. The entries are median out-of-sample R^2 of each category. The method that gives the highest out-of-sample R^2 is in bold. The category Consumption includes consumption, orders, and inventory variables.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

Industrial Production

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.893	**0.638**	0.777	0.731	0.727	**0.722**	0.833	0.775	0.898	0.902	0.877
$k = 2$	0.714	0.682	0.786	0.768	0.724	0.76	**0.708**	0.777	0.743	0.806	0.767
$k = 3$	**0.699**	0.838	0.759	0.726	**0.699**	0.775	0.749	**0.771**	**0.732**	**0.799**	0.765
$k = 4$	0.719	0.935	**0.717**	**0.719**	0.743	0.804	0.786	0.802	0.76	0.82	0.786
$k = 5$	0.747	1.018	0.733	0.741	0.788	0.767	0.741	0.802	0.782	0.845	**0.741**
$k = 6$	0.729	1.129	0.756	0.743	0.76	0.741	0.733	0.803	0.785	0.803	0.752
$k = 7$	0.725	1.188	0.762	0.765	0.775	0.75	0.738	0.81	0.779	0.81	0.786
$k = 8$	0.736	1.326	0.759	0.765	0.767	0.748	0.735	0.809	0.791	0.821	0.792
$k = 9$	0.732	1.467	0.775	0.774	0.786	0.756	0.732	0.817	0.788	0.843	0.772
$k = 10$	0.738	1.592	0.785	0.784	0.792	0.793	0.737	0.808	0.8	0.891	0.767
$k = 11$	0.75	1.718	0.789	0.794	0.789	0.801	0.751	0.82	0.815	0.893	0.783
$k = 12$	0.737	1.818	0.792	0.804	0.794	0.79	0.745	0.822	0.803	0.911	0.781

Real Personal Income

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.854	**0.68**	0.848	**0.776**	**0.778**	**0.768**	0.833	**0.925**	0.861	**0.875**	**0.838**
$k = 2$	**0.842**	0.787	0.865	0.882	0.882	0.863	**0.825**	0.956	**0.856**	0.892	0.868
$k = 3$	0.846	0.842	**0.834**	0.818	0.852	0.895	0.881	0.965	0.86	0.896	0.873
$k = 4$	0.869	1.065	0.846	0.85	0.864	0.895	0.885	0.977	0.887	0.925	0.884
$k = 5$	0.874	1.171	0.857	0.862	0.874	0.874	0.866	0.983	0.895	0.946	0.874
$k = 6$	0.862	1.278	0.878	0.874	0.892	0.877	0.864	0.979	0.889	0.934	0.896
$k = 7$	0.874	1.385	0.894	0.869	0.903	0.862	0.859	0.979	0.896	0.941	0.903
$k = 8$	0.869	1.475	0.897	0.909	0.911	0.86	0.865	0.977	0.883	0.921	0.908
$k = 9$	0.866	1.572	0.939	0.916	0.892	0.866	0.865	0.975	0.882	0.903	0.896
$k = 10$	0.884	1.668	0.915	0.938	0.899	0.898	0.881	1.008	0.909	0.925	0.905
$k = 11$	0.888	1.765	0.916	0.922	0.884	0.914	0.892	1.029	0.921	0.942	0.912
$k = 12$	0.888	1.834	0.934	0.953	0.885	0.909	0.889	1.02	0.917	0.96	0.915

Table B.1: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Given k, $k = 1, 2, ..., 12$: Real Variables

Real Manufacturing & Trade Industries Sales

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.95	**0.648**	0.767	**0.732**	0.719	**0.674**	0.917	0.761	0.95	0.948	0.927
$k = 2$	0.681	0.722	**0.739**	0.744	0.746	0.703	**0.663**	0.736	0.688	0.738	0.711
$k = 3$	**0.665**	0.863	0.74	0.733	**0.706**	0.714	0.694	**0.733**	**0.674**	**0.726**	0.718
$k = 4$	0.678	0.985	0.75	0.752	0.733	0.736	0.699	0.782	0.693	0.743	0.715
$k = 5$	0.676	1.116	0.797	0.797	0.769	0.697	0.682	0.777	0.7	0.772	**0.691**
$k = 6$	0.683	1.282	0.784	0.766	0.791	0.693	0.686	0.777	0.701	0.779	0.721
$k = 7$	0.698	1.421	0.772	0.779	0.756	0.688	0.675	0.785	0.708	0.771	0.745
$k = 8$	0.71	1.606	0.783	0.797	0.742	0.688	0.69	0.783	0.727	0.794	0.753
$k = 9$	0.707	1.792	0.796	0.787	0.75	0.706	0.692	0.796	0.729	0.795	0.735
$k = 10$	0.715	1.946	0.791	0.777	0.75	0.754	0.707	0.807	0.745	0.834	0.741
$k = 11$	0.717	2.088	0.799	0.771	0.743	0.753	0.722	0.816	0.759	0.848	0.755
$k = 12$	0.718	2.191	0.806	0.78	0.763	0.74	0.716	0.809	0.754	0.876	0.758

Nonagriculture Employment

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.586	**0.498**	0.538	**0.475**	0.543	**0.498**	0.571	0.539	0.592	0.593	0.591
$k = 2$	0.515	0.528	0.538	0.534	0.529	0.531	0.519	**0.538**	0.521	0.553	0.542
$k = 3$	**0.503**	0.673	**0.482**	0.476	0.498	0.528	0.524	0.539	**0.511**	**0.544**	0.562
$k = 4$	0.518	0.871	0.513	0.51	0.531	0.552	0.543	0.549	0.525	0.546	0.555
$k = 5$	0.518	1.031	0.504	0.507	0.509	0.51	**0.501**	0.547	0.53	0.555	0.519
$k = 6$	0.504	1.096	0.504	0.492	0.504	0.515	0.505	0.546	0.519	0.557	0.532
$k = 7$	0.516	1.194	0.506	0.506	0.502	0.51	0.507	0.549	0.528	0.565	0.56
$k = 8$	0.528	1.324	0.508	0.509	0.491	0.521	0.516	0.547	0.535	0.567	0.565
$k = 9$	0.525	1.439	0.506	0.499	0.493	0.53	0.527	0.555	0.537	0.602	0.535
$k = 10$	0.513	1.545	0.527	0.505	0.486	0.53	0.527	0.558	0.536	0.625	**0.512**
$k = 11$	0.521	1.648	0.517	0.5	**0.473**	0.54	0.516	0.565	0.541	0.655	0.523
$k = 12$	0.511	1.74	0.513	0.505	0.492	0.534	0.521	0.568	0.521	0.646	0.535

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the four real target variables (Industrial Production, Real Personal Income, Real Manufacturing & Trade Industries Sales, Nonagriculture Employment) by all factor estimations methods, with given k, $k = 1, 2, ..., 12$, is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

CPI

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.405	**0.392**	0.483	0.441	0.425	0.419	0.404	0.448	0.405	0.405	0.407
$k = 2$	0.41	0.397	0.466	0.435	0.402	0.406	0.408	**0.439**	0.413	0.413	0.413
$k = 3$	0.405	0.412	**0.406**	**0.4**	**0.398**	**0.385**	0.4	0.441	0.408	0.409	**0.398**
$k = 4$	**0.395**	0.409	0.409	0.407	0.4	0.402	**0.398**	0.446	**0.399**	0.402	0.4
$k = 5$	0.396	0.421	0.407	0.405	0.401	0.403	0.4	0.45	0.399	**0.398**	0.409
$k = 6$	0.405	0.412	0.408	0.405	0.409	0.405	0.406	0.45	0.408	0.407	0.408
$k = 7$	0.409	0.418	0.417	0.414	0.411	0.4	0.405	0.453	0.413	0.415	0.415
$k = 8$	0.408	0.42	0.416	0.416	0.408	0.403	0.406	0.454	0.412	0.415	0.416
$k = 9$	0.407	0.429	0.421	0.419	0.406	0.404	0.407	0.457	0.41	0.416	0.413
$k = 10$	0.42	0.442	0.411	0.425	0.408	0.416	0.411	0.458	0.424	0.452	0.419
$k = 11$	0.423	0.458	0.416	0.43	0.408	0.432	0.42	0.459	0.428	0.45	0.434
$k = 12$	0.424	0.48	0.429	0.431	0.408	0.43	0.422	0.464	0.43	0.457	0.435

Consumption Deflator

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	**0.483**	**0.492**	0.534	0.543	0.523	0.495	**0.48**	0.518	**0.48**	**0.482**	0.484
$k = 2$	0.496	0.502	**0.49**	0.516	0.503	0.492	0.494	**0.511**	0.496	0.497	0.499
$k = 3$	0.496	0.504	0.496	**0.507**	**0.499**	**0.475**	0.49	0.516	0.5	0.498	**0.482**
$k = 4$	0.485	0.507	0.508	0.511	0.503	0.502	0.492	0.528	0.489	0.488	0.492
$k = 5$	0.49	0.527	0.514	0.517	0.511	0.507	0.5	0.531	0.492	0.489	0.509
$k = 6$	0.511	0.523	0.511	0.515	0.518	0.515	0.515	0.534	0.51	0.505	0.52
$k = 7$	0.516	0.551	0.516	0.527	0.534	0.51	0.513	0.539	0.515	0.513	0.522
$k = 8$	0.518	0.566	0.517	0.532	0.531	0.513	0.516	0.54	0.52	0.518	0.527
$k = 9$	0.512	0.573	0.523	0.529	0.528	0.509	0.512	0.549	0.512	0.515	0.517
$k = 10$	0.529	0.594	0.528	0.531	0.53	0.526	0.521	0.547	0.529	0.55	0.527
$k = 11$	0.532	0.609	0.533	0.531	0.52	0.543	0.529	0.546	0.533	0.555	0.544
$k = 12$	0.534	0.63	0.54	0.527	0.512	0.542	0.53	0.553	0.535	0.566	0.544

Table B.2: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Given k, $k = 1, 2, ..., 12$: Nominal Variables

CPI excluding Food

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	**0.453**	**0.427**	0.452	0.49	**0.449**	0.476	**0.455**	0.497	**0.453**	**0.453**	**0.455**
$k = 2$	0.463	0.441	**0.417**	0.483	0.475	0.464	0.459	0.495	0.463	0.461	0.469
$k = 3$	0.464	0.465	0.464	0.466	0.473	**0.452**	0.466	0.494	0.465	0.463	0.457
$k = 4$	0.459	0.466	0.48	0.47	0.477	0.471	0.463	**0.493**	0.46	0.459	0.461
$k = 5$	0.464	0.477	0.48	0.467	0.483	0.473	0.474	0.498	0.465	0.458	0.474
$k = 6$	0.48	0.474	0.484	0.464	0.476	0.475	0.48	0.498	0.48	0.47	0.479
$k = 7$	0.479	0.478	0.478	0.47	0.471	0.469	0.478	0.5	0.479	0.472	0.479
$k = 8$	0.478	0.475	0.482	0.467	0.472	0.474	0.476	0.502	0.479	0.474	0.481
$k = 9$	0.479	0.484	0.484	0.469	0.479	0.477	0.479	0.509	0.479	0.487	0.479
$k = 10$	0.488	0.499	0.485	0.462	0.477	0.482	0.482	0.51	0.488	0.513	0.485
$k = 11$	0.489	0.513	0.493	**0.458**	0.472	0.494	0.488	0.506	0.488	0.51	0.494
$k = 12$	0.488	0.535	0.5	0.462	0.479	0.495	0.49	0.509	0.488	0.512	0.495

Producer Price Index

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.342	0.342	0.365	0.359	0.364	0.34	0.339	0.339	0.342	0.342	0.34
$k = 2$	0.338	0.337	0.358	0.357	0.355	0.336	0.341	**0.335**	0.34	0.342	0.34
$k = 3$	0.336	0.353	**0.35**	0.345	**0.345**	**0.333**	**0.335**	0.335	0.339	0.341	**0.336**
$k = 4$	**0.333**	0.34	0.354	0.347	0.346	0.336	0.336	0.336	**0.337**	**0.339**	0.342
$k = 5$	0.335	0.34	0.351	0.356	0.353	0.335	0.335	0.336	0.338	0.339	0.337
$k = 6$	0.337	0.335	0.352	0.357	0.356	0.338	0.338	0.337	0.34	0.343	0.339
$k = 7$	0.341	**0.334**	0.351	0.358	0.35	0.34	0.339	0.34	0.346	0.35	0.346
$k = 8$	0.343	0.34	0.352	0.359	0.349	0.342	0.343	0.34	0.349	0.352	0.349
$k = 9$	0.339	0.352	0.35	0.356	0.35	0.338	0.338	0.342	0.343	0.347	0.344
$k = 10$	0.343	0.36	0.352	0.352	0.35	0.345	0.339	0.341	0.348	0.371	0.346
$k = 11$	0.347	0.361	0.351	0.348	0.347	0.356	0.343	0.342	0.353	0.372	0.357
$k = 12$	0.348	0.371	0.353	**0.34**	0.35	0.353	0.346	0.344	0.354	0.369	0.357

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the four nominal target variables (CPI, Consumption Deflator, CPI excluding Food, Producer Price Index) by all factor estimations methods, with given k, $k = 1, 2, ..., 12$, is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

1. Output and Income

	PCA	PLS	Targeted Predictors			Weighted PC			Two Step	QMLE	One-sided
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.928	**0.685**	0.835	0.78	**0.76**	**0.755**	0.875	0.851	0.931	0.934	0.913
$k = 2$	0.793	0.704	0.836	0.824	0.803	0.822	0.767	**0.837**	0.808	0.854	0.82
$k = 3$	0.783	0.814	0.776	0.777	0.769	0.834	0.813	0.841	**0.798**	0.847	0.837
$k = 4$	0.798	0.942	0.769	**0.753**	0.772	0.844	0.83	0.873	0.82	0.867	0.84
$k = 5$	0.793	1.034	**0.744**	0.761	0.779	0.805	0.79	0.877	0.819	0.876	0.801
$k = 6$	0.779	1.127	0.759	0.753	0.76	0.803	0.782	0.876	0.807	0.861	0.82
$k = 7$	0.798	1.173	0.762	0.773	0.771	0.793	0.779	0.876	0.816	0.86	0.833
$k = 8$	0.799	1.296	0.764	0.771	0.766	0.795	0.788	0.875	0.822	0.85	0.837
$k = 9$	0.793	1.446	0.774	0.775	0.78	0.787	0.798	0.884	0.818	**0.845**	0.806
$k = 10$	0.771	1.585	0.781	0.785	0.786	0.802	0.781	0.888	0.82	0.888	**0.767**
$k = 11$	0.768	1.695	0.784	0.793	0.788	0.804	0.755	0.897	0.823	0.892	0.781
$k = 12$	**0.745**	1.811	0.796	0.81	0.791	0.789	**0.746**	0.896	0.806	0.917	0.782

2. Labor Market

	PCA	PLS	Targeted Predictors			Weighted PC			Two Step	QMLE	One-sided
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.763	**0.633**	0.708	0.691	0.704	0.676	0.751	0.741	0.768	0.772	0.77
$k = 2$	0.646	0.656	0.688	0.671	**0.661**	0.676	0.682	**0.706**	0.659	0.682	0.659
$k = 3$	**0.631**	0.81	**0.656**	**0.645**	0.662	0.686	0.659	0.713	**0.649**	**0.675**	0.699
$k = 4$	0.658	0.848	0.686	0.675	0.685	0.693	0.681	0.724	0.664	0.686	0.69
$k = 5$	0.656	0.876	0.683	0.692	0.674	**0.658**	**0.648**	0.727	0.666	0.688	**0.658**
$k = 6$	0.653	0.922	0.68	0.679	0.684	0.66	0.652	0.739	0.663	0.705	0.695
$k = 7$	0.672	0.908	0.682	0.685	0.679	0.668	0.662	0.739	0.672	0.708	0.721
$k = 8$	0.69	0.936	0.694	0.702	0.674	0.676	0.678	0.74	0.7	0.734	0.721
$k = 9$	0.686	0.989	0.692	0.703	0.684	0.685	0.686	0.751	0.701	0.753	0.691
$k = 10$	0.678	1.036	0.701	0.706	0.682	0.709	0.685	0.754	0.704	0.791	0.68
$k = 11$	0.689	1.085	0.712	0.711	0.687	0.72	0.685	0.761	0.719	0.8	0.694
$k = 12$	0.688	1.148	0.717	0.716	0.681	0.711	0.685	0.758	0.714	0.84	0.696

Table B.3: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Given k, $k = 1, 2, ..., 12$: Whole 144 Target Variables by Category

3. Housing

	PCA	PLS	Targeted Predictors			Weighted PC			Two Step	QMLE	One-sided
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.956	0.981	**0.932**	**0.924**	0.959	0.993	0.986	**0.906**	0.953	0.954	0.964
$k = 2$	0.931	0.941	0.953	0.948	0.94	0.922	**0.86**	0.925	0.919	0.922	**0.934**
$k = 3$	**0.921**	1.054	0.973	0.968	0.987	**0.915**	0.908	0.926	0.911	**0.913**	0.95
$k = 4$	0.925	0.874	0.995	0.993	0.997	0.921	0.932	0.935	**0.909**	0.919	0.938
$k = 5$	0.929	**0.853**	0.983	0.983	0.979	0.929	0.934	0.926	0.921	0.928	0.944
$k = 6$	0.939	0.892	0.979	0.971	0.964	0.937	0.943	0.917	0.933	0.952	0.972
$k = 7$	0.972	0.907	0.981	0.965	0.953	0.929	0.938	0.927	0.954	0.968	0.981
$k = 8$	0.956	0.9	0.977	0.954	0.947	0.942	0.94	0.939	0.946	1.001	0.948
$k = 9$	0.969	0.946	0.964	0.963	0.938	0.947	0.95	0.962	0.96	1.011	0.946
$k = 10$	0.957	0.976	0.96	0.962	**0.932**	0.93	0.955	0.972	0.952	1.022	0.937
$k = 11$	0.945	1.012	0.979	0.968	0.936	0.923	0.951	0.98	0.933	1.02	0.936
$k = 12$	0.934	1.072	0.981	0.982	0.946	0.938	0.928	0.971	0.923	1.041	0.945

4. Consumption, Orders, Inventories

	PCA	PLS	Targeted Predictors			Weighted PC			Two Step	QMLE	One-sided
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.988	**0.862**	0.891	0.924	0.907	0.912	0.977	0.912	0.99	0.988	0.983
$k = 2$	0.865	1.1	0.883	**0.885**	**0.892**	0.878	**0.835**	**0.902**	0.873	**0.892**	**0.878**
$k = 3$	**0.862**	1.268	**0.871**	0.888	0.894	0.894	0.867	0.918	**0.871**	0.897	0.893
$k = 4$	0.873	1.381	0.898	0.899	0.927	0.911	0.876	0.943	0.882	0.911	0.902
$k = 5$	0.884	1.53	0.912	0.916	0.909	0.888	0.878	0.94	0.898	0.931	0.883
$k = 6$	0.867	1.704	0.922	0.926	0.919	0.888	0.874	0.947	0.885	0.908	0.893
$k = 7$	0.874	1.83	0.927	0.927	0.914	**0.869**	0.865	0.948	0.885	0.901	0.922
$k = 8$	0.89	1.976	0.931	0.931	0.918	0.884	0.88	0.957	0.901	0.942	0.923
$k = 9$	0.905	2.087	0.952	0.941	0.936	0.894	0.899	0.968	0.908	0.951	0.913
$k = 10$	0.906	2.21	0.96	0.952	0.943	0.933	0.898	0.97	0.928	1.024	0.926
$k = 11$	0.909	2.31	0.966	0.957	0.955	0.941	0.913	0.974	0.932	1.028	0.931
$k = 12$	0.904	2.357	0.985	0.958	0.965	0.929	0.901	0.965	0.93	1.052	0.94

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by all factor estimations methods, with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

5. Money and Credit

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.592	0.555	0.593	0.581	**0.548**	0.558	0.596	0.543	0.592	0.59	0.585
$k = 2$	**0.516**	**0.53**	0.533	**0.522**	0.553	**0.516**	**0.515**	**0.523**	0.521	0.541	**0.516**
$k = 3$	0.517	0.553	**0.526**	0.526	0.548	0.518	0.517	0.533	**0.517**	0.535	0.517
$k = 4$	0.522	0.569	0.541	0.548	0.566	0.519	0.519	0.539	0.521	**0.534**	0.519
$k = 5$	0.522	0.581	0.566	0.561	0.589	0.521	0.522	0.537	0.521	0.535	0.526
$k = 6$	0.524	0.615	0.571	0.576	0.578	0.526	0.526	0.536	0.524	0.535	0.531
$k = 7$	0.534	0.633	0.563	0.582	0.583	0.529	0.529	0.537	0.537	0.537	0.541
$k = 8$	0.534	0.655	0.575	0.593	0.577	0.534	0.535	0.536	0.542	0.552	0.544
$k = 9$	0.537	0.676	0.583	0.591	0.584	0.547	0.535	0.541	0.543	0.574	0.545
$k = 10$	0.543	0.699	0.595	0.597	0.579	0.553	0.536	0.559	0.546	0.581	0.554
$k = 11$	0.551	0.715	0.599	0.592	0.576	0.557	0.551	0.563	0.558	0.591	0.568
$k = 12$	0.557	0.74	0.599	0.587	0.577	0.563	0.555	0.567	0.569	0.613	0.568

6. Interest and Exchange Rates

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.93	**0.915**	0.97	0.98	0.956	0.926	0.927	1.004	0.926	0.918	0.923
$k = 2$	0.931	0.964	0.933	0.944	0.952	0.898	0.907	**0.973**	0.912	0.902	0.903
$k = 3$	0.916	1.048	**0.912**	**0.895**	**0.9**	0.89	0.883	0.995	0.913	0.903	0.918
$k = 4$	0.916	1.177	0.956	0.944	0.944	0.874	0.875	0.993	0.92	0.921	0.919
$k = 5$	0.928	1.262	0.982	0.949	0.948	0.879	0.881	0.979	0.931	0.949	0.912
$k = 6$	0.889	1.287	0.98	0.97	0.979	0.877	0.879	0.976	0.874	**0.881**	**0.892**
$k = 7$	**0.864**	1.309	1.007	1.03	1.002	**0.825**	**0.862**	0.99	**0.862**	0.881	0.908
$k = 8$	0.903	1.36	1.017	1.045	1.022	0.869	0.887	0.993	0.911	0.931	0.927
$k = 9$	0.917	1.416	1.01	1.042	1.017	0.904	0.897	1.008	0.928	0.956	0.939
$k = 10$	0.952	1.438	1.029	1.033	1.028	0.939	0.909	1.004	0.948	1.014	0.959
$k = 11$	0.955	1.495	1.034	1.036	1.044	0.966	0.913	0.991	0.957	1.019	0.99
$k = 12$	0.95	1.519	1.052	1.048	1.043	0.958	0.956	1.01	0.949	1.089	0.985

Table B.4: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Given k, $k = 1, 2, ..., 12$: Whole 144 Target Variables by Category

7. Prices

	PCA	PLS	Targeted Predictors			Weighted PC			Two Step	QMLE	One-sided
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.404	**0.4**	0.418	0.421	0.414	0.415	0.403	0.444	0.405	0.405	0.406
$k = 2$	0.408	0.42	0.414	0.421	0.404	0.406	**0.397**	**0.438**	0.411	0.412	0.412
$k = 3$	0.403	0.434	**0.412**	**0.411**	**0.4**	**0.385**	0.399	0.441	0.405	0.407	**0.397**
$k = 4$	**0.396**	0.437	0.414	0.416	0.404	0.4	0.402	0.447	**0.399**	0.402	0.399
$k = 5$	0.4	0.448	0.424	0.422	0.418	0.401	0.399	0.447	0.401	**0.398**	0.409
$k = 6$	0.412	0.443	0.428	0.426	0.423	0.406	0.41	0.448	0.413	0.412	0.409
$k = 7$	0.417	0.448	0.439	0.434	0.426	0.406	0.409	0.45	0.418	0.415	0.424
$k = 8$	0.416	0.448	0.438	0.431	0.43	0.409	0.411	0.45	0.418	0.42	0.427
$k = 9$	0.418	0.459	0.443	0.436	0.433	0.417	0.411	0.453	0.419	0.427	0.425
$k = 10$	0.432	0.465	0.447	0.44	0.432	0.428	0.422	0.453	0.434	0.457	0.43
$k = 11$	0.44	0.475	0.448	0.444	0.437	0.442	0.432	0.456	0.44	0.457	0.443
$k = 12$	0.439	0.487	0.453	0.446	0.44	0.443	0.436	0.46	0.439	0.461	0.445

8. Stock Market

	PCA	PLS	Targeted Predictors			Weighted PC			Two Step	QMLE	One-sided
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static
$k = 1$	0.923	**0.903**	1.012	0.933	**0.885**	0.986	0.951	1.018	0.921	**0.917**	0.928
$k = 2$	0.917	0.905	**0.873**	**0.85**	0.888	0.893	0.896	1.021	0.915	0.929	0.902
$k = 3$	0.921	1.025	0.895	0.869	0.9	0.881	0.894	1.024	0.919	0.927	0.917
$k = 4$	0.928	1.147	0.944	0.902	0.915	0.906	0.91	1.009	0.925	0.932	0.923
$k = 5$	0.93	1.225	0.982	0.941	0.966	0.913	0.929	**1.006**	0.928	0.937	0.936
$k = 6$	0.943	1.289	1.017	1.003	1.023	0.918	0.934	1.009	0.941	0.947	0.931
$k = 7$	0.903	1.362	1.036	1.005	1.014	0.886	0.893	1.012	0.905	0.927	0.939
$k = 8$	**0.899**	1.471	1.043	1.016	1.055	**0.878**	**0.887**	1.015	0.903	0.94	**0.893**
$k = 9$	0.902	1.568	1.045	1.02	1.074	0.879	0.89	1.03	**0.9**	0.939	0.898
$k = 10$	0.922	1.654	1.04	1.032	1.084	0.914	0.893	1.033	0.92	0.988	0.939
$k = 11$	0.935	1.725	1.057	1.07	1.078	0.935	0.924	1.05	0.947	0.997	0.986
$k = 12$	0.934	1.775	1.072	1.083	1.072	0.927	0.937	1.042	0.946	1.009	0.981

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by all factor estimations methods, with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

Industrial Production												
	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.638**										
$\hat{k}_{BIC}$	0.75	1.326	0.765	0.752	0.761	0.777	0.743	0.781	0.817	0.903	0.751	
$\hat{k}_{BN-p1}$	0.744	1.297	0.766	0.804	0.794	0.768	0.74	0.791	0.801	0.847	0.782	
$\hat{k}_{BN-p2}$	0.713	1.009	0.757	0.804	0.794	0.73	0.716	0.794	0.768	0.777	0.731	
$\hat{k}_{BN-p3}$	0.737	1.818	0.781	0.804	0.794	0.79	0.745	0.83	0.803	0.911	0.781	
$\hat{k}_{BN-BIC}$	0.685	0.847	**0.701**	0.757	0.808	0.757	0.723	**0.774**	0.721	0.784	0.77	
$\hat{k}_{AH}$	0.687	0.694	0.767	**0.707**	**0.692**	0.706	0.694	0.789	0.721	0.768	0.711	
$\hat{k}_{ON}$	**0.659**	0.677	0.759	0.741	0.706	**0.671**	**0.659**	0.776	**0.685**	**0.709**	**0.658**	
$\hat{k}_{ABC-L}$	0.708	0.945	0.734	0.727	0.729	0.74	0.719	0.789	0.757	0.807	0.719	
$\hat{k}_{ABC-S}$	0.731	1.348	0.76	0.776	0.78	0.765	0.739	0.808	0.785	0.835	0.756	
BN2007												**0.733**
Mean	0.713	1.06	0.754	0.764	0.762	0.745	0.72	0.792	0.762	0.816	0.74	0.733
Best k	0.699	0.638	0.717	0.719	0.699	0.722	0.708	0.771	0.732	0.799	0.741	0.741

Table B.5: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Information Criteria: Real Variables

Real Personal Income

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.68**										
$\hat{k}_{BIC}$	0.89	1.363	0.878	0.917	0.921	0.902	0.899	**0.953**	0.924	0.966	0.901	
$\hat{k}_{BN-p1}$	0.876	1.496	0.901	0.967	0.885	0.873	0.87	0.995	0.898	0.931	0.905	
$\hat{k}_{BN-p2}$	0.86	1.308	0.886	0.965	0.884	0.863	0.858	0.979	0.883	0.912	0.889	
$\hat{k}_{BN-p3}$	0.888	1.834	0.944	0.953	0.885	0.909	0.889	0.999	0.917	0.96	0.915	
$\hat{k}_{BN-BIC}$	0.826	0.922	0.848	0.851	0.906	0.861	0.843	0.963	0.842	0.879	0.865	
$\hat{k}_{AH}$	0.821	0.757	**0.841**	**0.789**	**0.797**	0.829	0.818	0.954	0.838	0.869	0.836	
$\hat{k}_{ON}$	**0.772**	0.736	0.863	0.881	0.874	**0.764**	**0.761**	0.964	**0.783**	**0.805**	**0.769**	
$\hat{k}_{ABC-L}$	0.834	1.161	0.859	0.846	0.871	0.846	0.84	0.989	0.857	0.904	0.844	
$\hat{k}_{ABC-S}$	0.867	1.44	0.898	0.905	0.914	0.865	0.866	0.981	0.889	0.918	0.878	
BN2007												**0.857**
Mean	0.848	1.17	0.88	0.897	0.882	0.857	0.849	0.975	0.87	0.905	0.867	0.857
Best k	0.745	0.685	0.744	0.753	0.76	0.755	0.746	0.837	0.798	0.845	0.767	0.767

Real Manufacturing & Trade Industries Sales

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.648**										
$\hat{k}_{BIC}$	0.694	1.64	0.789	0.771	0.785	0.776	0.706	0.784	0.776	0.918	0.795	
$\hat{k}_{BN-p1}$	0.711	1.473	0.741	0.78	0.763	0.719	0.701	0.757	0.74	0.797	0.743	
$\hat{k}_{BN-p2}$	0.69	1.067	0.729	0.779	0.763	0.691	0.676	0.764	0.705	0.768	0.7	
$\hat{k}_{BN-p3}$	0.718	2.191	0.765	0.78	0.763	0.74	0.716	0.835	0.754	0.876	0.758	
$\hat{k}_{BN-BIC}$	0.652	0.861	0.778	0.801	0.773	0.705	0.681	**0.736**	0.662	0.714	0.713	
$\hat{k}_{AH}$	0.655	0.717	0.765	0.741	0.729	0.649	0.646	0.744	0.667	0.7	0.655	
$\hat{k}_{ON}$	**0.644**	0.702	**0.728**	**0.732**	**0.707**	**0.625**	**0.629**	0.736	**0.647**	**0.663**	**0.626**	
$\hat{k}_{ABC-L}$	0.674	0.98	0.802	0.779	0.796	0.678	0.665	0.769	0.693	0.76	0.68	
$\hat{k}_{ABC-S}$	0.688	1.615	0.782	0.786	0.752	0.706	0.696	0.794	0.715	0.779	0.718	
BN2007												**0.702**
Mean	0.681	1.189	0.764	0.772	0.759	0.699	0.68	0.769	0.707	0.775	0.71	0.702
Best k	0.631	0.633	0.656	0.645	0.661	0.658	0.648	0.706	0.649	0.675	0.658	0.658

Nonagriculture Employment												
	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.498**										
$\hat{k}_{BIC}$	0.494	1.178	0.527	0.518	**0.486**	0.522	**0.494**	0.544	0.505	0.611	**0.5**	
$\hat{k}_{BN-p1}$	0.529	1.326	0.511	0.504	0.492	0.515	0.521	0.548	0.543	0.587	0.535	
$\hat{k}_{BN-p2}$	0.506	0.971	0.528	0.504	0.492	**0.5**	0.506	0.549	0.517	0.548	0.507	
$\hat{k}_{BN-p3}$	0.511	1.74	**0.5**	0.505	0.492	0.534	0.521	0.569	0.521	0.646	0.535	
$\hat{k}_{BN-BIC}$	**0.484**	0.67	0.505	0.509	0.492	0.516	0.507	**0.54**	**0.49**	0.52	0.55	
$\hat{k}_{AH}$	0.511	0.531	0.512	**0.495**	0.543	0.512	0.511	0.543	0.517	0.535	0.523	
$\hat{k}_{ON}$	0.5	0.545	0.537	0.518	0.545	0.502	0.498	0.541	0.503	**0.509**	0.511	
$\hat{k}_{ABC-L}$	0.509	0.877	0.517	0.512	0.517	0.506	0.508	0.543	0.522	0.552	0.525	
$\hat{k}_{ABC-S}$	0.503	1.297	0.521	0.51	0.49	0.523	0.512	0.554	0.524	0.59	0.518	
BN2007												**0.5**
Mean	0.505	0.963	0.518	0.508	0.505	0.514	0.509	0.548	0.516	0.566	0.523	0.5
Best k	0.921	0.853	0.932	0.924	0.932	0.915	0.86	0.906	0.909	0.913	0.934	0.934

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the four real target variables (Industrial Production, Real Personal Income, Real Manufacturing & Trade Industries Sales, Nonagriculture Employment) by all factor estimations methods is considered. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The two information criteria, PLS1 and BN2007 are only applied to PLS and One-sided estimation, respectively. Therefore the other factor estimations for those two criteria remain blank. The lag of target variables, p, is determined by BIC. The row Mean is the mean of the method over 11 information criteria. Best k is the best results in **bold** in Table B.1. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

	CPI											
	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.392**										
$\hat{k}_{BIC}$	0.407	0.436	0.43	0.426	0.412	0.401	**0.399**	0.445	0.411	0.446	0.427	
$\hat{k}_{BN-p1}$	0.409	0.426	0.406	0.411	0.406	0.406	0.406	0.444	0.413	0.415	0.415	
$\hat{k}_{BN-p2}$	**0.406**	0.405	0.413	0.416	0.406	0.403	0.404	0.449	**0.409**	**0.405**	0.416	
$\hat{k}_{BN-p3}$	0.424	0.48	0.417	0.414	0.408	0.43	0.422	0.456	0.43	0.457	0.435	
$\hat{k}_{BN-BIC}$	0.41	0.398	0.409	0.404	0.403	**0.399**	0.416	0.441	0.413	0.414	**0.406**	
$\hat{k}_{AH}$	0.412	0.395	0.452	0.439	0.401	0.406	0.409	0.443	0.413	0.411	0.414	
$\hat{k}_{ON}$	0.411	0.393	0.423	0.415	**0.396**	0.403	0.407	**0.44**	0.412	0.41	0.412	
$\hat{k}_{ABC-L}$	0.406	0.413	**0.402**	**0.403**	0.401	0.402	0.407	0.449	0.409	0.408	0.411	
$\hat{k}_{ABC-S}$	0.407	0.427	0.419	0.421	0.412	0.405	0.405	0.458	0.411	0.419	0.409	
BN2007												**0.408**
Mean	0.41	0.417	0.419	0.417	0.405	0.406	0.408	0.447	0.413	0.421	0.416	0.408
Best k	0.862	0.862	0.871	0.885	0.892	0.869	0.835	0.902	0.871	0.892	0.878	0.878

Table B.6: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Information Criteria: Nominal Variables

Consumption Deflator

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.492**										
$\hat{k}_{BIC}$	**0.492**	0.595	0.514	0.521	0.504	0.515	0.5	0.514	0.499	0.55	0.528	
$\hat{k}_{BN-p1}$	0.515	0.57	0.513	0.512	0.506	0.513	0.51	0.518	0.516	0.515	0.52	
$\hat{k}_{BN-p2}$	0.497	0.511	0.507	**0.51**	0.506	0.5	0.497	0.521	0.498	**0.492**	0.512	
$\hat{k}_{BN-p3}$	0.534	0.63	0.523	0.526	**0.501**	0.542	0.53	0.539	0.535	0.566	0.544	
$\hat{k}_{BN-BIC}$	0.497	0.496	**0.504**	0.515	0.519	**0.489**	0.507	0.514	0.5	0.495	**0.49**	
$\hat{k}_{AH}$	0.497	0.504	0.528	0.521	0.501	0.492	0.496	**0.512**	0.496	0.496	0.501	
$\hat{k}_{ON}$	0.497	0.503	0.522	0.536	0.501	0.489	**0.495**	0.514	**0.495**	0.496	0.498	
$\hat{k}_{ABC-L}$	0.502	0.528	0.506	0.514	0.505	0.496	0.505	0.529	0.504	0.5	0.505	
$\hat{k}_{ABC-S}$	0.514	0.575	0.522	0.53	0.522	0.514	0.515	0.542	0.515	0.519	0.517	
BN2007												**0.5**
Mean	0.505	0.54	0.515	0.521	0.507	0.506	0.506	0.523	0.506	0.514	0.513	0.5
Best k	0.516	0.53	0.526	0.522	0.548	0.516	0.515	0.523	0.517	0.534	0.516	0.516

CPI excluding Food

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.427**										
$\hat{k}_{BIC}$	0.468	0.515	0.489	0.478	0.486	0.471	0.473	0.5	0.468	0.51	0.476	
$\hat{k}_{BN-p1}$	0.48	0.486	0.493	0.467	0.482	0.476	0.479	0.498	0.48	0.476	0.48	
$\hat{k}_{BN-p2}$	0.474	0.459	0.492	0.473	0.482	0.471	0.474	0.505	0.474	0.464	0.476	
$\hat{k}_{BN-p3}$	0.488	0.535	0.492	**0.46**	0.479	0.495	0.49	0.508	0.488	0.512	0.495	
$\hat{k}_{BN-BIC}$	0.468	0.454	0.479	0.471	0.482	0.459	0.474	**0.491**	0.469	0.466	0.463	
$\hat{k}_{AH}$	0.458	0.437	**0.422**	0.494	0.489	0.458	0.46	0.495	0.458	0.454	0.464	
$\hat{k}_{ON}$	**0.454**	0.437	0.478	0.477	**0.469**	**0.454**	**0.456**	0.491	**0.454**	**0.451**	**0.46**	
$\hat{k}_{ABC-L}$	0.464	0.461	0.483	0.474	0.475	0.462	0.466	0.501	0.466	0.458	0.47	
$\hat{k}_{ABC-S}$	0.478	0.488	0.481	0.465	0.476	0.473	0.478	0.504	0.479	0.482	0.477	
BN2007												**0.469**
Mean	0.47	0.47	0.479	0.473	0.48	0.469	0.472	0.499	0.471	0.475	0.473	0.469
Best k	0.864	0.915	0.912	0.895	0.9	0.825	0.862	0.973	0.862	0.881	0.892	0.892

Producer Price Index

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		0.342										
$\hat{k}_{BIC}$	0.341	0.345	0.354	0.345	0.35	0.343	0.341	0.335	0.345	0.374	0.352	
$\hat{k}_{BN-p1}$	0.339	0.346	0.354	**0.34**	**0.348**	**0.337**	0.337	0.334	0.343	0.347	0.345	
$\hat{k}_{BN-p2}$	0.339	0.337	**0.349**	0.341	0.349	0.34	0.34	0.334	0.345	0.345	0.349	
$\hat{k}_{BN-p3}$	0.348	0.371	0.353	0.341	0.349	0.353	0.346	0.34	0.354	0.369	0.357	
$\hat{k}_{BN-BIC}$	0.339	0.34	0.349	0.358	0.354	0.339	0.343	**0.333**	0.342	**0.343**	0.341	
$\hat{k}_{AH}$	0.343	0.335	0.355	0.355	0.362	0.339	0.341	0.335	0.344	0.344	0.343	
$\hat{k}_{ON}$	0.345	**0.334**	0.355	0.355	0.349	0.339	0.342	0.334	0.345	0.345	0.344	
$\hat{k}_{ABC-L}$	0.341	0.344	0.354	0.353	0.351	0.339	0.341	0.336	0.345	0.346	0.342	
$\hat{k}_{ABC-S}$	**0.337**	0.348	0.35	0.354	0.351	0.338	**0.336**	0.339	**0.34**	0.344	**0.339**	
BN2007												**0.345**
Mean	0.341	0.344	0.353	0.349	0.351	0.341	0.341	0.336	0.345	0.351	0.346	0.345
Best k	0.396	0.4	0.412	0.411	0.4	0.385	0.397	0.438	0.399	0.398	0.397	0.397

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the four nominal target variables (CPI, Consumption Deflator, CPI excluding Food, Producer Price Index) by all factor estimations methods is considered. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The two information criteria, PLS1 and BN2007 are only applied to PLS and One-sided estimation, respectively. Therefore the other factor estimations for those two criteria remain blank. The row Mean is the mean of the method over 11 information criteria. Best k is the best results in **bold** in Table B.2. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

1. Output and Income

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.685**										
$\hat{k}_{BIC}$	0.749	1.327	0.764	**0.755**	0.762	0.789	0.757	0.855	0.813	0.888	0.766	
$\hat{k}_{BN-p1}$	0.804	1.28	0.774	0.809	0.794	0.807	0.798	0.865	0.835	0.865	0.82	
$\hat{k}_{BN-p2}$	0.787	1.014	0.767	0.803	0.793	0.789	0.78	0.869	0.805	0.856	0.797	
$\hat{k}_{BN-p3}$	0.745	1.811	0.793	0.809	0.793	0.789	0.746	0.887	0.806	0.917	0.782	
$\hat{k}_{BN-BIC}$	0.765	0.827	**0.748**	0.756	0.793	0.826	0.797	0.847	0.782	0.832	0.829	
$\hat{k}_{AH}$	0.761	0.713	0.816	0.779	**0.757**	0.772	0.752	0.842	0.775	0.801	0.759	
$\hat{k}_{ON}$	**0.73**	0.708	0.825	0.814	0.789	**0.715**	**0.712**	**0.839**	**0.742**	**0.759**	**0.719**	
$\hat{k}_{ABC-L}$	0.77	0.937	0.775	0.761	0.782	0.786	0.77	0.866	0.8	0.856	0.776	
$\hat{k}_{ABC-S}$	0.781	1.321	0.772	0.777	0.783	0.801	0.786	0.873	0.812	0.848	0.79	
BN2007												**0.769**
Mean	0.766	1.062	0.782	0.785	0.783	0.786	0.766	0.86	0.797	0.847	0.782	0.769
Best k	0.745	0.685	0.744	0.753	0.76	0.755	0.746	0.837	0.798	0.845	0.767	0.767

Table B.7: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Information Criteria: Whole 144 Target Variables by Category

2. Labor Market

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.633**										
$\hat{k}_{BIC}$	0.667	1.016	0.694	0.693	**0.665**	0.693	0.669	0.714	0.678	0.789	0.681	
$\hat{k}_{BN-p1}$	0.687	0.968	0.682	0.708	0.681	0.687	0.679	0.723	0.696	0.752	0.688	
$\hat{k}_{BN-p2}$	0.659	0.905	0.684	0.705	0.685	0.652	0.656	0.726	0.671	0.713	0.672	
$\hat{k}_{BN-p3}$	0.688	1.148	0.686	0.713	0.681	0.711	0.685	0.746	0.714	0.84	0.696	
$\hat{k}_{BN-BIC}$	**0.621**	0.813	0.69	0.674	0.67	0.673	**0.646**	0.714	**0.627**	**0.657**	0.684	
$\hat{k}_{AH}$	0.666	0.659	**0.672**	0.67	0.705	0.673	0.672	**0.711**	0.678	0.691	0.684	
$\hat{k}_{ON}$	0.665	0.644	0.679	**0.659**	0.683	0.654	0.658	0.718	0.663	0.665	**0.664**	
$\hat{k}_{ABC-L}$	0.645	0.869	0.679	0.669	0.671	**0.645**	0.646	0.722	0.657	0.692	0.678	
$\hat{k}_{ABC-S}$	0.661	0.927	0.693	0.693	0.673	0.682	0.669	0.75	0.685	0.729	0.67	
BN2007												**0.666**
Mean	0.662	0.858	0.684	0.687	0.679	0.674	0.664	0.725	0.674	0.725	0.68	0.666
Best k	0.631	0.633	0.656	0.645	0.661	0.658	0.648	0.706	0.649	0.675	0.658	0.658

3. Housing

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		0.981										
$\hat{k}_{BIC}$	0.949	0.975	0.965	0.956	0.947	0.943	0.964	0.932	0.929	1.018	0.95	
$\hat{k}_{BN-p1}$	0.967	0.937	0.986	0.979	0.946	0.952	0.951	0.926	0.965	1.003	0.942	
$\hat{k}_{BN-p2}$	0.99	0.993	0.994	0.983	0.946	0.955	0.969	0.938	0.998	1.022	0.963	
$\hat{k}_{BN-p3}$	0.934	1.072	0.981	0.985	0.946	0.938	0.928	0.979	0.923	1.041	0.945	
$\hat{k}_{BN-BIC}$	0.934	0.927	0.982	0.972	0.989	0.943	0.929	0.928	0.918	0.923	0.961	
$\hat{k}_{AH}$	**0.858**	0.97	**0.93**	**0.944**	0.96	**0.829**	**0.848**	0.92	**0.853**	**0.85**	**0.84**	
$\hat{k}_{ON}$	0.875	0.983	0.968	0.971	0.957	0.842	0.863	**0.918**	0.858	0.85	0.858	
$\hat{k}_{ABC-L}$	0.938	**0.905**	1.007	0.983	0.991	0.941	0.935	0.952	0.928	0.951	0.958	
$\hat{k}_{ABC-S}$	0.956	0.922	0.968	0.951	**0.94**	0.943	0.953	0.951	0.952	0.997	0.967	
BN2007												**0.966**
Mean	0.933	0.966	0.976	0.969	0.958	0.921	0.927	0.938	0.925	0.962	0.932	0.966
Best k	0.921	0.853	0.932	0.924	0.932	0.915	0.86	0.906	0.909	0.913	0.934	0.934

4. Consumption, Orders, Inventories

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.862**										
$\hat{k}_{BIC}$	0.899	1.763	0.936	0.924	0.924	0.914	0.88	0.909	0.938	1.037	0.937	
$\hat{k}_{BN-p1}$	0.904	1.946	0.945	0.95	0.958	0.905	0.895	0.925	0.923	0.948	0.924	
$\hat{k}_{BN-p2}$	0.864	1.664	0.936	0.955	0.955	0.865	0.865	0.925	0.886	0.918	0.895	
$\hat{k}_{BN-p3}$	0.904	2.357	0.975	0.964	0.964	0.929	0.901	0.959	0.93	1.052	0.94	
$\hat{k}_{BN-BIC}$	0.848	1.255	0.901	0.916	0.926	0.881	0.879	0.914	0.861	0.881	0.89	
$\hat{k}_{AH}$	**0.835**	1.045	0.865	0.907	0.921	0.832	0.832	**0.902**	**0.844**	0.863	0.845	
$\hat{k}_{ON}$	0.838	1.048	**0.849**	**0.858**	**0.879**	**0.827**	**0.823**	0.915	0.848	**0.856**	**0.836**	
$\hat{k}_{ABC-L}$	0.864	1.454	0.925	0.91	0.918	0.862	0.865	0.937	0.882	0.9	0.868	
$\hat{k}_{ABC-S}$	0.892	1.949	0.929	0.936	0.946	0.888	0.887	0.953	0.899	0.934	0.911	
BN2007												**0.889**
Mean	0.872	1.534	0.918	0.924	0.932	0.878	0.87	0.927	0.89	0.932	0.894	0.889
Best k	0.862	0.862	0.871	0.885	0.892	0.869	0.835	0.902	0.871	0.892	0.878	0.878

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by all factor estimations methods is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The two information criteria, PLS1 and BN2007 are only applied to PLS and One-sided estimation, respectively. Therefore the other factor estimations for those two criteria remain blank. The row Mean is the mean of the method over 11 information criteria. Best k is the best results in **bold** in Table B.3. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

5. Money and Credit

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		0.555										
$\hat{k}_{BIC}$	0.54	0.677	0.574	0.568	0.57	0.54	0.519	0.53	0.55	0.58	0.559	
$\hat{k}_{BN-p1}$	0.53	0.643	0.543	0.57	0.578	0.53	0.532	0.541	0.539	0.567	0.544	
$\hat{k}_{BN-p2}$	0.53	0.594	0.546	0.56	0.577	0.534	0.531	0.546	0.535	0.555	0.558	
$\hat{k}_{BN-p3}$	0.557	0.74	0.574	0.578	0.577	0.563	0.555	0.553	0.569	0.613	0.568	
$\hat{k}_{BN-BIC}$	0.52	0.547	0.536	0.541	0.573	**0.516**	0.517	0.532	0.52	**0.531**	0.519	
$\hat{k}_{AH}$	**0.515**	**0.534**	0.587	0.582	0.56	0.519	**0.514**	**0.52**	**0.515**	0.534	**0.515**	
$\hat{k}_{ON}$	0.523	0.554	**0.532**	**0.522**	**0.532**	0.528	0.524	0.527	0.529	0.536	0.521	
$\hat{k}_{ABC-L}$	0.522	0.597	0.543	0.544	0.564	0.521	0.522	0.536	0.521	0.535	0.524	
$\hat{k}_{ABC-S}$	0.536	0.663	0.574	0.592	0.587	0.535	0.536	0.551	0.544	0.559	0.544	
BN2007												**0.544**
Mean	0.53	0.61	0.557	0.562	0.569	0.532	0.528	0.537	0.536	0.557	0.539	0.544
Best k	0.516	0.53	0.526	0.522	0.548	0.516	0.515	0.523	0.517	0.534	0.516	0.516

Table B.8: 12-Month-Ahead DIAR Forecasts by All Factor Estimations With Information Criteria: Whole 144 Target Variables by Category

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.915**										
$\hat{k}_{BIC}$	0.939	1.463	1.001	1	0.966	0.965	0.968	1.004	0.935	1.066	0.951	
$\hat{k}_{BN-p1}$	0.92	1.37	0.991	1.037	1.044	0.907	**0.89**	1.01	0.925	0.923	0.936	
$\hat{k}_{BN-p2}$	0.926	1.278	0.979	1.017	1.036	0.915	0.916	1.01	0.92	0.941	0.923	
$\hat{k}_{BN-p3}$	0.95	1.519	1.041	1.048	1.043	0.958	0.956	1.014	0.949	1.089	0.985	
$\hat{k}_{BN-BIC}$	0.932	1.086	0.941	0.985	0.999	0.922	0.916	0.994	0.929	0.918	**0.903**	
$\hat{k}_{AH}$	0.923	0.96	0.955	0.952	0.931	**0.899**	0.904	**0.976**	**0.904**	**0.899**	0.904	
$\hat{k}_{ON}$	0.917	0.916	**0.924**	**0.908**	**0.9**	0.904	0.911	0.985	0.914	0.911	0.912	
$\hat{k}_{ABC-L}$	0.926	1.249	0.943	0.948	0.936	0.914	0.917	0.995	0.924	0.941	0.915	
$\hat{k}_{ABC-S}$	**0.904**	1.382	1.019	1.023	1.027	0.908	0.895	0.994	0.921	0.951	0.925	
BN2007												**0.965**
Mean	0.926	1.214	0.977	0.991	0.987	0.921	0.919	0.998	0.925	0.96	0.928	0.965
Best k	0.864	0.915	0.912	0.895	0.9	0.825	0.862	0.973	0.862	0.881	0.892	0.892

Table caption row: **6. Interest and Exchange Rates**

7. Prices

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		**0.4**										
$\hat{k}_{BIC}$	0.425	0.463	0.442	0.441	0.428	0.42	0.417	0.446	0.422	0.451	0.425	
$\hat{k}_{BN-p1}$	0.42	0.454	0.409	0.414	0.424	0.418	0.411	0.442	0.421	0.423	0.426	
$\hat{k}_{BN-p2}$	0.405	0.433	**0.408**	0.416	0.426	0.402	0.403	0.445	0.408	0.404	0.414	
$\hat{k}_{BN-p3}$	0.439	0.487	0.421	0.429	0.433	0.443	0.436	0.452	0.439	0.461	0.445	
$\hat{k}_{BN-BIC}$	0.407	0.426	0.408	0.41	0.411	**0.397**	0.412	0.44	0.409	0.412	0.405	
$\hat{k}_{AH}$	0.398	0.416	0.416	0.422	0.401	0.401	0.398	0.441	0.399	0.4	0.405	
$\hat{k}_{ON}$	**0.395**	0.412	0.412	**0.409**	**0.399**	0.397	**0.395**	**0.439**	**0.396**	**0.397**	**0.401**	
$\hat{k}_{ABC-L}$	0.405	0.437	0.408	0.411	0.41	0.4	0.405	0.449	0.408	0.407	0.409	
$\hat{k}_{ABC-S}$	0.417	0.453	0.44	0.433	0.436	0.419	0.413	0.453	0.419	0.427	0.421	
BN2007												**0.407**
Mean	0.412	0.438	0.418	0.421	0.419	0.411	0.41	0.445	0.413	0.42	0.417	0.407
Best k	0.396	0.4	0.412	0.411	0.4	0.385	0.397	0.438	0.399	0.398	0.397	0.397

8. Stock Market

	PCA	**PLS**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
			$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
PLS1		0.903										
$\hat{k}_{BIC}$	0.908	1.703	1.074	1.049	1.006	0.924	0.901	**1.003**	0.941	1.043	0.98	
$\hat{k}_{BN-p1}$	0.899	1.496	1.053	1.084	1.072	0.873	0.889	1.004	0.898	0.946	0.895	
$\hat{k}_{BN-p2}$	0.944	1.383	1.053	1.082	1.071	0.926	0.941	1.006	0.951	0.961	0.945	
$\hat{k}_{BN-p3}$	0.934	1.775	1.071	1.089	1.072	0.927	0.937	1.043	0.946	1.009	0.981	
$\hat{k}_{BN-BIC}$	0.924	1.065	1.004	0.97	1.013	0.889	0.904	1.021	0.925	0.932	0.924	
$\hat{k}_{AH}$	0.915	0.913	1.007	0.908	**0.849**	0.901	0.899	1.005	0.916	0.926	0.905	
$\hat{k}_{ON}$	**0.873**	**0.894**	**0.895**	**0.875**	0.885	**0.862**	**0.869**	1.027	**0.864**	**0.875**	**0.864**	
$\hat{k}_{ABC-L}$	0.937	1.188	0.948	0.94	0.945	0.92	0.93	1.033	0.933	0.945	0.942	
$\hat{k}_{ABC-S}$	0.915	1.419	1.027	1.034	1.054	0.893	0.901	1.017	0.918	0.936	0.919	
BN2007												**0.964**
Mean	0.917	1.274	1.015	1.003	0.996	0.902	0.908	1.018	0.921	0.953	0.928	0.964
Best k	0.899	0.903	0.873	0.85	0.885	0.878	0.887	1.006	0.9	0.917	0.893	0.893

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by all factor estimations methods is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The two information criteria, PLS1 and BN2007 are only applied to PLS and One-sided estimation, respectively. Therefore the other factor estimations for those two criteria remain blank. The row Mean is the mean of the method over 11 information criteria. Best k is the best results in **bold** in Table B.4. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : Mean	**PCA**	**PLS1**	**PLS, other IC**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
				$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
Industrial Production	0.713	**0.638**	1.107	0.754	0.764	0.762	0.745	0.72	0.792	0.762	0.816	0.74	0.733
Personal Income	0.848	**0.68**	1.224	0.88	0.897	0.882	0.857	0.849	0.975	0.87	0.905	0.867	0.857
Mfg & Trade Sales	0.681	**0.648**	1.25	0.764	0.772	0.759	0.699	0.68	0.769	0.707	0.775	0.71	0.702
Nonag. Employment	0.505	**0.498**	1.015	0.518	0.508	0.505	0.514	0.509	0.548	0.516	0.566	0.523	0.5
CPI	0.41	**0.392**	0.419	0.419	0.417	0.405	0.406	0.408	0.447	0.413	0.421	0.416	0.408
Consumption Deflator	0.505	**0.492**	0.546	0.515	0.521	0.507	0.506	0.506	0.523	0.506	0.514	0.513	0.5
CPI exc. Food	0.47	**0.427**	0.475	0.479	0.473	0.48	0.469	0.472	0.499	0.471	0.475	0.473	0.469
Producer Price Index	0.341	0.342	0.344	0.353	0.349	0.351	0.341	0.341	**0.336**	0.345	0.351	0.346	0.345

Table B.9: 12-Month-Ahead DIAR Forecasts For 8 Target Variables: Mean of All Factor-Augmented Forecasts by Information Criteria

Notes: The entries are mean relative mean squared errors (RMSE) of the method relative to a forecast based on the target's historical mean. 12-month-ahead DIAR forecasting by all factor estimations methods is considered, for the eight target variables: (Industrial Production, Real Personal Income, Real Manufacturing & Trade Industries Sales, Nonagriculture Employment, CPI, Consumption Deflator, CPI excluding Food, Producer Price Index) The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The mean over 11 information criteria for the given method is presented. The entries are identical to the Mean rows in Table B.5 to Table B.6. This experiment is recursive out-of-sample forecast. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : Mean	**PCA**	**PLS1**	**PLS, other IC**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
				$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
Industrial Production	0.659	**0.638**	0.677	0.701	0.707	0.692	0.671	0.659	0.774	0.685	0.709	0.658	0.733
Personal Income	0.772	**0.68**	0.736	0.841	0.789	0.797	0.764	0.761	0.953	0.783	0.805	0.769	0.857
Mfg & Trade Sales	0.644	0.648	0.702	0.728	0.732	0.707	**0.625**	0.629	0.736	0.647	0.663	0.626	0.702
Nonag. Employment	**0.484**	0.498	0.531	0.5	0.495	0.486	0.5	0.494	0.54	0.49	0.509	0.5	0.5
CPI	0.406	**0.392**	0.393	0.402	0.403	0.396	0.399	0.399	0.44	0.409	0.405	0.406	0.408
Consumption Deflator	0.492	0.492	0.496	0.504	0.51	0.501	**0.489**	0.495	0.512	0.495	0.492	0.49	0.5
CPI exc. Food	0.454	0.427	0.437	**0.422**	0.46	0.469	0.454	0.456	0.491	0.454	0.451	0.46	0.469
Producer Price Index	0.337	0.342	0.334	0.349	0.34	0.348	0.337	0.336	**0.333**	0.34	0.343	0.339	0.345

Table B.10: 12-Month-Ahead DIAR Forecasts For 8 Target Variables: The Best Results of All Factor-Augmented Forecasts by Information Criteria

Notes: The entries are minimum relative mean squared errors (RMSE) of the method relative to a forecast based on the target's historical mean. 12-month-ahead DIAR forecasting by all factor estimations methods is considered, for the eight target variables: (Industrial Production, Real Personal Income, Real Manufacturing & Trade Industries Sales, Nonagriculture Employment, CPI, Consumption Deflator, CPI excluding Food, Producer Price Index) The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The minimum RMSE, or the best result, over 11 information criteria for the given method is presented. The entries are identical to the best results in **bold** in Table B.5 to Table B.6. This experiment is recursive out-of-sample forecast. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : Mean	**PCA**	**PLS1**	**PLS, other IC**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
				$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
Overall	0.756	**0.716**	0.991	0.788	0.788	0.781	0.768	0.758	0.82	0.774	0.806	0.763	0.758
Output and Income	0.766	**0.685**	1.104	0.782	0.785	0.783	0.786	0.766	0.86	0.797	0.847	0.782	0.769
Labor Market	0.662	**0.633**	0.883	0.684	0.687	0.679	0.674	0.664	0.725	0.674	0.725	0.68	0.666
Housing	0.933	0.981	0.965	0.976	0.969	0.958	**0.921**	0.927	0.938	0.925	0.962	0.932	0.966
Consumption, Orders, Inventories	0.872	**0.862**	1.609	0.918	0.924	0.932	0.878	0.87	0.927	0.89	0.932	0.894	0.889
Money and Credit	0.53	0.555	0.617	0.557	0.562	0.569	0.532	**0.528**	0.537	0.536	0.557	0.539	0.544
Interest and Exchange Rates	0.926	**0.915**	1.247	0.977	0.991	0.987	0.921	0.919	0.998	0.925	0.96	0.928	0.965
Prices	0.412	**0.4**	0.442	0.418	0.421	0.419	0.411	0.41	0.445	0.413	0.42	0.417	0.407
Stock Market	0.917	0.903	1.315	1.015	1.003	0.996	**0.902**	0.908	1.018	0.921	0.953	0.928	0.964

Table B.11: 12-Month-Ahead DIAR Forecasts For The Whole 144 Target Variables by Categories: Mean of All Factor-Augmented Forecasts by Information Criteria

Notes: The entries are mean relative mean squared errors (RMSE) of the method relative to a forecast based on the target's historical mean. 12-month-ahead DIAR forecasting by all factor estimations methods is considered, for the whole 144 target variables. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The mean RMSE over 11 information criteria for the given method for each target variable is calculated. Then the whole 144 mean RMSE for each factor estimation methods are divided into eight categories, according to target variables. The median of given method in the category is presented. The entries are identical to the Mean rows in Table B.7 to Table B.8. This experiment is recursive out-of-sample forecast. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : Minimum	**PCA**	**PLS1**	**PLS, other IC**	**Targeted Predictors**			**Weighted PC**			**Two Step**	**QMLE**	**One-sided**	
				$\lambda_2 = 0.25$	$\lambda_2 = 0.5$	$\lambda_2 = 1.5$	SWa	SWb	Rule B			Static	Dynamic
Overall	0.736	**0.716**	0.74	0.77	0.768	0.763	0.748	0.744	0.795	0.75	0.761	0.741	0.758
Output and Income	0.73	**0.685**	0.708	0.748	0.755	0.757	0.715	0.712	0.839	0.742	0.759	0.719	0.769
Labor Market	**0.621**	0.633	0.644	0.672	0.659	0.665	0.645	0.646	0.711	0.627	0.657	0.664	0.666
Housing	0.858	0.981	0.905	0.93	0.944	0.94	**0.829**	0.848	0.918	0.853	0.85	0.84	0.966
Consumption, Orders, Inventories	0.835	0.862	1.045	0.849	0.858	0.879	0.827	**0.823**	0.902	0.844	0.856	0.836	0.889
Money and Credit	0.515	0.555	0.534	0.532	0.522	0.532	0.516	**0.514**	0.52	0.515	0.531	0.515	0.544
Interest and Exchange Rates	0.904	0.915	0.916	0.924	0.908	0.9	0.899	**0.89**	0.976	0.904	0.899	0.903	0.965
Prices	**0.395**	0.4	0.412	0.408	0.409	0.399	0.397	0.395	0.439	0.396	0.397	0.401	0.407
Stock Market	0.873	0.903	0.894	0.895	0.875	**0.849**	0.862	0.869	1.003	0.864	0.875	0.864	0.964

Table B.12: 12-Month-Ahead DIAR Forecasts For The Whole 144 Target Variables by Categories: The Best Results of All Factor-Augmented Forecasts by Information Criteria

Notes: The entries are minimum relative mean squared errors (RMSE) of the method relative to a forecast based on the target's historical mean. 12-month-ahead DIAR forecasting by all factor estimations methods is considered, for the whole 144 target variables. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. The minimum RMSE, or the best result, over 11 information criteria for the given method for each target variable is calculated. Then the whole 144 minimum RMSE for each factor estimation methods are divided into eight categories, according to target variables. The median of given method in the category is presented. The entries are identical to the best results in **bold** in Table B.7 to Table B.8. This experiment is recursive out-of-sample forecast. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : Mean	**PCA**	**PLS1**	**PLS, other IC**	**Targeted Predictors**	**Weighted PC**	**Two Step**	**QMLE**	**One-sided**	**Range exc. PLS, other IC**	**Range**
	Percentile = 25									
Overall	0.468	**0.459**	0.537	0.479	0.482	0.469	0.488	0.476	0.069	0.078
Output and Income	0.695	**0.646**	0.984	0.746	0.745	0.746	0.796	0.723	0.289	0.338
Labor Market	0.467	**0.449**	0.595	0.462	0.524	0.464	0.512	0.479	0.133	0.146
Housing	0.403	**0.365**	0.414	0.446	0.406	0.399	0.417	0.406	0.047	0.081
Consumption, Orders, Inventories	0.745	**0.675**	1.273	0.781	0.775	0.767	0.81	0.762	0.528	0.598
Money and Credit	0.451	**0.43**	0.503	0.467	0.45	0.454	0.456	0.453	0.053	0.073
Interest and Exchange Rates	0.795	**0.753**	1.106	0.861	0.847	0.789	0.837	0.808	0.317	0.353
Prices	0.346	**0.342**	0.344	0.351	0.345	0.35	0.356	0.351	0.012	0.014
Stock Market	0.796	**0.785**	0.893	0.799	0.85	0.804	0.83	0.797	0.097	0.108
	Percentile = 50									
Overall	0.657	**0.628**	0.862	0.691	0.697	0.668	0.701	0.667	0.205	0.234
Output and Income	0.786	**0.68**	1.111	0.769	0.815	0.809	0.862	0.793	0.342	0.431
Labor Market	0.561	**0.553**	0.792	0.6	0.604	0.577	0.625	0.58	0.231	0.239
Housing	0.459	0.497	0.49	0.466	0.46	**0.451**	0.469	0.46	0.039	0.046
Consumption, Orders, Inventories	0.856	**0.805**	1.55	0.907	0.86	0.874	0.913	0.868	0.694	0.745
Money and Credit	**0.517**	0.555	0.619	0.556	0.525	0.531	0.561	0.531	0.102	0.102
Interest and Exchange Rates	0.908	**0.906**	1.209	0.986	0.922	0.914	0.946	0.926	0.301	0.303
Prices	0.414	**0.4**	0.444	0.421	0.425	0.414	0.421	0.414	0.03	0.044
Stock Market	0.912	**0.903**	1.315	1.008	0.938	0.92	0.952	0.929	0.403	0.412

Table B.13: 12-Month-Ahead DIAR Forecasts For The Whole 144 Target Variables by Categories: Mean of All Factor-Augmented Forecasts by Information Criteria, 25th, 50th and 75th Percentiles

	Percentile = 75									
Overall	0.846	**0.788**	1.248	0.889	0.862	0.858	0.897	0.859	0.402	0.46
Output and Income	0.856	**0.733**	1.231	0.816	0.865	0.87	0.905	0.865	0.415	0.498
Labor Market	0.661	**0.624**	0.955	0.686	0.727	0.675	0.718	0.674	0.294	0.331
Housing	0.495	0.569	0.554	0.51	0.495	**0.487**	0.502	0.49	0.067	0.082
Consumption, Orders, Inventories	**0.987**	0.99	1.74	1.004	1.013	1.002	1.031	0.991	0.753	0.753
Money and Credit	0.701	**0.696**	0.742	0.718	0.705	0.713	0.717	0.707	0.041	0.046
Interest and Exchange Rates	1.044	**1.021**	1.394	1.141	1.032	1.027	1.074	1.051	0.367	0.373
Prices	0.487	**0.458**	0.539	0.514	0.483	0.49	0.497	0.489	0.056	0.081
Stock Market	1.004	**0.991**	1.539	1.14	1.007	1.014	1.044	1.032	0.535	0.548

Notes: The entries are relative mean squared errors (RMSE) of the respective method relative to a forecast based on the target's historical mean. First, 12-month-ahead DIAR forecasting by all factor estimations methods is considered, for the whole 144 target variables. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. Second, the mean RMSE over 11 information criteria for the given method for each target variable is calculated. Third, then the whole 144 mean RMSE for each factor estimation methods are divided into eight categories, according to target variables. Finally, the 25th, 50th and 75th percentiles of each method in the category is presented. The Rnage exc. PLS, other IC column is the range of each row, excluding PLS, other IC. The Range column is the range of each row This experiment is recursive out-of-sample forecast. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : Mean	**PCA**	**PLS1**	**PLS, other IC**	**Targeted Predictors**	**Weighted PC**	**Two Step**	**QMLE**	**One-sided**	**Range**
				Percentile = 25					
Overall	0.442	0.459	0.455	0.451	**0.434**	0.441	0.448	0.451	0.021
Output and Income	0.659	**0.646**	0.666	0.682	0.659	0.683	0.704	0.66	0.045
Labor Market	0.424	0.449	0.44	0.436	0.436	**0.416**	0.453	0.466	0.05
Housing	0.385	**0.365**	0.375	0.413	0.374	0.38	0.379	0.378	0.039
Consumption, Orders, Inventories	0.695	0.675	0.728	0.726	**0.671**	0.701	0.728	0.676	0.057
Money and Credit	0.431	0.43	0.429	0.436	**0.423**	0.431	0.432	0.429	0.013
Interest and Exchange Rates	0.765	0.753	0.784	**0.735**	0.75	0.759	0.764	0.763	0.049
Prices	0.341	0.342	**0.334**	0.34	0.336	0.344	0.347	0.345	0.013
Stock Market	0.732	0.785	0.716	0.738	**0.704**	0.741	0.752	0.72	0.048
				Percentile = 50					
Overall	**0.613**	0.628	0.639	0.62	0.617	0.619	0.626	0.617	0.026
Output and Income	0.726	0.68	**0.679**	0.706	0.706	0.748	0.767	0.706	0.088
Labor Market	**0.529**	0.553	0.552	0.56	0.537	0.543	0.561	0.549	0.032
Housing	0.424	0.497	0.458	0.428	**0.407**	0.416	0.414	0.41	0.051
Consumption, Orders, Inventories	0.808	0.805	0.861	0.833	**0.801**	0.826	0.843	0.819	0.06
Money and Credit	**0.493**	0.555	0.528	0.517	0.495	0.509	0.531	0.5	0.038
Interest and Exchange Rates	0.872	0.906	0.897	0.847	**0.842**	0.88	0.881	0.881	0.055
Prices	0.393	0.4	0.412	0.396	**0.39**	0.395	0.394	0.396	0.022
Stock Market	0.867	0.903	0.894	**0.84**	0.849	0.863	0.874	0.856	0.054

Table B.14: 12-Month-Ahead DIAR Forecasts For The Whole 144 Target Variables by Categories: The Best Results of All Factor-Augmented Forecasts by Information Criteria, 25th, 50th and 75th Percentiles

	Percentile = 75								
Overall	0.795	0.788	0.819	0.795	0.784	0.803	0.809	**0.783**	0.036
Output and Income	0.814	**0.733**	0.736	0.761	0.784	0.83	0.813	0.78	0.094
Labor Market	**0.616**	0.624	0.638	0.626	0.626	0.627	0.665	0.63	0.049
Housing	0.442	0.569	0.521	0.477	**0.425**	0.435	0.432	0.429	0.096
Consumption, Orders, Inventories	**0.932**	0.99	1.087	0.941	0.941	0.94	0.952	0.94	0.155
Money and Credit	0.67	0.696	0.69	0.672	0.658	0.663	0.679	**0.641**	0.049
Interest and Exchange Rates	0.983	1.021	1.029	1.012	0.969	0.98	**0.967**	0.975	0.062
Prices	**0.454**	0.458	0.458	0.471	0.454	0.459	0.473	0.46	0.019
Stock Market	0.962	0.991	0.993	**0.894**	0.937	0.954	0.96	0.94	0.099

Notes: The entries are relative mean squared errors (RMSE) of the respective method relative to a forecast based on the target's historical mean. First, 12-month-ahead DIAR forecasting by all factor estimations methods is considered, for the whole 144 target variables. The number of contemporaneous factors, k is determined by 11 information criteria, PLS1, $\hat{k}_{BIC}$, to BN2007. Second, the minimum RMSE, or the best results, over 11 information criteria for the given method for each target variable is calculated. Third, then the whole 144 minimum RMSE for each factor estimation methods are divided into eight categories, according to target variables. Finally, the 25th, 50th and 75th percentiles of each method in the category is presented. The Range column is the range of each row This experiment is recursive out-of-sample forecast. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DI : PCA	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Industrial Production	0.893	0.699	0.684	0.7	0.703	0.666	0.664	0.669	**0.663**	0.668	0.68	0.672
Personal Income	0.845	**0.841**	0.846	0.869	0.873	0.862	0.875	0.869	0.866	0.884	0.888	0.888
Mfg & Trade Sales	0.95	0.698	0.683	0.694	**0.677**	0.697	0.716	0.722	0.72	0.724	0.727	0.716
Nonag. Employment	0.583	0.565	0.561	0.553	0.53	**0.516**	0.527	0.538	0.545	0.538	0.543	0.533
CPI	0.992	0.909	0.746	**0.745**	0.749	0.763	0.773	0.755	0.749	0.772	0.769	0.759
Consumption Deflator	1	0.957	**0.829**	0.835	0.85	0.873	0.872	0.858	0.846	0.863	0.859	0.843
CPI exc. Food	0.999	0.963	**0.827**	0.839	0.849	0.853	0.856	0.851	0.859	0.869	0.866	0.864
Producer Price Index	1.001	0.937	0.759	0.761	0.754	0.762	0.764	0.763	0.76	0.752	0.747	**0.731**

Table B.15: 12-Month-Ahead DI Forecasts by PCA With Given k, $k = 1, 2, ..., 12$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of PCA factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the eight target variables by PCA, with given k, $k = 1, 2, ..., 12$, is presented. The forecast with the minimum RMSE for the method is in **bold**.

DIAR : PCA	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Industrial Production	0.893	0.714	**0.699**	0.719	0.747	0.729	0.725	0.736	0.732	0.738	0.75	0.737
Personal Income	0.854	**0.842**	0.846	0.869	0.874	0.862	0.874	0.869	0.866	0.884	0.888	0.888
Mfg & Trade Sales	0.95	0.681	**0.665**	0.678	0.676	0.683	0.698	0.71	0.707	0.715	0.717	0.718
Nonag. Employment	0.586	0.515	**0.503**	0.518	0.518	0.504	0.516	0.528	0.525	0.513	0.521	0.511
CPI	0.405	0.41	0.405	**0.395**	0.396	0.405	0.409	0.408	0.407	0.42	0.423	0.424
Consumption Deflator	**0.483**	0.496	0.496	0.485	0.49	0.511	0.516	0.518	0.512	0.529	0.532	0.534
CPI exc. Food	**0.453**	0.463	0.464	0.459	0.464	0.48	0.479	0.478	0.479	0.488	0.489	0.488
Producer Price Index	0.342	0.338	0.336	**0.333**	0.335	0.337	0.341	0.343	0.339	0.343	0.347	0.348

Table B.16: 12-Month-Ahead DIAR Forecasts by PCA With Given k, $k = 1, 2, ..., 12$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of PCA factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the eight target variables by PCA, with given k, $k = 1, 2, ..., 12$, is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DIAR-LAG : PCA	k =1	k =2	k =3	k =4
Industrial Production	0.885	0.716	0.707	**0.685**
Personal Income	0.904	**0.838**	0.845	0.867
Mfg & Trade Sales	0.944	0.659	0.659	**0.653**
Nonag. Employment	0.595	0.522	0.527	**0.507**
CPI	0.406	0.407	0.409	**0.404**
Consumption Deflator	**0.485**	0.502	0.503	0.485
CPI exc. Food	**0.459**	0.47	0.467	0.459
Producer Price Index	0.347	0.341	0.337	**0.333**

Table B.17: 12-Month-Ahead DIAR-LAG Forecasts by PCA With Given k, $k = 1, 2, ..., 4$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of PCA factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the eight target variables by PCA, with given k, $k = 1, 2, ..., 4$, is presented. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DI : PCA	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Overall	0.998	0.947	0.899	0.902	0.905	0.883	**0.878**	0.895	0.901	0.912	0.918	0.909
Output and Income	0.947	0.773	0.767	0.785	0.781	0.764	0.787	0.79	0.783	0.76	0.758	**0.735**
Labor Market	0.921	0.754	0.749	0.751	**0.723**	0.751	0.771	0.773	0.781	0.784	0.777	0.773
Housing	2.004	2.077	2.073	1.955	1.866	1.744	1.622	1.558	1.534	1.362	1.266	**1.235**
Consumption, Orders, Inventories	0.984	0.915	**0.9**	0.913	0.93	0.916	0.91	0.906	0.915	0.925	0.922	0.92
Money and Credit	1.013	**0.995**	1.006	1	1.003	1.006	1.003	1.012	1.01	1.022	1.035	1.036
Interest and Exchange Rates	1.007	0.927	0.93	0.911	0.932	0.895	**0.882**	0.914	0.924	0.958	0.964	0.958
Prices	0.999	0.958	0.87	0.876	0.878	0.88	0.878	0.873	0.88	0.881	0.876	**0.864**
Stock Market	0.923	**0.882**	0.883	0.889	0.895	0.906	0.895	0.889	0.898	0.914	0.918	0.928

Table B.18: 12-Month-Ahead DI Forecasts by PCA With Given k, $k = 1, 2, ..., 12$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the whole 144 target variables by PCA, with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The forecast with the minimum RMSE for the method is in **bold**.

DIAR : PCA	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Overall	0.859	0.757	0.754	**0.725**	0.744	0.741	0.754	0.777	0.773	0.771	0.775	0.767
Output and Income	0.928	0.793	0.783	0.798	0.793	0.779	0.798	0.799	0.793	0.771	0.768	**0.745**
Labor Market	0.763	0.646	**0.631**	0.658	0.656	0.653	0.672	0.69	0.686	0.678	0.689	0.688
Housing	0.956	0.931	**0.921**	0.925	0.929	0.939	0.972	0.956	0.969	0.957	0.945	0.934
Consumption, Orders, Inventories	0.988	0.865	**0.862**	0.873	0.884	0.867	0.874	0.89	0.905	0.906	0.909	0.904
Money and Credit	0.592	**0.516**	0.517	0.522	0.522	0.524	0.534	0.534	0.537	0.543	0.551	0.557
Interest and Exchange Rates	0.93	0.931	0.916	0.916	0.928	0.889	**0.864**	0.903	0.917	0.952	0.955	0.95
Prices	0.404	0.408	0.403	**0.396**	0.4	0.412	0.417	0.416	0.418	0.432	0.44	0.439
Stock Market	0.923	0.917	0.921	0.928	0.93	0.943	0.903	**0.899**	0.902	0.922	0.935	0.934

Table B.19: 12-Month-Ahead DIAR Forecasts by PCA With Given k, $k = 1, 2, ..., 12$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by PCA, with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DIAR-LAG : PCA	k =1	k =2	k =3	k =4
Overall	0.861	0.741	0.751	**0.735**
Output and Income	0.915	**0.788**	0.802	0.79
Labor Market	0.788	**0.635**	0.644	0.643
Housing	0.969	0.937	**0.924**	0.931
Consumption, Orders, Inventories	0.991	**0.87**	0.879	0.88
Money and Credit	0.593	**0.516**	0.517	0.522
Interest and Exchange Rates	0.963	0.937	**0.916**	0.916
Prices	0.405	0.405	0.406	**0.403**
Stock Market	**0.917**	0.918	0.918	0.933

Table B.20: 12-Month-Ahead DIAR-LAG Forecasts by PCA With Given k, $k = 1, 2, ..., 4$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the whole 144 target variables by PCA, with given k, $k = 1, 2, ..., 4$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DI : PCA	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Industrial Production	0.691	0.672	0.668	0.672	0.658	0.678	**0.646**	0.686	0.663	0.67	0.663
Personal Income	0.89	0.876	0.86	0.888	0.826	0.82	**0.771**	0.834	0.867	0.848	0.841
Mfg & Trade Sales	0.689	0.723	0.707	0.716	0.668	0.682	**0.655**	0.689	0.703	0.692	0.677
Nonag. Employment	0.522	0.55	0.526	0.533	0.535	0.545	**0.499**	0.521	0.526	0.529	0.516
CPI	**0.753**	0.759	0.77	0.759	0.786	0.952	0.974	0.784	0.769	0.812	0.745
Consumption Deflator	0.86	0.859	**0.837**	0.843	0.849	0.988	1.005	0.86	0.862	0.885	0.829
CPI exc. Food	0.842	0.861	0.85	0.864	**0.831**	0.998	1.014	0.834	0.869	0.885	0.827
Producer Price Index	0.751	0.764	0.772	**0.731**	0.791	0.976	0.987	0.794	0.757	0.814	0.731

Table B.21: 12-Month-Ahead DI Forecasts by PCA With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the eight target variables by PCA is considered. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.15. The forecast with the minimum RMSE for the method is in **bold**.

DIAR : PCA	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Industrial Production	0.75	0.744	0.713	0.737	0.685	0.687	**0.659**	0.708	0.731	0.713	0.699
Personal Income	0.89	0.876	0.86	0.888	0.826	0.821	**0.772**	0.834	0.867	0.848	0.842
Mfg & Trade Sales	0.694	0.711	0.69	0.718	0.652	0.655	**0.644**	0.674	0.688	0.681	0.665
Nonag. Employment	0.494	0.529	0.506	0.511	**0.484**	0.511	0.5	0.509	0.503	0.505	0.503
CPI	0.407	0.409	**0.406**	0.424	0.41	0.412	0.411	0.406	0.407	0.41	0.395
Consumption Deflator	**0.492**	0.515	0.497	0.534	0.497	0.497	0.497	0.502	0.514	0.505	0.483
CPI exc. Food	0.468	0.48	0.474	0.488	0.468	0.458	**0.454**	0.464	0.478	0.47	0.453
Producer Price Index	0.341	0.339	0.339	0.348	0.339	0.343	0.345	0.341	**0.337**	0.341	0.333

Table B.22: 12-Month-Ahead DIAR Forecasts by PCA With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the eight target variables by PCA is considered. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, is determined by BIC. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.16. The forecast with the minimum RMSE for the method is in **bold**.

DIAR-LAG : PCA	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Industrial Production	0.682	0.685	0.682	0.685	0.726	0.678	**0.652**	0.703	0.706	0.689	0.685
Personal Income	0.844	0.867	0.851	0.867	0.85	0.82	**0.774**	0.836	0.838	0.839	0.838
Mfg & Trade Sales	0.637	0.653	0.643	0.653	0.691	**0.632**	0.634	0.653	0.66	0.651	0.653
Nonag. Employment	0.512	0.508	0.509	0.507	0.522	0.515	**0.475**	0.515	0.522	0.509	0.507
CPI	0.407	**0.404**	0.408	0.404	0.404	0.414	0.414	0.406	0.408	0.408	0.404
Consumption Deflator	0.493	**0.485**	0.486	0.485	0.501	0.507	0.506	0.491	0.495	0.494	0.485
CPI exc. Food	0.463	**0.459**	0.46	0.459	0.469	0.469	0.466	0.466	0.466	0.464	0.459
Producer Price Index	0.345	**0.333**	0.334	0.333	0.341	0.349	0.351	0.337	0.336	0.34	0.333

Table B.23: 12-Month-Ahead DIAR-LAG Forecasts by PCA With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the eight target variables by PCA is considered. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.17. The forecast with the minimum RMSE for the method is in **bold**.

DI : PCA	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Overall	0.891	0.898	**0.886**	0.909	0.901	0.96	0.956	0.896	0.895	0.91	0.878
Output and Income	0.736	0.79	0.767	0.735	0.734	0.738	**0.708**	0.757	0.772	0.749	0.735
Labor Market	0.75	0.792	0.783	0.773	0.743	0.782	0.774	**0.723**	0.765	0.765	0.723
Housing	1.285	1.459	1.562	**1.235**	1.946	1.914	1.921	1.863	1.629	1.646	1.235
Consumption, Orders, Inventories	0.917	0.906	0.893	0.92	0.916	0.902	**0.877**	0.905	0.914	0.906	0.9
Money and Credit	1.012	1.006	1.004	1.036	0.998	0.995	**0.993**	1.007	1.01	1.007	0.995
Interest and Exchange Rates	0.934	**0.926**	0.949	0.958	0.944	0.966	0.998	0.93	0.931	0.948	0.882
Prices	0.877	0.879	0.875	**0.864**	0.882	0.989	0.995	0.889	0.881	0.903	0.864
Stock Market	0.885	0.891	0.914	0.928	0.888	0.892	**0.864**	0.897	0.89	0.894	0.882

Table B.24: 12-Month-Ahead DI Forecasts by PCA With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the whole 144 target variables by PCA is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.18. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : PCA	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Overall	0.75	0.767	0.742	0.767	0.744	0.766	0.756	**0.736**	0.775	0.756	0.725
Output and Income	0.749	0.804	0.787	0.745	0.765	0.761	**0.73**	0.77	0.781	0.766	0.745
Labor Market	0.667	0.687	0.659	0.688	**0.621**	0.666	0.665	0.645	0.661	0.662	0.631
Housing	0.949	0.967	0.99	0.934	0.934	**0.858**	0.875	0.938	0.956	0.933	0.921
Consumption, Orders, Inventories	0.899	0.904	0.864	0.904	0.848	**0.835**	0.838	0.864	0.892	0.872	0.862
Money and Credit	0.54	0.53	0.53	0.557	0.52	**0.515**	0.523	0.522	0.536	0.53	0.516
Interest and Exchange Rates	0.939	0.92	0.926	0.95	0.932	0.923	0.917	0.926	**0.904**	0.926	0.864
Prices	0.425	0.42	0.405	0.439	0.407	0.398	**0.395**	0.405	0.417	0.412	0.396
Stock Market	0.908	0.899	0.944	0.934	0.924	0.915	**0.873**	0.937	0.915	0.917	0.899

Table B.25: 12-Month-Ahead DIAR Forecasts by PCA With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by PCA is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.19. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR-LAG : PCA	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Overall	0.734	0.735	**0.733**	0.735	0.749	0.756	0.756	0.735	0.743	0.742	0.735
Output and Income	0.76	0.789	0.773	0.79	0.802	0.749	**0.722**	0.79	0.796	0.775	0.788
Labor Market	0.646	0.643	0.641	0.643	0.646	0.661	0.677	**0.639**	0.64	0.648	0.635
Housing	0.944	0.931	0.946	0.931	0.928	**0.871**	0.88	0.928	0.929	0.921	0.924
Consumption, Orders, Inventories	0.876	0.88	0.87	0.88	0.888	**0.834**	0.853	0.866	0.876	0.869	0.87
Money and Credit	**0.515**	0.522	0.521	0.522	0.515	0.515	0.526	0.516	0.516	0.519	0.516
Interest and Exchange Rates	0.922	0.916	0.941	0.916	**0.906**	0.933	0.941	0.932	0.923	0.926	0.916
Prices	0.411	**0.403**	0.406	0.403	0.404	0.405	0.404	0.404	0.405	0.405	0.403
Stock Market	0.924	0.933	0.936	0.933	0.919	0.918	**0.872**	0.917	0.918	0.919	0.917

Table B.26: 12-Month-Ahead DIAR-LAG Forecasts by PCA With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the whole 144 target variables by PCA is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.20. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DI : PLS	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Industrial Production	**0.594**	0.662	0.772	0.98	1.083	1.2	1.31	1.472	1.63	1.801	1.952	2.074
Personal Income	**0.681**	0.789	0.846	1.07	1.178	1.285	1.394	1.484	1.581	1.678	1.776	1.841
Mfg & Trade Sales	**0.631**	0.717	0.839	0.989	1.113	1.274	1.424	1.61	1.796	1.955	2.094	2.191
Nonag. Employment	**0.469**	0.535	0.648	0.884	1.038	1.103	1.199	1.338	1.455	1.569	1.683	1.771
CPI	0.762	0.677	0.687	0.635	0.63	**0.629**	0.639	0.643	0.656	0.677	0.703	0.739
Consumption Deflator	0.851	0.757	0.748	**0.721**	0.721	0.739	0.759	0.778	0.794	0.817	0.842	0.869
CPI exc. Food	0.854	0.776	0.777	0.724	**0.709**	0.713	0.725	0.727	0.742	0.761	0.78	0.802
Producer Price Index	0.738	0.651	0.648	0.612	**0.605**	0.618	0.62	0.634	0.651	0.674	0.69	0.712

Table B.27: 12-Month-Ahead DI Forecasts by PLS With Given k, $k = 1, 2, ..., 12$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of PLS factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the eight target variables by PLS, with given k, $k = 1, 2, ..., 12$, is presented. The forecast with the minimum RMSE for the method is in **bold**.

DIAR : PLS	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Industrial Production	**0.638**	0.682	0.838	0.935	1.018	1.129	1.188	1.326	1.467	1.592	1.718	1.818
Personal Income	**0.68**	0.787	0.842	1.065	1.171	1.278	1.385	1.475	1.572	1.668	1.765	1.834
Mfg & Trade Sales	**0.648**	0.722	0.863	0.985	1.116	1.282	1.421	1.606	1.792	1.946	2.088	2.191
Nonag. Employment	**0.498**	0.528	0.673	0.871	1.031	1.096	1.194	1.324	1.439	1.545	1.648	1.74
CPI	**0.392**	0.397	0.412	0.409	0.421	0.412	0.418	0.42	0.429	0.442	0.458	0.48
Consumption Deflator	**0.492**	0.502	0.504	0.507	0.527	0.523	0.551	0.566	0.573	0.594	0.609	0.63
CPI exc. Food	**0.427**	0.441	0.465	0.466	0.477	0.474	0.478	0.475	0.484	0.499	0.513	0.535
Producer Price Index	0.342	0.337	0.353	0.34	0.34	0.335	**0.334**	0.34	0.352	0.36	0.361	0.371

Table B.28: 12-Month-Ahead DIAR Forecasts by PLS With Given k, $k = 1, 2, ..., 12$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of PLS factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the eight target variables by PLS, with given k, $k = 1, 2, ..., 12$, is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DIAR-LAG : PLS	k =1	k =2	k =3	k =4
Industrial Production	**0.636**	0.651	0.791	0.968
Personal Income	**0.671**	0.796	0.858	1.163
Mfg & Trade Sales	**0.646**	0.688	0.884	1.043
Nonag. Employment	**0.494**	0.534	0.684	0.998
CPI	**0.382**	0.382	0.414	0.422
Consumption Deflator	**0.489**	0.499	0.528	0.529
CPI exc. Food	**0.434**	0.441	0.483	0.476
Producer Price Index	0.339	**0.334**	0.359	0.345

Table B.29: 12-Month-Ahead DIAR-LAG Forecasts by PLS With Given k, $k = 1, 2, ..., 4$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of PLS factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the eight target variables by PLS, with given k, $k = 1, 2, ..., 4$, is presented. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DI : PLS	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Overall	0.86	**0.847**	0.973	1.062	1.132	1.19	1.232	1.286	1.328	1.356	1.394	1.434
Output and Income	**0.679**	0.707	0.816	0.978	1.062	1.164	1.272	1.411	1.548	1.684	1.802	1.889
Labor Market	0.705	**0.674**	0.803	0.852	0.911	0.942	0.979	1.026	1.093	1.151	1.192	1.22
Housing	1.435	1.473	1.678	1.289	1.247	1.315	1.294	1.301	1.358	1.283	**1.235**	1.309
Consumption, Orders, Inventories	**0.891**	1.111	1.301	1.479	1.586	1.711	1.928	2.105	2.219	2.349	2.452	2.535
Money and Credit	1.02	**0.995**	1.071	1.137	1.155	1.172	1.202	1.222	1.242	1.258	1.275	1.288
Interest and Exchange Rates	0.956	**0.929**	1.046	1.182	1.262	1.287	1.309	1.358	1.416	1.438	1.446	1.459
Prices	0.85	0.776	0.762	0.723	**0.715**	0.726	0.732	0.739	0.756	0.773	0.786	0.807
Stock Market	0.905	**0.881**	1.022	1.144	1.223	1.288	1.36	1.471	1.568	1.654	1.725	1.775

Table B.30: 12-Month-Ahead DI Forecasts by PLS With Given k, $k = 1, 2, ..., 12$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the whole 144 target variables by PLS, with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The forecast with the minimum RMSE for the method is in **bold**.

DIAR : PLS	k =1	k =2	k =3	k =4	k =5	k =6	k =7	k =8	k =9	k =10	k =11	k =12
Overall	**0.716**	0.746	0.842	0.93	0.968	1.042	1.081	1.119	1.18	1.229	1.265	1.276
Output and Income	**0.685**	0.704	0.814	0.942	1.034	1.127	1.173	1.296	1.446	1.585	1.695	1.811
Labor Market	**0.633**	0.656	0.81	0.848	0.876	0.922	0.908	0.936	0.989	1.036	1.085	1.148
Housing	0.981	0.941	1.054	0.874	**0.853**	0.892	0.907	0.9	0.946	0.976	1.012	1.072
Consumption, Orders, Inventories	**0.862**	1.1	1.268	1.381	1.53	1.704	1.83	1.976	2.087	2.21	2.31	2.357
Money and Credit	0.555	**0.53**	0.553	0.569	0.581	0.615	0.633	0.655	0.676	0.699	0.715	0.74
Interest and Exchange Rates	**0.915**	0.964	1.048	1.177	1.262	1.287	1.309	1.36	1.416	1.438	1.495	1.519
Prices	**0.4**	0.42	0.434	0.437	0.448	0.443	0.448	0.448	0.459	0.465	0.475	0.487
Stock Market	**0.903**	0.905	1.025	1.147	1.225	1.289	1.362	1.471	1.568	1.654	1.725	1.775

Table B.31: 12-Month-Ahead DIAR Forecasts by PLS With Given k, $k = 1, 2, ..., 12$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by PLS, with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DIAR-LAG : PLS	k =1	k =2	k =3	k =4
Overall	**0.709**	0.753	0.861	0.952
Output and Income	0.691	**0.69**	0.821	1.022
Labor Market	**0.637**	0.661	0.827	0.883
Housing	1.022	0.99	1.103	**0.879**
Consumption, Orders, Inventories	**0.869**	1.102	1.284	1.448
Money and Credit	0.538	**0.519**	0.567	0.582
Interest and Exchange Rates	**0.916**	0.983	1.083	1.203
Prices	**0.393**	0.411	0.443	0.444
Stock Market	0.909	**0.907**	1.1	1.252

Table B.32: 12-Month-Ahead DIAR-LAG Forecasts by PLS With Given k, $k = 1, 2, ..., 4$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the whole 144 target variables by PLS, with given k, $k = 1, 2, ..., 4$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The forecast with the minimum RMSE for the method is in **bold**.

DI : PLS	PLS1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Industrial Production	**0.594**	1.375	1.385	0.997	2.074	0.774	0.652	0.628	0.926	1.492	1.09	0.594
Personal Income	**0.681**	1.37	1.505	1.318	1.841	0.927	0.757	0.736	1.169	1.448	1.175	0.681
Mfg & Trade Sales	**0.631**	1.611	1.475	1.062	2.191	0.834	0.693	0.673	0.966	1.618	1.175	0.631
Nonag. Employment	**0.469**	1.176	1.329	0.971	1.771	0.664	0.533	0.513	0.87	1.315	0.961	0.469
CPI	0.762	0.719	0.655	**0.618**	0.739	0.646	0.721	0.737	0.623	0.656	0.688	0.629
Consumption Deflator	0.851	0.825	0.779	**0.697**	0.869	0.716	0.809	0.84	0.719	0.786	0.789	0.721
CPI exc. Food	0.854	0.754	0.73	0.699	0.802	0.729	0.825	0.851	**0.689**	0.74	0.767	0.709
Producer Price Index	0.738	0.67	0.642	**0.609**	0.712	0.621	0.714	0.727	0.619	0.636	0.669	0.605

Table B.33: 12-Month-Ahead DI Forecasts by PLS With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the eight target variables by PLS is considered. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.27. The forecast with the minimum RMSE for the method is in **bold**.

DIAR : PLS	PLS1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Industrial Production	**0.638**	1.326	1.297	1.009	1.818	0.847	0.694	0.677	0.945	1.348	1.06	0.638
Personal Income	**0.68**	1.363	1.496	1.308	1.834	0.922	0.757	0.736	1.161	1.44	1.17	0.68
Mfg & Trade Sales	**0.648**	1.64	1.473	1.067	2.191	0.861	0.717	0.702	0.98	1.615	1.189	0.648
Nonag. Employment	**0.498**	1.178	1.326	0.971	1.74	0.67	0.531	0.545	0.877	1.297	0.963	0.498
CPI	**0.392**	0.436	0.426	0.405	0.48	0.398	0.395	0.393	0.413	0.427	0.417	0.392
Consumption Deflator	**0.492**	0.595	0.57	0.511	0.63	0.496	0.504	0.503	0.528	0.575	0.54	0.492
CPI exc. Food	**0.427**	0.515	0.486	0.459	0.535	0.454	0.437	0.437	0.461	0.488	0.47	0.427
Producer Price Index	0.342	0.345	0.346	0.337	0.371	0.34	0.335	**0.334**	0.344	0.348	0.344	0.334

Table B.34: 12-Month-Ahead DIAR Forecasts by PLS With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the eight target variables by PLS is considered. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, is determined by BIC. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.28. The forecast with the minimum RMSE for the method is in **bold**.

DIAR-LAG : PLS	PLS1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Industrial Production	0.636	0.931	0.968	0.93	0.968	**0.606**	0.675	0.658	0.762	0.788	0.792	0.636
Personal Income	**0.671**	1.133	1.163	1.12	1.163	0.789	0.76	0.73	0.88	0.868	0.928	0.671
Mfg & Trade Sales	0.646	1.021	1.042	0.94	1.043	**0.637**	0.698	0.688	0.808	0.852	0.837	0.646
Nonag. Employment	**0.494**	0.919	0.998	0.915	0.998	0.502	0.529	0.539	0.653	0.667	0.721	0.494
CPI	0.382	0.422	0.422	0.411	0.422	**0.373**	0.389	0.392	0.409	0.42	0.404	0.382
Consumption Deflator	0.489	0.534	0.529	0.511	0.529	**0.488**	0.509	0.511	0.522	0.53	0.515	0.489
CPI exc. Food	0.434	0.479	0.476	0.473	0.476	**0.433**	0.448	0.45	0.478	0.481	0.463	0.434
Producer Price Index	0.339	0.344	0.345	0.346	0.345	**0.337**	0.34	0.341	0.356	0.356	0.345	0.334

Table B.35: 12-Month-Ahead DIAR-LAG Forecasts by PLS With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the eight target variables by PLS is considered. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.29. The forecast with the minimum RMSE for the method is in **bold**.

DI : PLS	PLS1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Overall	**0.86**	1.316	1.293	1.147	1.434	0.949	0.873	0.873	1.085	1.276	1.111	0.847
Output and Income	**0.679**	1.355	1.32	1.01	1.889	0.81	0.705	0.696	0.932	1.418	1.081	0.679
Labor Market	0.705	1.105	1.034	0.896	1.22	0.794	**0.682**	0.689	0.892	1.052	0.907	0.674
Housing	1.435	**1.246**	1.333	1.397	1.309	1.453	1.59	1.487	1.326	1.28	1.386	1.235
Consumption, Orders, Inventories	**0.891**	1.802	1.995	1.733	2.535	1.267	1.026	1.021	1.511	2.058	1.584	0.891
Money and Credit	1.02	1.184	1.216	1.179	1.288	1.068	**0.995**	1.006	1.147	1.236	1.134	0.995
Interest and Exchange Rates	0.956	1.463	1.365	1.278	1.459	1.087	0.933	**0.9**	1.251	1.382	1.207	0.929
Prices	0.85	0.761	0.748	**0.698**	0.807	0.723	0.824	0.842	0.716	0.75	0.772	0.715
Stock Market	0.905	1.703	1.496	1.379	1.775	1.062	0.891	**0.883**	1.189	1.419	1.27	0.881

Table B.36: 12-Month-Ahead DI Forecasts by PLS With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the whole 144 target variables by PCA is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.30. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR : PLS	PLS1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Overall	**0.716**	1.118	1.125	0.999	1.276	0.857	0.748	0.74	0.93	1.13	0.964	0.716
Output and Income	**0.685**	1.327	1.28	1.014	1.811	0.827	0.713	0.708	0.937	1.321	1.062	0.685
Labor Market	**0.633**	1.016	0.968	0.905	1.148	0.813	0.659	0.644	0.869	0.927	0.858	0.633
Housing	0.981	0.975	0.937	0.993	1.072	0.927	0.97	0.983	**0.905**	0.922	0.966	0.853
Consumption, Orders, Inventories	**0.862**	1.763	1.946	1.664	2.357	1.255	1.045	1.048	1.454	1.949	1.534	0.862
Money and Credit	0.555	0.677	0.643	0.594	0.74	0.547	**0.534**	0.554	0.597	0.663	0.61	0.53
Interest and Exchange Rates	**0.915**	1.463	1.37	1.278	1.519	1.086	0.96	0.916	1.249	1.382	1.214	0.915
Prices	**0.4**	0.463	0.454	0.433	0.487	0.426	0.416	0.412	0.437	0.453	0.438	0.4
Stock Market	0.903	1.703	1.496	1.383	1.775	1.065	0.913	**0.894**	1.188	1.419	1.274	0.903

Table B.37: 12-Month-Ahead DIAR Forecasts by PLS With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by PCA is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.31. The lag of target variables, p, is determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

DIAR-LAG : PLS	PLS1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$	Mean	Best k
Overall	**0.709**	0.952	0.952	0.928	0.952	0.75	0.752	0.746	0.828	0.849	0.842	0.709
Output and Income	0.691	0.976	1.021	0.928	1.022	**0.671**	0.705	0.704	0.792	0.815	0.832	0.69
Labor Market	0.637	0.885	0.884	0.872	0.883	0.653	0.651	**0.632**	0.754	0.803	0.765	0.637
Housing	1.022	0.916	0.88	0.935	**0.879**	1.066	1.028	1.022	1.165	1.135	1.005	0.879
Consumption, Orders, Inventories	**0.869**	1.431	1.447	1.39	1.448	1.057	1.052	1.048	1.271	1.276	1.229	0.869
Money and Credit	0.538	0.581	0.584	0.589	0.582	**0.51**	0.528	0.551	0.566	0.56	0.559	0.519
Interest and Exchange Rates	0.916	1.184	1.204	1.145	1.203	0.961	0.969	**0.91**	1.064	1.073	1.063	0.916
Prices	**0.393**	0.441	0.444	0.437	0.444	0.403	0.407	0.41	0.436	0.442	0.426	0.393
Stock Market	0.909	1.253	1.252	1.205	1.252	0.918	0.912	**0.896**	1.022	1.065	1.068	0.907

Table B.38: 12-Month-Ahead DIAR-LAG Forecasts by PLS With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the whole 144 target variables by PCA is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The Mean column is the mean of the method over 9 information criteria. Best k is the best results in **bold** in Table B.32. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. The forecast with the minimum RMSE for corresponding factor estimation method is in **bold**.

	PLS 1	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5	PCA 6	PCA 7	PCA 8	PCA 9	PCA 10	PCA 11	PCA 12
Industrial Production	***0.594***	0.893	0.699	0.684	0.7	0.703	0.666	0.664	0.669	**0.663**	0.668	0.68	0.672
Personal Income	***0.681***	0.845	**0.841**	0.846	0.869	0.873	0.862	0.875	0.869	0.866	0.884	0.888	0.888
Mfg & Trade Sales	***0.631***	0.95	0.698	0.683	0.694	**0.677**	0.697	0.716	0.722	0.72	0.724	0.727	0.716
Nonag. Employment	***0.469***	0.583	0.565	0.561	0.553	0.53	**0.516**	0.527	0.538	0.545	0.538	0.543	0.533
CPI	0.762	0.992	0.909	0.746	**0.745**	0.749	0.763	0.773	0.755	0.749	0.772	0.769	0.759
Consumption Deflator	0.851	1	0.957	**0.829**	0.835	0.85	0.873	0.872	0.858	0.846	0.863	0.859	0.843
CPI exc. Food	0.854	0.999	0.963	**0.827**	0.839	0.849	0.853	0.856	0.851	0.859	0.869	0.866	0.864
Producer Price Index	0.738	1.001	0.937	0.759	0.761	0.754	0.762	0.764	0.763	0.76	0.752	0.747	**0.731**

Table B.39: 12-Month-Ahead DI Forecasts by PLS 1 and PCA With Given k, $k = 1, 2, ..., 12$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the eight target variables by PLS 1 and PCA, with given k, $k = 1, 2, ..., 12$, is presented. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

	PLS 1	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5	PCA 6	PCA 7	PCA 8	PCA 9	PCA 10	PCA 11	PCA 12
Industrial Production	***0.638***	0.893	0.714	**0.699**	0.719	0.747	0.729	0.725	0.736	0.732	0.738	0.75	0.737
Personal Income	***0.68***	0.854	**0.842**	0.846	0.869	0.874	0.862	0.874	0.869	0.866	0.884	0.888	0.888
Mfg & Trade Sales	***0.648***	0.95	0.681	**0.665**	0.678	0.676	0.683	0.698	0.71	0.707	0.715	0.717	0.718
Nonag. Employment	***0.498***	0.586	0.515	**0.503**	0.518	0.518	0.504	0.516	0.528	0.525	0.513	0.521	0.511
CPI	***0.392***	0.405	0.41	0.405	**0.395**	0.396	0.405	0.409	0.408	0.407	0.42	0.423	0.424
Consumption Deflator	0.492	**0.483**	0.496	0.496	0.485	0.49	0.511	0.516	0.518	0.512	0.529	0.532	0.534
CPI exc. Food	***0.427***	**0.453**	0.463	0.464	0.459	0.464	0.48	0.479	0.478	0.479	0.488	0.489	0.488
Producer Price Index	0.342	0.342	0.338	0.336	**0.333**	0.335	0.337	0.341	0.343	0.339	0.343	0.347	0.348

Table B.40: 12-Month-Ahead DIAR Forecasts by PLS 1 and PCA With Given k, $k = 1, 2, ..., 12$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the eight target variables by PLS 1 and PCA, with given k, $k = 1, 2, ..., 12$, is presented. The lag of target variables, p, is determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

	PLS 1	PCA 1	PCA 2	PCA 3	PCA 4
Industrial Production	***0.636***	0.885	0.716	0.707	**0.685**
Personal Income	***0.671***	0.904	**0.838**	0.845	0.867
Mfg & Trade Sales	***0.646***	0.944	0.659	0.659	**0.653**
Nonag. Employment	***0.494***	0.595	0.522	0.527	**0.507**
CPI	***0.382***	0.406	0.407	0.409	**0.404**
Consumption Deflator	0.489	**0.485**	0.502	0.503	0.485
CPI exc. Food	***0.434***	**0.459**	0.47	0.467	0.459
Producer Price Index	0.339	0.347	0.341	0.337	**0.333**

Table B.41: 12-Month-Ahead DIAR-LAG Forecasts by PLS 1 and PCA With Given k, $k = 1, 2, ..., 4$: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the eight target variables by PLS 1 and PCA, with given k, $k = 1, 2, ..., 4$, is presented. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DI : PLS 1 & PCA *k*	PLS 1	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5	PCA 6	PCA 7	PCA 8	PCA 9	PCA 10	PCA 11	PCA 12
Overall	***0.86***	0.998	0.947	0.899	0.902	0.905	0.883	**0.878**	0.895	0.901	0.912	0.918	0.909
Output and Income	***0.679***	0.947	0.773	0.767	0.785	0.781	0.764	0.787	0.79	0.783	0.76	0.758	**0.735**
Labor Market	***0.705***	0.921	0.754	0.749	0.751	**0.723**	0.751	0.771	0.773	0.781	0.784	0.777	0.773
Housing	1.435	2.004	2.077	2.073	1.955	1.866	1.744	1.622	1.558	1.534	1.362	1.266	**1.235**
Consumption, Orders, Inventories	***0.891***	0.984	0.915	**0.9**	0.913	0.93	0.916	0.91	0.906	0.915	0.925	0.922	0.92
Money and Credit	1.02	1.013	**0.995**	1.006	1	1.003	1.006	1.003	1.012	1.01	1.022	1.035	1.036
Interest and Exchange Rates	0.956	1.007	0.927	0.93	0.911	0.932	0.895	**0.882**	0.914	0.924	0.958	0.964	0.958
Prices	***0.85***	0.999	0.958	0.87	0.876	0.878	0.88	0.878	0.873	0.88	0.881	0.876	**0.864**
Stock Market	0.905	0.923	**0.882**	0.883	0.889	0.895	0.906	0.895	0.889	0.898	0.914	0.918	0.928

Table B.42: 12-Month-Ahead DI Forecasts by PLS 1 and PCA With Given k, $k = 1, 2, ..., 12$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the whole 144 target variables by PLS 1 and PCA with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DIAR : PLS 1 & PCA *k*	PLS 1	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5	PCA 6	PCA 7	PCA 8	PCA 9	PCA 10	PCA 11	PCA 12
Overall	🟩 ***0.716***	0.859	0.757	0.754	**0.725**	0.744	0.741	0.754	0.777	0.773	0.771	0.775	0.767
Output and Income	🟩 ***0.685***	0.928	0.793	0.783	0.798	0.793	0.779	0.798	0.799	0.793	0.771	0.768	**0.745**
Labor Market	0.633	0.763	0.646	**0.631**	0.658	0.656	0.653	0.672	0.69	0.686	0.678	0.689	0.688
Housing	0.981	0.956	0.931	**0.921**	0.925	0.929	0.939	0.972	0.956	0.969	0.957	0.945	0.934
Consumption, Orders, Inventories	🟩 ***0.862***	0.988	0.865	**0.862**	0.873	0.884	0.867	0.874	0.89	0.905	0.906	0.909	0.904
Money and Credit	0.555	0.592	**0.516**	0.517	0.522	0.522	0.524	0.534	0.534	0.537	0.543	0.551	0.557
Interest and Exchange Rates	0.915	0.93	0.931	0.916	0.916	0.928	0.889	**0.864**	0.903	0.917	0.952	0.955	0.95
Prices	0.4	0.404	0.408	0.403	**0.396**	0.4	0.412	0.417	0.416	0.418	0.432	0.44	0.439
Stock Market	0.903	0.923	0.917	0.921	0.928	0.93	0.943	0.903	**0.899**	0.902	0.922	0.935	0.934

Table B.43: 12-Month-Ahead DIAR Forecasts by PLS 1 and PCA With Given k, $k = 1, 2, ..., 12$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by PLS 1 and PCA with given k, $k = 1, 2, ..., 12$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, is determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DIAR-LAG : PLS 1 & PCA *k*	PLS 1	PCA 1	PCA 2	PCA 3	PCA 4
Overall	***0.709***	0.861	0.741	0.751	**0.735**
Output and Income	***0.691***	0.915	**0.788**	0.802	0.79
Labor Market	0.637	0.788	**0.635**	0.644	0.643
Housing	1.022	0.969	0.937	**0.924**	0.931
Consumption, Orders, Inventories	***0.869***	0.991	**0.87**	0.879	0.88
Money and Credit	0.538	0.593	**0.516**	0.517	0.522
Interest and Exchange Rates	***0.916***	0.963	0.937	**0.916**	0.916
Prices	***0.393***	0.405	0.405	0.406	**0.403**
Stock Market	***0.909***	**0.917**	0.918	0.918	0.933

Table B.44: 12-Month-Ahead DIAR-LAG Forecasts by PLS 1 and PCA With Given k, $k = 1, 2, ..., 4$: The Whole 144 Variables by Categories

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the whole 144 target variables by PLS 1 and PCA with given k, $k = 1, 2, ..., 4$, is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DI : PLS 1 & PCA by IC	**PLS**	**PCA**								
	PLS 1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$
Industrial Production	***0.594***	0.691	0.672	0.668	0.672	0.658	0.678	**0.646**	0.686	0.663
Personal Income	***0.681***	0.89	0.876	0.86	0.888	0.826	0.82	**0.771**	0.834	0.867
Mfg & Trade Sales	***0.631***	0.689	0.723	0.707	0.716	0.668	0.682	**0.655**	0.689	0.703
Nonag. Employment	***0.469***	0.522	0.55	0.526	0.533	0.535	0.545	**0.499**	0.521	0.526
CPI	0.762	**0.753**	0.759	0.77	0.759	0.786	0.952	0.974	0.784	0.769
Consumption Deflator	0.851	0.86	0.859	**0.837**	0.843	0.849	0.988	1.005	0.86	0.862
CPI exc. Food	0.854	0.842	0.861	0.85	0.864	**0.831**	0.998	1.014	0.834	0.869
Producer Price Index	0.738	0.751	0.764	0.772	**0.731**	0.791	0.976	0.987	0.794	0.757

Table B.45: 12-Month-Ahead DI Forecasts by PLS 1 and PCA With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the eight target variables by PLS 1 and PCA with information criteria is considered. For PCA, the number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DIAR : PLS 1 & PCA by IC	**PLS**	**PCA**								
	PLS 1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$
Industrial Production	***0.638***	0.75	0.744	0.713	0.737	0.685	0.687	**0.659**	0.708	0.731
Personal Income	***0.68***	0.89	0.876	0.86	0.888	0.826	0.821	**0.772**	0.834	0.867
Mfg & Trade Sales	0.648	0.694	0.711	0.69	0.718	0.652	0.655	**0.644**	0.674	0.688
Nonag. Employment	0.498	0.494	0.529	0.506	0.511	**0.484**	0.511	0.5	0.509	0.503
CPI	***0.392***	0.407	0.409	**0.406**	0.424	0.41	0.412	0.411	0.406	0.407
Consumption Deflator	***0.492***	**0.492**	0.515	0.497	0.534	0.497	0.497	0.497	0.502	0.514
CPI exc. Food	***0.427***	0.468	0.48	0.474	0.488	0.468	0.458	**0.454**	0.464	0.478
Producer Price Index	0.342	0.341	0.339	0.339	0.348	0.339	0.343	0.345	0.341	**0.337**

Table B.46: 12-Month-Ahead DIAR Forecasts by PLS 1 and PCA With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the eight target variables by PLS 1 and PCA with information criteria is considered. For PCA, the number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, is determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DIAR-LAG : PLS 1 & PCA by IC	**PLS**	**PCA**								
	PLS 1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$
Industrial Production	***0.636***	0.682	0.685	0.682	0.685	0.726	0.678	**0.652**	0.703	0.706
Personal Income	***0.671***	0.844	0.867	0.851	0.867	0.85	0.82	**0.774**	0.836	0.838
Mfg & Trade Sales	0.646	0.637	0.653	0.643	0.653	0.691	**0.632**	0.634	0.653	0.66
Nonag. Employment	0.494	0.512	0.508	0.509	0.507	0.522	0.515	**0.475**	0.515	0.522
CPI	***0.382***	0.407	**0.404**	0.408	0.404	0.404	0.414	0.414	0.406	0.408
Consumption Deflator	0.489	0.493	**0.485**	0.486	0.485	0.501	0.507	0.506	0.491	0.495
CPI exc. Food	***0.434***	0.463	**0.459**	0.46	0.459	0.469	0.469	0.466	0.466	0.466
Producer Price Index	0.339	0.345	**0.333**	0.334	0.333	0.341	0.349	0.351	0.337	0.336

Table B.47: 12-Month-Ahead DIAR-LAG Forecasts by PLS 1 and PCA With Information Criteria: 8 Target Variables

Notes: The entries are relative mean squared errors (RMSE) of respective factor estimation method relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the eight target variables by PLS 1 and PCA with information criteria is considered. For PCA, the number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DI : **PLS 1 & PCA by IC**	**PLS**	**PCA**								
	PLS 1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$
Overall	***0.86***	0.891	0.898	**0.886**	0.909	0.901	0.96	0.956	0.896	0.895
Output and Income	***0.679***	0.736	0.79	0.767	0.735	0.734	0.738	**0.708**	0.757	0.772
Labor Market	***0.705***	0.75	0.792	0.783	0.773	0.743	0.782	0.774	**0.723**	0.765
Housing	1.435	1.285	1.459	1.562	**1.235**	1.946	1.914	1.921	1.863	1.629
Consumption, Orders, Inventories	0.891	0.917	0.906	0.893	0.92	0.916	0.902	**0.877**	0.905	0.914
Money and Credit	1.02	1.012	1.006	1.004	1.036	0.998	0.995	**0.993**	1.007	1.01
Interest and Exchange Rates	0.956	0.934	**0.926**	0.949	0.958	0.944	0.966	0.998	0.93	0.931
Prices	***0.85***	0.877	0.879	0.875	**0.864**	0.882	0.989	0.995	0.889	0.881
Stock Market	0.905	0.885	0.891	0.914	0.928	0.888	0.892	**0.864**	0.897	0.89

Table B.48: 12-Month-Ahead DI Forecasts by PLS 1 and PCA With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DI forecasting for the whole 144 target variables by PLS 1 and PCA with information criteria is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. For PCA, the number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DIAR : PLS 1 & PCA by IC	**PLS**	**PCA**								
	PLS 1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$
Overall	***0.716***	0.75	0.767	0.742	0.767	0.744	0.766	0.756	**0.736**	0.775
Output and Income	***0.685***	0.749	0.804	0.787	0.745	0.765	0.761	**0.73**	0.77	0.781
Labor Market	0.633	0.667	0.687	0.659	0.688	**0.621**	0.666	0.665	0.645	0.661
Housing	0.981	0.949	0.967	0.99	0.934	0.934	**0.858**	0.875	0.938	0.956
Consumption, Orders, Inventories	0.862	0.899	0.904	0.864	0.904	0.848	**0.835**	0.838	0.864	0.892
Money and Credit	0.555	0.54	0.53	0.53	0.557	0.52	**0.515**	0.523	0.522	0.536
Interest and Exchange Rates	0.915	0.939	0.92	0.926	0.95	0.932	0.923	0.917	0.926	**0.904**
Prices	0.4	0.425	0.42	0.405	0.439	0.407	0.398	**0.395**	0.405	0.417
Stock Market	0.903	0.908	0.899	0.944	0.934	0.924	0.915	**0.873**	0.937	0.915

Table B.49: 12-Month-Ahead DIAR Forecasts by PLS 1 and PCA With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR forecasting for the whole 144 target variables by PLS 1 and PCA with information criteria is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. For PCA, the number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, is determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

DIAR-LAG : PLS 1 & PCA by IC	**PLS**	**PCA**								
	PLS 1	$\hat{k}_{BIC}$	$\hat{k}_{BN-p1}$	$\hat{k}_{BN-p2}$	$\hat{k}_{BN-p3}$	$\hat{k}_{BN-BIC}$	$\hat{k}_{AH}$	$\hat{k}_{ON}$	$\hat{k}_{ABC-L}$	$\hat{k}_{ABC-S}$
Overall	***0.709***	0.734	0.735	**0.733**	0.735	0.749	0.756	0.756	0.735	0.743
Output and Income	***0.691***	0.76	0.789	0.773	0.79	0.802	0.749	**0.722**	0.79	0.796
Labor Market	***0.637***	0.646	0.643	0.641	0.643	0.646	0.661	0.677	**0.639**	0.64
Housing	1.022	0.944	0.931	0.946	0.931	0.928	**0.871**	0.88	0.928	0.929
Consumption, Orders, Inventories	0.869	0.876	0.88	0.87	0.88	0.888	**0.834**	0.853	0.866	0.876
Money and Credit	0.538	**0.515**	0.522	0.521	0.522	0.515	0.515	0.526	0.516	0.516
Interest and Exchange Rates	0.916	0.922	0.916	0.941	0.916	**0.906**	0.933	0.941	0.932	0.923
Prices	***0.393***	0.411	**0.403**	0.406	0.403	0.404	0.405	0.404	0.404	0.405
Stock Market	0.909	0.924	0.933	0.936	0.933	0.919	0.918	**0.872**	0.917	0.918

Table B.50: 12-Month-Ahead DIAR-LAG Forecasts by PLS 1 and PCA With Information Criteria: Whole 144 Target Variables by Category

Notes: The entries are median relative mean squared errors (RMSE) in the 8 categories: 1. Output and Income, 2. Labor Market, 3. Housing, 4. Consumption, Orders, Inventories, 5. Money and Credit, 6. Interest and Exchange Rates, 7. Prices, 8. Stock Market. The Overall row is the median RMSE across all the 144 target variables. RMSE is defined relative to a forecast based on the target variable's historical mean. This experiment is recursive out-of-sample forecast. 12-month-ahead DIAR-LAG forecasting for the whole 144 target variables by PLS 1 and PCA with information criteria is considered. The 144 forecasts are divided into eight categories according to the target variable and the median RMSE of each category is presented. For PCA, the number of contemporaneous factors, k is determined by 9 information criteria, $\hat{k}_{BIC}$, to $\hat{k}_{ABC-S}$. The lag of target variables, p, and the lag of the factors, m, are determined by BIC. If PLS 1 gives lower RMSE than the best PCA forecast in **bold**, the PLS forecast is shaded with green, ***underlined and in bold italic***.

	DI		DIAR		DIAR-LAG	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
PLS 1	199	45.64%	184	42.2%	193	44.27%
PCA 1	47	10.78%	61	13.99%	69	15.83%
PCA 2	13	2.98%	18	4.13%	43	9.86%
PCA 3	26	5.96%	40	9.17%	61	13.99%
PCA 4	37	8.49%	43	9.86%	70	16.06%
PCA 5	16	3.67%	19	4.36%		
PCA 6	11	2.52%	21	4.82%		
PCA 7	23	5.28%	20	4.59%		
PCA 8	8	1.83%	5	1.15%		
PCA 9	4	0.92%	10	2.29%		
PCA 10	1	0.23%	0	0%		
PCA 11	11	2.52%	0	0%		
PCA 12	40	9.17%	15	3.44%		

Table B.51: PLS & PCA k: Whole 148 Variables, 6-,12- and 24-month-ahead DI, DIAR, DIAR-LAG forecast

Notes: PLS 1 and PCA k, $k = 1, 2, ..., 12$, for the whole 148 target variables in 6-, 12- and 24-month ahead DI, DIAR and DIAR-LAG forecasting are considered. For DIAR-LAG, the maximum k is 4, so $k = 1, 2, ..., 4$ are considered. DIAR-LAG PCA k forecasts with $4 < k$ are left empty. The frequency is how many times the given method achieved the minimum RMSE in 436 variable-horizon combinations. Forecast results for 148 target variables (144 for 6- and 12-month ahead forecasts) are counted. The percentage is the frequency divided by 436.

	DI		DIAR		DIAR-LAG	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
PLS 1	202	46.33%	183	41.97%	161	36.93%
$PCA : BIC$	26	5.96%	18	4.13%	28	6.42%
$PCA : BN - p1$	11	2.52%	11	2.52%	44	10.09%
$PCA : BN - p2$	12	2.75%	9	2.06%	26	5.96%
$PCA : BN - p3$	48	11.01%	8	1.83%	3	0.69%
$PCA : BN - BIC$	37	8.49%	67	15.37%	29	6.65%
$PCA : AH$	15	3.44%	30	6.88%	34	7.8%
$PCA : ON$	41	9.4%	61	13.99%	58	13.3%
$PCA : ABC - L$	23	5.28%	32	7.34%	31	7.11%
$PCA : ABC - S$	21	4.82%	17	3.9%	22	5.05%

Table B.52: PLS & PCA by Information Criteria: Whole 148 Variables, 6-,12- and 24-month-ahead DI, DIAR, DIAR-LAG forecast

Notes: PLS 1 and PCA by information criteria, for the whole 148 target variables in 6-, 12- and 24-month ahead DI, DIAR and DIAR-LAG forecasting are considered. The frequency is how many times the given method achieved the minimum RMSE in 436 variable-horizon combinations. Forecast results for 148 target variables (144 for 6- and 12-month ahead forecasts) are counted. The percentage is the frequency divided by 436.

B.2 Data Appendix

The variables used in this study and their categories are presented here. The tcode column denotes the transformation type. Denote the time-series at t as x_t. 1 = no transformation, 2 = first difference, 3 = second difference, 4 = logarithm, 5 = first difference of logarithms 6 = second difference of logarithms, 7 = $\Delta(x_t/x_{t-1} - 1)$. The following abbrevation is used to denote seasonality of a series. SA = Seasonally Adjusted, SSAR = Seasonally Adjusted Annual Rate, NSA = Not Seasonally Adjusted. The data is taken from FRED, FRED-MD (McCracken and Ng (2016)) and ISM. The source column indicates where the variable is taken from. If a variable belongs to both FRED and FRED-MD, the source is denoted as FRED. If a variable in FRED-MD is adjustmented from the raw data available in FRED, the source is denoted as FRED-MD.

ID	Variable	tcode	Description	Period	Seasonality	Source
1	RPI	5	Real Personal Income	1959:01-2019:12	SAAR	FRED
2	W875RX1	5	Real personal income excluding current transfer receipts	1959:01-2019:12	SAAR	FRED
6	INDPRO	5	Industrial Production Index	1959:01-2019:12	SA	FRED
7	IPFPNSS	5	Industrial Production: Final Products and Nonindustrial Supplies	1959:01-2019:12	SA	FRED
8	IPFINAL	5	Industrial Production: Final Products (Market Group)	1959:01-2019:12	SA	FRED
9	IPCONGD	5	Industrial Production: Consumer Goods	1959:01-2019:12	SA	FRED
10	IPDCONGD	5	Industrial Production: Durable Consumer Goods	1959:01-2019:12	SA	FRED
11	IPNCONGD	5	Industrial Production: Nondurable Consumer Goods	1959:01-2019:12	SA	FRED
12	IPBUSEQ	5	Industrial Production: Business Equipment	1959:01-2019:12	SA	FRED
13	IPMAT	5	Industrial Production: Materials	1959:01-2019:12	SA	FRED
14	IPDMAT	5	Industrial Production: Durable Materials	1959:01-2019:12	SA	FRED
15	IPNMAT	5	Industrial Production: Nondurable Materials	1959:01-2019:12	SA	FRED
16	IPMANSICS	5	Industrial Production: Manufacturing (SIC)	1959:01-2019:12	SA	FRED
17	IPB51222S	5	Industrial Production: Residential utilities	1959:01-2019:12	SA	FRED
18	IPFUELS	5	Industrial Production: Fuels	1959:01-2019:12	SA	FRED
19	CUMFNS	2	Capacity Utilization: Manufacturing (SIC)	1959:01-2019:12	SA	FRED
129	IPNMAN	5	Industrial Production: Nondurable Manufacturing (NAICS)	1972:01-2019:12	SA	FRED
130	IPDMAN	5	Industrial Production: Durable Manufacturing (NAICS)	1972:01-2019:12	SA	FRED
131	IPMINE	5	Industrial Production: Mining	1959:01-2019:12	SA	FRED
132	TCU	1	Capacity Utilization: Total Industry	1967:01-2019:12	SA	FRED
133	CAPUTLGMFDS	1	Capacity Utilization: Durable manufacturing	1967:01-2019:12	SA	FRED
134	CAPUTLGMFNS	1	Capacity Utilization: Nondurable manufacturing	1967:01-2019:12	SA	FRED
135	CAPUTLG21S	1	Capacity Utilization: Mining	1967:01-2019:12	SA	FRED
136	CAPUTLG2211A2S	1	Capacity Utilization: Electric and gas utilities	1967:01-2019:12	SA	FRED
173	ISM \MAN_PROD	1	Manufacturing Production Index	1959:01-2019:12	SA	ISM

Table B.53: Category 1. Output and income

ID	Variable	tcode	Description	Period	Seasonality	Source
20	HWI	2	Help-Wanted Index for United States	1959:01-2019:12		FRED-MD
21	HWIURATIO	2	Ratio of Help Wanted/No. Unemployed	1959:01-2019:12		FRED-MD
22	CLF16OV	5	Civilian Labor Force Level	1959:01-2019:12	SA	FRED
23	CE16OV	5	Employment Level	1959:01-2019:12	SA	FRED
24	UNRATE	2	Unemployment Rate	1959:01-2019:12	SA	FRED
25	UEMPMEAN	2	Average Weeks Unemployed	1959:01-2019:12	SA	FRED
26	UEMPLT5	5	Number Unemployed for Less Than 5 Weeks	1959:01-2019:12	SA	FRED
27	UEMP5TO14	5	Number Unemployed for 5-14 Weeks	1959:01-2019:12	SA	FRED
28	UEMP15OV	5	Number Unemployed for 15 Weeks & Over	1959:01-2019:12	SA	FRED
29	UEMP15T26	5	Number Unemployed for 15-26 Weeks	1959:01-2019:12	SA	FRED
30	UEMP27OV	5	Number Unemployed for 27 Weeks & Over	1959:01-2019:12	SA	FRED
31	CLAIMSx	5	Initial Claims	1959:01-2019:12		FRED-MD
32	PAYEMS	5	All Employees, Total Nonfarm	1959:01-2019:12	SA	FRED
33	USGOOD	5	All Employees, Goods-Producing	1959:01-2019:12	SA	FRED
34	CES1021000001	5	All Employees, Mining	1959:01-2019:12	SA	FRED
35	USCONS	5	All Employees, Construction	1959:01-2019:12	SA	FRED
36	MANEMP	5	All Employees, Manufacturing	1959:01-2019:12	SA	FRED
37	DMANEMP	5	All Employees, Durable Goods	1959:01-2019:12	SA	FRED
38	NDMANEMP	5	All Employees, Nondurable Goods	1959:01-2019:12	SA	FRED
39	SRVPRD	5	All Employees, Service-Providing	1959:01-2019:12	SA	FRED
40	USTPU	5	All Employees, Trade, Transportation, and Utilities	1959:01-2019:12	SA	FRED
41	USWTRADE	5	All Employees, Wholesale Trade	1959:01-2019:12	SA	FRED
42	USTRADE	5	All Employees, Retail Trade	1959:01-2019:12	SA	FRED
43	USFIRE	5	All Employees, Financial Activities	1959:01-2019:12	SA	FRED
44	USGOVT	5	All Employees, Government	1959:01-2019:12	SA	FRED
45	CES0600000007	1	Average Weekly Hours of Production and Nonsupervisory Employees, Goods-Producing	1959:01-2019:12	SA	FRED
46	AWOTMAN	2	Average Weekly Overtime Hours of Production and Nonsupervisory Employees, Manufacturing	1959:01-2019:12	SA	FRED
47	AWHMAN	1	Average Weekly Hours of Production and Nonsupervisory Employees, Manufacturing	1959:01-2019:12	SA	FRED
120	CES0600000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Goods-Producing	1959:01-2019:12	SA	FRED
121	CES2000000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Construction	1959:01-2019:12	SA	FRED
122	CES3000000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Manufacturing	1959:01-2019:12	SA	FRED
137	USPRIV	5	All Employees, Total Private	1959:01-2019:12	SA	FRED
138	CES5552000001	5	All Employees, Finance and Insurance	1990:01-2019:12	SA	FRED
139	CES5553100001	5	All Employees, Real Estate	1990:01-2019:12	SA	FRED
140	SRVPRD	5	All Employees, Service-Providing	1959:01-2019:12	SA	FRED
141	AWHNONAG	1	Average Weekly Hours of Production and Nonsupervisory Employees, Total Private	1964:01-2019:12	SA	FRED
166	AHETPI	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Total Private	1964:01-2019:12	SA	FRED
167	CES4000000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Trade, Transportation, and Utilities	1964:01-2019:12	SA	FRED
168	CES4200000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Retail Trade	1972:01-2019:12	SA	FRED
169	CES4142000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Wholesale Trade	1972:01-2019:12	SA	FRED
170	CES5500000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Financial Activities	1964:01-2019:12	SA	FRED
171	CES0800000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees, Private Service-Providing	1964:01-2019:12	SA	FRED
174	ISM \MAN_EMPL	1	Manufacturing Employment Index	1959:01-2019:12	SA	ISM

Table B.54: Category 2. Labor market

ID	Variable	tcode	Description	Period	Seasonality	Source
48	HOUST	4	Housing Starts: Total: New Privately Owned Housing Units Started	1959:01-2019:12	SAAR	FRED
49	HOUSTNE	4	Housing Starts in Northeast Census Region	1959:01-2019:12	SAAR	FRED
50	HOUSTMW	4	Housing Starts in Midwest Census Region	1959:01-2019:12	SAAR	FRED
51	HOUSTS	4	Housing Starts in South Census Region	1959:01-2019:12	SAAR	FRED
52	HOUSTW	4	Housing Starts in West Census Region	1959:01-2019:12	SAAR	FRED
53	PERMIT	4	New Private Housing Units Authorized by Building Permits	1960:01-2019:12	SAAR	FRED
54	PERMITNE	4	New Private Housing Units Authorized by Building Permits in the Northeast Census Region	1960:01-2019:12	SAAR	FRED
55	PERMITMW	4	New Private Housing Units Authorized by Building Permits in the Midwest Census Region	1960:01-2019:12	SAAR	FRED
56	PERMITS	4	New Private Housing Units Authorized by Building Permits in the South Census Region	1960:01-2019:12	SAAR	FRED
57	PERMITW	4	New Private Housing Units Authorized by Building Permits in the West Census Region	1960:01-2019:12	SAAR	FRED
150	HSN1F	4	New One Family Houses Sold: United States	1963:01-2019:12	SAAR	FRED
151	HSN1FNE	4	New One Family Houses Sold in Northeast Census Region	1973:01-2019:12	SAAR	FRED
152	HSN1FMW	4	New One Family Houses Sold in Midwest Census Region	1973:01-2019:12	SAAR	FRED
153	HSN1FS	4	New One Family Houses Sold in South Census Region	1973:01-2019:12	SAAR	FRED
154	HSN1FW	4	New One Family Houses Sold in West Census Region	1973:01-2019:12	SAAR	FRED
155	MSACSR	4	Monthly Supply of Houses in the United States	1963:01-2019:12	SA	FRED
156	HNFSEPUSSA	4	New One Family Homes for Sale in the United States	1963:01-2019:12	SA	FRED
157	UNDCONTSA	4	New Privately-Owned Housing Units Under Construction: Total	1970:01-2019:12	SA	FRED

Table B.55: Category 3. Housing

ID	Variable	tcode	Description	Period	Seasonality	Source
3	DPCERA3M086SBEA	5	Real personal consumption expenditures (chain-type quantity index)	1959:01-2019:12	SA	FRED
4	CMRMTSPLx	5	Real Manu. and Trade Industries Sales	1959:01-2019:12		FRED-MD
5	RETAILx	5	Retail and Food Services Sales	1959:01-2019:12		FRED-MD
58	ACOGNO	5	Manufacturers' New Orders: Consumer Goods	1992:02-2019:12	SA	FRED
59	AMDMNOx	5	New Orders for Durable Goods	1959:01-2019:12		FRED-MD
60	ANDENOx	5	New Orders for Nondefense Capital Goods	1968:02-2019:12		FRED-MD
61	AMDMUOx	5	Unfilled Orders for Durable Goods	1959:01-2019:12		FRED-MD
62	BUSINVx	5	Total Business Inventories	1959:01-2019:12		FRED-MD
63	ISRATIOx	2	Total Business: Inventories to Sales Ratio	1959:01-2019:12		FRED-MD
123	UMCSENTx	2	Consumer Sentiment Index	1959:05-2019:12		FRED-MD
142	USASLMNTO02MLSAM	5	Sales: Manufacturing: Total manufacturing: Value for the United States	1960:01-2019:12	SA	FRED
143	USASLRTTO02MLSAM	5	Sales: Retail trade: Total retail trade: Value for the United States	1960:01-2019:12	SA	FRED
144	USASLWHTO02MLSAM	5	Sales: Wholesale trade: Total wholesale trade: Value for the United States	1960:01-2019:12	SA	FRED
145	USASARTMISMEI	1	Total Retail Trade in United States	1960:01-2019:12	SA	FRED
146	DDURRA3M086SBEA	5	Real personal consumption expenditures: Durable goods (chain-type quantity index)	1959:01-2019:12	SA	FRED
147	DNDGRA3M086SBEA	5	Real personal consumption expenditures: Nondurable goods (chain-type quantity index)	1959:01-2019:12	SA	FRED
148	DSERRA3M086SBEA	5	Real personal consumption expenditures: Services (chain-type quantity index)	1959:01-2019:12	SA	FRED
149	DGDSRA3M086SBEA	5	Real personal consumption expenditures: Goods (chain-type quantity index)	1959:01-2019:12	SA	FRED
158	INVCMRMTSPL	5	Real Manufacturing and Trade Inventories	1967:01-2019:12	SA	FRED
159	SOANDI	1	Chicago Fed National Activity Index: Sales, Orders and Inventories	1967:03-2019:12	NSA	FRED
160	USAODMNTO02MLSAM	5	Orders: Manufacturing: Total orders: Value for the United States	1960:01-2019:12	SA	FRED
172	ISM \MAN_PMI	1	PMI Composite Index	1959:01-2019:12	SA	ISM
175	ISM \MAN_NEWORDERS	1	Manufacturing New Orders Index	1959:01-2019:12	SA	ISM
176	ISM \MAN_DELIV	1	Manufacturing Supplier Deliveries Index	1959:01-2019:12	SA	ISM
177	ISM \MAN_INVENT	1	Manufacturing Inventories Index	1959:01-2019:12	NSA	ISM

Table B.56: Category 4. Consumption, orders, and inventories

ID	Variable	tcode	Description	Period	Seasonality	Source
64	M1SL	6	M1 Money Stock	1959:01-2019:12	SA	FRED
65	M2SL	6	M2 Money Stock	1959:01-2019:12	SA	FRED
66	M2REAL	5	Real M2 Money Stock	1959:01-2019:12	SA	FRED
67	BOGMBASE	6	Monetary Base; Total	1959:01-2019:12	NSA	FRED
68	TOTRESNS	6	Total Reserves of Depository Institutions	1959:01-2019:12	NSA	FRED
69	NONBORRES	7	Reserves of Depository Institutions, Nonborrowed	1959:01-2019:12	NSA	FRED
70	BUSLOANS	6	Commercial and Industrial Loans, All Commercial Banks	1959:01-2019:12	SA	FRED
71	REALLN	6	Real Estate Loans, All Commercial Banks	1959:01-2019:12	SA	FRED
72	NONREVSL	6	Total Nonrevolving Credit Owned and Securitized, Outstanding	1959:01-2019:12	SA	FRED
73	CONSPI	2	Nonrevolving consumer credit to Personal Income	1959:01-2019:12		FRED-MD
124	MZMSL	6	MZM Money Stock	1959:01-2019:12	SA	FRED
125	DTCOLNVHFNM	6	Consumer Motor Vehicle Loans Owned by Finance Companies, Outstanding	1959:01-2019:12	NSA	FRED
126	DTCTHFNM	6	Total Consumer Loans and Leases Owned and Securitized by Finance Companies, Outstanding	1959:01-2019:12	NSA	FRED
127	INVEST	6	Securities in Bank Credit, All Commercial Banks	1959:01-2019:12	SA	FRED
162	USGSEC	5	Treasury and Agency Securities, All Commercial Banks	1959:01-2019:12	SA	FRED

Table B.57: Category 5. Money and credit

ID	Variable	tcode	Description	Period	Seasonality	Source
78	FEDFUNDS	2	Effective Federal Funds Rate	1959:01-2019:12	NSA	FRED
79	CP3Mx	2	3-Month AA Financial Commercial Paper Rate	1959:01-2019:12		FRED-MD
80	TB3MS	2	3-Month Treasury Bill: Secondary Market Rate	1959:01-2019:12	NSA	FRED
81	TB6MS	2	6-Month Treasury Bill: Secondary Market Rate	1959:01-2019:12	NSA	FRED
82	GS1	2	1-Year Treasury Constant Maturity Rate	1959:01-2019:12	NSA	FRED
83	GS5	2	5-Year Treasury Constant Maturity Rate	1959:01-2019:12	NSA	FRED
84	GS10	2	10-Year Treasury Constant Maturity Rate	1959:01-2019:12	NSA	FRED
85	AAA	2	Moody's Seasoned Aaa Corporate Bond Yield	1959:01-2019:12	NSA	FRED
86	BAA	2	Moody's Seasoned Baa Corporate Bond Yield	1959:01-2019:12	NSA	FRED
87	COMPAPFFx	1	3-Month Commercial Paper Minus FEDFUNDS	1959:01-2019:12		FRED-MD
88	TB3SMFFM	1	3-Month Treasury Bill Minus Federal Funds Rate	1959:01-2019:12	NSA	FRED
89	TB6SMFFM	1	6-Month Treasury Bill Minus Federal Funds Rate	1959:01-2019:12	NSA	FRED
90	T1YFFM	1	1-Year Treasury Constant Maturity Minus Federal Funds Rate	1959:01-2019:12	NSA	FRED
91	T5YFFM	1	5-Year Treasury Constant Maturity Minus Federal Funds Rate	1959:01-2019:12	NSA	FRED
92	T10YFFM	1	10-Year Treasury Constant Maturity Minus Federal Funds Rate	1959:01-2019:12	NSA	FRED
93	AAAFFM	1	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	1959:01-2019:12	NSA	FRED
94	BAAFFM	1	Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate	1959:01-2019:12	NSA	FRED
95	TWEXAFEGSMTHx	5	Trade Weighted U.S. Dollar Index	1973:01-2019:12		FRED-MD
96	EXSZUSx	5	Switzerland / U.S. Foreign Exchange Rate	1959:01-2019:12		FRED-MD
97	EXJPUSx	5	Japan / U.S. Foreign Exchange Rate	1959:01-2019:12		FRED-MD
98	EXUSUKx	5	U.S. / U.K. Foreign Exchange Rate	1959:01-2019:12		FRED-MD
99	EXCAUSx	5	Canada / U.S. Foreign Exchange Rate	1959:01-2019:12		FRED-MD
161	RNUSBIS	5	Real Narrow Effective Exchange Rate for United States	1964:01-2019:12	NSA	FRED

Table B.58: Category 6. Interest and exchange rates

ID	Variable	tcode	Description	Period	Seasonality	Source
100	WPSFD49207	6	Producer Price Index by Commodity: Final Demand: Finished Goods	1959:01-2019:12	SA	FRED
101	WPSFD49502	6	Producer Price Index by Commodity: Final Demand: Personal Consumption Goods (Finished Consumer Goods)	1959:01-2019:12	SA	FRED
102	WPSID61	6	Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Processed Goods for Intermediate Demand	1959:01-2019:12	SA	FRED
103	WPSID62	6	Producer Price Index by Commodity: Intermediate Demand by Commodity Type: Unprocessed Goods for Intermediate Demand	1959:01-2019:12	SA	FRED
104	OILPRICEx	6	Crude Oil, spliced WTI and Cushing	1959:01-2019:12		FRED-MD
105	PPICMM	6	Producer Price Index by Commodity: Metals and Metal Products: Primary Nonferrous Metals	1959:01-2019:12	NSA	FRED
106	CPIAUCSL	6	Consumer Price Index for All Urban Consumers: All Items in U.S. City Average	1959:01-2019:12	SA	FRED
107	CPIAPPSL	6	Consumer Price Index for All Urban Consumers: Apparel in U.S. City Average	1959:01-2019:12	SA	FRED
108	CPITRNSL	6	Consumer Price Index for All Urban Consumers: Transportation in U.S. City Average	1959:01-2019:12	SA	FRED
109	CPIMEDSL	6	Consumer Price Index for All Urban Consumers: Medical Care in U.S. City Average	1959:01-2019:12	SA	FRED
110	CUSR0000SAC	6	Consumer Price Index for All Urban Consumers: Commodities in U.S. City Average	1959:01-2019:12	SA	FRED
111	CUSR0000SAD	6	Consumer Price Index for All Urban Consumers: Durables in U.S. City Average	1959:01-2019:12	SA	FRED
112	CUSR0000SAS	6	Consumer Price Index for All Urban Consumers: Services in U.S. City Average	1959:01-2019:12	SA	FRED
113	CPIULFSL	6	Consumer Price Index for All Urban Consumers: All Items Less Food in U.S. City Average	1959:01-2019:12	SA	FRED
114	CUSR0000SA0L2	6	Consumer Price Index for All Urban Consumers: All Items Less Shelter in U.S. City Average	1959:01-2019:12	SA	FRED
115	CUSR0000SA0L5	6	Consumer Price Index for All Urban Consumers: All Items Less Medical Care in U.S. City Average	1959:01-2019:12	SA	FRED
116	PCEPI	6	Personal Consumption Expenditures: Chain-type Price Index	1959:01-2019:12	SA	FRED
117	DDURRG3M086SBEA	6	Personal consumption expenditures: Durable goods (chain-type price index)	1959:01-2019:12	SA	FRED
118	DNDGRG3M086SBEA	6	Personal consumption expenditures: Nondurable goods (chain-type price index)	1959:01-2019:12	SA	FRED
119	DSERRG3M086SBEA	6	Personal consumption expenditures: Services (chain-type price index)	1959:01-2019:12	SA	FRED
163	WPSFD49209	6	Producer Price Index by Commodity: Final Demand: Finished Goods, Excluding Foods	1967:01-2019:12	SA	FRED
164	CPIUFDSL	6	Consumer Price Index for All Urban Consumers: Food in U.S. City Average	1959:01-2019:12	SA	FRED
165	CPIHOSSL	6	Consumer Price Index for All Urban Consumers: Housing in U.S. City Average	1967:01-2019:12	SA	FRED
178	ISM \MAN_PRICES	1	Manufacturing Prices Index	1959:01-2019:12	NSA	ISM

Table B.59: Category 7. Prices

ID	Variable	tcode	Description	Period	Seasonality	Source
74	SP_500	5	S&P's Common Stock Price Index: Composite	1959:01-2019:12		FRED-MD
75	SP_indust	5	S&P's Common Stock Price Index: Industrials	1959:01-2019:12		FRED-MD
76	SP_div_yield	2	S&P's Composite Common Stock: Dividend Yield	1959:01-2019:12		FRED-MD
77	SP_PE_ratio	5	S&P's Composite Common Stock: Price-Earnings Ratio	1959:01-2019:12		FRED-MD
128	VXOCLSx	1	VXO	1962:07-2019:12		FRED-MD

Table B.60: Category 8. Stock market