Forecasting with Partial Least Squares When a Large Number of Predictors Are Available*

Seung C. Ahn† Juhee Bae‡

February 12, 2022

Abstract

We consider Partial Least Squares (PLS) estimation of a time-series forecasting model with the data containing a large number (T) of time series observations on each of a large number (N) of predictor variables. In the model, a subset or a whole set of the latent common factors in predictors are determinants of a single target variable to be forecasted. The factors relevant for forecasting the target variable, which we refer to as PLS factors, can be sequentially generated by a method called "Nonlinear Iterative Partial Least Squares" (NIPLS) algorithm. Two main findings from our asymptotic analysis are the following. First, the optimal number of the PLS factors for forecasting could be much smaller than the number of the common factors in the original predictor variables relevant for the target variable. Second, as more than the optimal number of PLS factors is used, the out-of-sample forecasting power of the factors could rather decrease while their in-sample explanatory power may increase. Our Monte Carlo simulation results confirm these asymptotic results. In addition, our simulation results indicate that unless very large samples are used, 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. We find that 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.

JEL Classification Codes: C51, C53, C55.

Keywords: Partial least squares, Factors, Forecasting.

*We thank Seth Pruitt for generously sharing codes with us. We are also grateful to Domenico Ferraro, Nicolai Kuminoff, Berthold Herrendorf, Bart Hobijn, Nazim Tamkoc, and Gustavo Ventura for their feedbacks and comments. This paper was presented in seminars and workshops at Arizona State University. We would like to thank the participants in the seminars and workshops.

†Department of Economics, W.P Carey School of Business, Arizona State University, Tempe, AZ 85255, USA, email: miniahn@asu.edu

‡Adam Smith Business School, University of Glasgow, G12 8QQ, UK, email: Juhee.Bae@glasgow.ac.uk

1 Introduction

Regressions with a large number of predictor variables (N) could produce poor forecasting results because of high multicollinearity among the predictors, especially when the number of time series observations (T) is not sufficiently larger than N. A treatment to this large-dimensionality problem is the use of shrinkage estimation methods such as Ridge, Bayesian, and Principal Component (PC) regressions. [1] Another possible choice is the Partial Least Squares (PLS) regression that was originally introduced and developed by Wold (1966; 1973; 1982). [2][3] The PLS regression has been popularly used in chemometrics, bioinformatics, machine learning and marketing research. Recently, use of the PLS regression has been increasingly popular in the fields of finance and economics, especially for the analysis of the data with both large N and large T; see, for example, Groen and Kapetanios (2009; 2016), Kelly and Pruitt (2013; 2015), Huang, Jiang, Tu, and Zhou (2015), Carrasco and Rossi (2016), Light, Maslov, and Rytchkov (2017), Tu and Lee (2019), and Rytchkov and Zhong (2021).

The PC and PLS regressions are similar in the sense that both use a small number of estimated factors correlated with the true common latent factors in predictor variables, as regressors. However, these regressions use different approaches to extract factors from predictor variables. Specifically, the PC regression estimates and uses for forecasting all 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 in the sense that it isolates and estimates the relevant factors that are correlated with a target variable from the latent factors that governs predictors; see Mehmood, Liland, Snipen, and Sæbø (2012). For this reason, many previous studies have conjectured that the PLS factors may have higher predictive power than the PC factors. The purpose of this paper is to revisit this conjecture by 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 Kapetanio (2016). Kelly and Pruitt (2015) consider the cases in which individual predictor variables are correlated with a target variable only through the common factors, and a subset or a whole set of these common factors in predictors are the determinant of the target variable. Groen and Kapetanio (2016) examine the forecasting power of PLS factors for the cases in which predictor variables are directly correlated with the target variable, not indirectly through the latent factors. We do not consider the model of Groen and Kapetanio (2016) in this paper. Our asymptotic analysis is conducted for a

[1] see De Mol, Giannone, and Reichlin (2008)

[2] The PLS regression is also a shrinkage estimation method; see, for example, De Jong (1993) and Phatak and De Hoog (2002).

[3] 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).

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 predictive accuracy for the target variable. We also conduct 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.

It is known that the PLS regression may use a smaller number of factors than the PC regression to reach the maximum prediction power. For the cases where 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 largest 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 predictors. For example, if all 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 N/T ratio is sufficiently small. While using more PLS factors can inflate the PLS regression's in-sample fit, 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 N/T ratio is sufficiently small, the out-of-sample prediction power of the PLS regression is often peaked when a fewer number of factors are used than what asymptotic theory suggests. 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 finding 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 2 introduces the model we consider and states the asymptotic properties of the PLS factors. Our Monte Carlo simulation results are reported in Section 3, while some results from a topical empirical study are reported in Section 4. Some concluding remarks follow in Section 5. Proofs of the theorems and lemmas are all given in Appendix.

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

2 Model and Asymptotic Properties of PLS factors

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:

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

$$y_{t+1} = \Sigma_{j=1}^{J} \boldsymbol{f}'_{(j)t}\boldsymbol{\beta}_{(j)} + u_{t+1} = \boldsymbol{f}'_{\bullet t}\boldsymbol{\beta} + u_{t+1}, \tag{2}$$

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 $\boldsymbol{f}_{(j)t}$, $\boldsymbol{f}_{\bullet t} = (\boldsymbol{f}'_{(1)t}, ..., \boldsymbol{f}'_{(J)t})'$, $\boldsymbol{\phi}_{\bullet i} = (\boldsymbol{\phi}'_{(1)i}, ..., \boldsymbol{\phi}'_{(J)i})'$, $\boldsymbol{\beta}_{(j)}$ 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 $\boldsymbol{f}_{\bullet t}$ are sorted into the D different groups, $\boldsymbol{f}_{(1)t}$, ... , $\boldsymbol{f}_{(D)t}$. Without loss of generality, we assume that $\mathrm{E}(\boldsymbol{f}_{\bullet t}) = \mathbf{0}_{K\times 1}$ and $\mathrm{E}(e_{it}) = \mathrm{E}(u_{t+1}) = 0$, for all i and t. For the cases in which $\boldsymbol{f}_{\bullet t}$, x_{it} and y_{t+1} have non-zero means, we can replace them in (1) and (2) respectively by their demeaned versions, $\boldsymbol{f}_{\bullet t} - \bar{\boldsymbol{f}}_{\bullet}$, $x_{it} - \bar{x}_i$, and $y_{t+1} - \bar{y}$, where $\bar{x}_i = T^{-1}\Sigma_{t=1}^{T} x_{it}$, $\bar{\boldsymbol{f}}_{\bullet} = T^{-1}\Sigma_{t=1}^{T}\boldsymbol{f}_{\bullet t}$, and $\bar{y} = T^{-1}\Sigma_{t=1}^{T} y_{t+1}$. Stacking the equations for individual predictors in (1) vertically, we have

$$\boldsymbol{x}_{\bullet t} = \Sigma_{j=1}^{J} \boldsymbol{\Phi}_{(j)} \boldsymbol{f}_{(j)t} + \boldsymbol{e}_{\bullet t} = \boldsymbol{\Phi} \boldsymbol{f}_{\bullet t} + \boldsymbol{e}_{\bullet t}, \tag{3}$$

where $\boldsymbol{x}_{\bullet t} = (x_{1t}, ..., x_{Nt})'$ and $\boldsymbol{e}_{\bullet t} = (e_{1t}, ..., e_{Nt})'$, $\boldsymbol{\Phi}_{(j)} = (\boldsymbol{\phi}_{(j)1}, ..., \boldsymbol{\phi}_{(j)N})'$, and $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_{(1)}, ..., \boldsymbol{\Phi}_{(J)})$. The equations in (3) and (2) can be rewritten by the following two matrix equations:

$$\boldsymbol{X} = \Sigma_{j=1}^{J} \boldsymbol{F}_{(j)} \boldsymbol{\Phi}'_{(j)} + \boldsymbol{E} = \boldsymbol{F}\boldsymbol{\Phi}' + \boldsymbol{E}; \tag{4}$$

$$\boldsymbol{y} = \Sigma_{j=1}^{J} \boldsymbol{F}_{(j)} \boldsymbol{\beta}_{(j)} + \boldsymbol{u} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{u}, \tag{5}$$

where $\boldsymbol{X} = (\boldsymbol{x}_{\bullet 1}, ..., \boldsymbol{x}_{\bullet T})'$, $\boldsymbol{F}_{(j)} = (\boldsymbol{f}_{(j)1}, ..., \boldsymbol{f}_{(j)T})'$, $\boldsymbol{F} = (\boldsymbol{F}_{(1)}, ..., \boldsymbol{F}_{(J)})$, $\boldsymbol{E} = (\boldsymbol{e}_{\bullet 1}, ..., \boldsymbol{e}_{\bullet T})'$, $\boldsymbol{y} = (y_2, ..., y_{T+1})'$, and $\boldsymbol{u}$ is similarly defined. For the model given in (4) and (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 (4) and (5) under some preliminary assumptions that are unrealistically restrictive.

Preliminary Assumption (PA): (i) $\boldsymbol{E} = \boldsymbol{0}_{T \times N}$. (ii) The variable groups, $\boldsymbol{f}_{\cdot t}$ 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 $\boldsymbol{\Phi}'\boldsymbol{\Phi}/N = \boldsymbol{I}_K$.

Some remarks follow on PA. First, under (i), the predictors x_{it} do not have 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 $\boldsymbol{f}_{\cdot t}^*$ have an unrestrictive variance-covariance matrix $\boldsymbol{\Sigma}^*$ and the factor loading matrix $\boldsymbol{\Phi}^*$does not satisfy the assumption (v). Let $\boldsymbol{f}_{\bullet t} = \boldsymbol{f}_{\bullet t}^* (N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{1/2} \boldsymbol{\Xi}^*$ and $\boldsymbol{\Phi} = \boldsymbol{\Phi}^* (N^{-1/2}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1/2} \boldsymbol{\Xi}^*$, where

$$\Xi^* = \Xi((N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{1/2} \boldsymbol{\Sigma}^* (N^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi})^{1/2} | 1 : K).$$

Then, we can easily see that $\boldsymbol{\Phi}\boldsymbol{f}_{\bullet t} = \boldsymbol{\Phi}^* \boldsymbol{f}_{\bullet t}^*$ and $N^{-1/2}\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$. 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.[4]

Under condition (v), the explanatory power of a factor in $\boldsymbol{f}_{\bullet t}$ for individual predictor variables x_{it} are on average proportional to the factor's variance. In the literature, it is often assumed that the individual factors in $\boldsymbol{f}_{\bullet t}$ 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 generalization 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: $

$$\boldsymbol{S}_{NT} = \frac{\boldsymbol{X}'\boldsymbol{X}}{NT}; \boldsymbol{b}_{NT} = \frac{\boldsymbol{X}'\boldsymbol{y}}{N^{1/2}T}. \tag{6}$$

[4]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.

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 $q = 1, ...,$

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

$$\hat{\boldsymbol{P}}_{1:q}^{PC} = \boldsymbol{X}\hat{\boldsymbol{A}}_{1:q}^{PC};$$

$$\hat{\boldsymbol{\delta}}_{1:q}^{PC} = \left(\hat{\boldsymbol{P}}_{1:q}^{PC\prime}\hat{\boldsymbol{P}}_{1:q}^{PC}\right)^{-1}\hat{\boldsymbol{P}}_{1:q}^{PC\prime}\boldsymbol{y};$$

$$\hat{y}_{T+2|q}^{PC} = \boldsymbol{x}_{\bullet T+1}^{\prime}\hat{\boldsymbol{A}}_{1:q}^{PC}\hat{\boldsymbol{\delta}}_{1:q}^{PC}.$$

Here, $\hat{\boldsymbol{A}}_{1:q}^{PC}$ is the $N \times q$ matrix of the PC factor loadings, $\hat{\boldsymbol{P}}_{1:q}^{PC}$ is a $T \times q$ matrix of the first q PC factors, $\hat{\boldsymbol{\delta}}_{1:K}^{PC}$ is the OLS estimator obtained by regressing $\boldsymbol{y}$ on $\hat{\boldsymbol{F}}_{1:q}^{PC}$, and $\hat{y}_{T+2|q}^{PC}$ denotes the forecast for y_{T+2} obtained by the first q PC factors. Under PA, if both $\boldsymbol{f}_{\bullet T+1}$ and $\boldsymbol{\beta}$ were observable, the best forecast for y_{T+2} is $y_{T+2}^{*} \equiv \boldsymbol{f}_{\bullet T+1}^{\prime}\boldsymbol{\beta} = \Sigma_{j=1}^{J}\boldsymbol{f}_{(j)T+1}^{\prime}\boldsymbol{\beta}_{(j)}$. By Bai and Ng (2006), the forecast $\hat{y}_{T+2:K}^{PC}$ that is obtained using the first K PC factors is a consistent estimator of the best forecast y_{T+2}^{*}.

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

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

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

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

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

$$\tilde{y}_{T+2|q}^{PLS} = \boldsymbol{x}_{\bullet T+1}^{\prime}\tilde{\boldsymbol{A}}_{1:q}^{PLS}\tilde{\boldsymbol{\delta}}_{1:q}^{PLS}.$$

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

The factors in $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$ 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 factor vectors in $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$ span the same space as the factor vectors generated by the NIPLS algorithm, and both factors produce the same forecasts. Thus, we refer to the factors of form $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$ 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 $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$, 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 $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$ are more likely to contain serious numerical errors. 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 factors is how to find the optimal q (say, q^*_{PLS}) for forecasting y_{T+2}. Helland (1990) finds that q^*_{PLS} 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 $\hat{\boldsymbol{A}}_{1:q}^{PC}$, $\hat{\boldsymbol{\delta}}_{1:q}^{PC}$, $\hat{\boldsymbol{A}}_{1:q}^{PLS}$, and $\tilde{\boldsymbol{\delta}}_{1:q}^{PLS}$, which are computed replacing $\boldsymbol{b}_{NT}$ and $\boldsymbol{S}_{NT}$ by $\mathrm{E}(\boldsymbol{b}_{NT})$ and $\mathrm{E}(\boldsymbol{S}_{NT})$, respectively. Let us denote them by $\boldsymbol{A}_{1:q}^{PC}$, $\boldsymbol{\delta}_{1:q}^{PC}$, $\boldsymbol{A}_{1:q}^{PLS}$, and $\boldsymbol{\delta}_{1:q}^{PLS}$, respectively. Under PA, we can easily find that

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

With these, we can easily show

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

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

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

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

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 y^*_{T+2}: $y^{PC}_{T+2:K} \equiv \boldsymbol{p}^{PC}_{T+1|K}\delta^{PC}_{1:K} = \boldsymbol{f}'_{\bullet T+1}\boldsymbol{\beta} = y^*_{T+2}$. The optimal number of the PC factors for forecasting y_{T+2} is K (the total number of the common factors in $\boldsymbol{f}_{\bullet t}$).

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

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

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,

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

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

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

It can be also shown that

$$\boldsymbol{\delta}_{1:J}^{PLS} \equiv [\mathrm{E}(\boldsymbol{P}_{1:J}^{PLS\prime}\boldsymbol{P}_{1:J}^{PLS})]^{-1}\mathrm{E}(\boldsymbol{P}_{1:J}^{PLS\prime}\boldsymbol{y}) = N^{-1/2}[\bar{\boldsymbol{D}}_0^*]^{-1}\boldsymbol{1}_J,$$

where $\boldsymbol{1}_J$ 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

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

The optimal number of the PLS factors for forecasting y_{T+2} is J, which is the number of the distinct factor variances, unless some of the $\boldsymbol{\beta}_{(j)}$ are zero vectors or scalar. Thus, unless all the factors in $\boldsymbol{f}_{\bullet t}$ 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 estimate the optimal forecast.

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 $j = 1, \ldots, J$) 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 Kapetanio (2016) consider an alternative model in which the predictor variables x_{it} are directly correlated with y_{t+1}, not indirectly through the latent factors $\boldsymbol{f}_{\bullet t}$. Specifically, they consider a model that consists of equation (4) and a forecast model $y_{t+1} = \boldsymbol{x}'_{\bullet t}\boldsymbol{\beta}^x + u_{t+1}$, where $\boldsymbol{\beta}^x$ is an $N \times 1$ coefficient vector. For this case, $\hat{y}^*_{T+2} = \boldsymbol{f}'_{\bullet T+1}\boldsymbol{\beta}$ is no longer optimal forecast even if both $\boldsymbol{f}_{\bullet T+1}$ and $\boldsymbol{\beta}$ are known. With some restrictive assumptions on $\boldsymbol{E}$ and $\boldsymbol{\beta}^x$, Groen and Kapetanio (2016) show that the PLS regression could generate more accurate forecasting results than the PC regression. For the model given in equations (4) and (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 interesting that idiosyncratic components of some predictors are correlated with y_{t+1}, so we consider some of such cases in our simulation experiments.

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, $\{\boldsymbol{f}_{\bullet t}\}$, $\{\boldsymbol{\phi}_{\bullet i}\}$, $\{e_{it}\}$, 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) $\mathrm{E}(\boldsymbol{f}_{\bullet t}) = \mathbf{0}_{K\times 1}$, $\mathrm{E}(e_{it}) = 0$, and $\mathrm{E}(u_{t+1}) = 0$, for all i and t.

Assumption 2 (A.2): For $j,\ j' = 1, \ldots, J$ 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 $\sigma_1^2 > \sigma_2^2 > \ldots > \sigma_J^2 > 0$, $ks(j) = \Sigma_{h=1}^{j} k(h)$, and $K = ks(J)$. That is, $\hat{\boldsymbol{\Omega}}_{\boldsymbol{F}} = T^{-1}\boldsymbol{F}'\boldsymbol{F} \to_p \boldsymbol{\Omega}_{\boldsymbol{F}} = \boldsymbol{diag}(\sigma_1^2\boldsymbol{I}_{k(1)}, \ldots, \sigma_R^2\boldsymbol{I}_{k(J)})$.

Assumption 3 (A.3): For $j,\ j' = 1, \ldots, J$ 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, $\hat{\mathbf{\Omega}}_{\mathbf{\Phi}} = N^{-1}\mathbf{\Phi}'\mathbf{\Phi} \to_p \boldsymbol{I}_K$.

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

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

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

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

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

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 $\boldsymbol{f}_{\boldsymbol{\cdot}t}$. 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 $\{\boldsymbol{f}_{\boldsymbol{\cdot}t}\}$ and $\{e_{it}\}$ as long as (A.6) holds. As discussed in the previous subsection, the zero-mean assumption on the $\boldsymbol{f}_{\boldsymbol{\cdot}t}$ in 2.2 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 $\mathbf{\Phi}'\mathbf{\Phi} = \boldsymbol{I}_K$ 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 $\hat{\mathbf{\Omega}}_F$ and $\hat{\mathbf{\Omega}}_{\mathbf{\Phi}}$ are T^{γ}-consistent and N^{γ}-consistent estimators of $\mathbf{\Omega}_{\boldsymbol{F}}$ and $\mathbf{\Omega}_{\mathbf{\Phi}}$, respectively, while the elements in $\hat{\mathbf{\Omega}}_{\boldsymbol{F}}$ and $\hat{\mathbf{\Omega}}_{\mathbf{\Phi}}$ 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 $\hat{\mathbf{\Omega}}_{\boldsymbol{F}}$ and $\hat{\mathbf{\Omega}}_{\mathbf{\Phi}}$ could be also T^{γ}-consistent and N^{γ}-consistent for the eigenvalues of $\mathbf{\Omega}_{\boldsymbol{F}}$ and $\mathbf{\Omega}_{\mathbf{\Phi}}$, respectively. For example, Anderson (1963) has shown that the eigenvalues of $\hat{\mathbf{\Omega}}_{\boldsymbol{F}}$ are $T^{1/2}$-consistent if the $\boldsymbol{f}_{\boldsymbol{\cdot}t}$ are *iid* multivariate normal vectors, In fact, the eigenvalues of $\hat{\mathbf{\Omega}}_{\boldsymbol{F}}$ are$T^{1/2}$-consistent even if the $\boldsymbol{f}_{\boldsymbol{\cdot}t}$ are not normal; see Fang and Krishnaiah (1982). It is too restrictive to assume that $\{\boldsymbol{f}_{\boldsymbol{\cdot}t}\}$ is an *iid* process. Taniguchi and Krishnaiah (1987) have shown that the eigenvalues of $\hat{\mathbf{\Omega}}_{\boldsymbol{F}}$ are $T^{1/2}$-consistent if $\{\boldsymbol{f}_{\boldsymbol{\cdot}t}\}$ 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 $M^{-1}\boldsymbol{E}'\boldsymbol{E}$ 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), $\Sigma_{h=1}^{m}\lambda_h((NT)^{-1}\boldsymbol{E}'\boldsymbol{E}) \geq (m_c/m)(c + o_p(1)) > 0$.

Sufficient conditions for (A.6) are the following: As $N \to \infty$ for each t and as $T \to \infty$, for each i,

$$N^{-1/2}\Sigma_{i=1}^{N}\boldsymbol{\phi}_{\bullet i}e_{it} \to_d N(\mathbf{0}_{K\times 1}, \boldsymbol{\Gamma}_t); \tag{7}$$

$$T^{-1/2}\Sigma_{t=1}^{T}\boldsymbol{f}_{\bullet t}e_{it} \to_d N(\mathbf{0}_{K}, \boldsymbol{\Gamma}_i) \tag{8}$$

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 (8), Bai (2003) and have derived the asymptotic distributions of the principcomponent factors and factor loadings. Imagine that the factor loading matrix Φ is observable. For such cases, the factor vector $\boldsymbol{f}_{\bullet t}$ can be consistently estimated by the OLS regression of $x_{\bullet t}$ on Φ. The conditions (A.2) and (7) are the sufficient conditions under which the resulting OLS estimators are asymptotically normal. Similarly, for the cases in which the latent factor matrix $\boldsymbol{F}$ is observable, the conditions (A.2) and (8) are the sufficient conditions under which the OLS estimators of $\boldsymbol{\phi}_{\bullet i}$ obtained by regressing $\boldsymbol{x}_{i\bullet} = (x_{i1}, ..., x_{iT})'$ 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 $\left\|(NT)^{-1/2}\boldsymbol{F}'\boldsymbol{E}\right\|_F = O_p(1)$, $\left\|(NT)^{-1/2}\boldsymbol{\Phi}'\boldsymbol{E}\right\| = O_p(1)$, and $\left\|(NT)^{-1/2}\boldsymbol{\Phi}'\boldsymbol{E}'\boldsymbol{F}\right\| = O_p(1)$. It is shown in Appendix (Lemma C.3) that these conditions hold under (A.1) and (A.6). 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 $y^*_{T+2} = \Sigma_{j=1}^{J}\boldsymbol{f}'_{(j)T+1}\boldsymbol{\beta}_{(j)}$. Strictly speaking, y^*_{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 y^*_{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 $j = 1, \ldots, R$ (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}.

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

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

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

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

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

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

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

$$\tilde{y}_{T+2|1}^{PLS} = \boldsymbol{x}_{\bullet T+1}'\tilde{\boldsymbol{\alpha}}_1^{PLS}\tilde{\boldsymbol{\delta}}_{1:1}^{PLS} = \frac{m^{1/2}}{T^{1/2}}\frac{\boldsymbol{c}_L'\boldsymbol{c}_L}{\boldsymbol{c}_L^{*\prime}\boldsymbol{\Lambda}_L^*\boldsymbol{c}_L^*}\frac{\boldsymbol{e}_{\bullet T+1}'}{N^{1/2}}\frac{\boldsymbol{E}'\boldsymbol{u}}{M^{1/2}}.$$

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

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

because

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

These results indicate that $\left|\tilde{y}_{T+2|1}^{PLS} - y_{T+2}^*\right| = \left|\tilde{y}_{T+2|1}^{PLS}\right| = O_p((m/T)^{1/2})$, where $y_{T+2}^* = 0$. Thus, if $N/T \to 0$, then,$\left|\tilde{y}_{T+2|1}^{PLS} - y_{T+2}^*\right| \to_p 0$ 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 y^*_{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.

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

$$\begin{aligned}\boldsymbol{\Lambda}^{\boldsymbol{S}_{NT}}_{(j)} &= \boldsymbol{\Lambda}(\boldsymbol{S}_{NT}|ks(j-1)+1 : ks(j));\\ \boldsymbol{\Xi}^{\boldsymbol{S}_{NT}}_{\boldsymbol{(j)}} &= \boldsymbol{\Xi}(\boldsymbol{S}_{NT}|ks(j-1)+1 : ks(j));\\ \boldsymbol{c}^{\boldsymbol{S}_{NT}}_{(j)} &= \boldsymbol{\Xi}^{\boldsymbol{S}_{NT}\prime}_{(j)}\boldsymbol{b}_{NT}\end{aligned}$$

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

Similarly, we also define

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

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

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

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

With the above notation and result, we can show that

$$\tilde{\boldsymbol{\alpha}}^{PLS}_q = (\boldsymbol{S}_{NT})^{q-1}\boldsymbol{b}_{NT} = \Sigma^J_{j=1}\boldsymbol{\Xi}^{\boldsymbol{S}_{NT}}_{(j)}(\boldsymbol{\Lambda}^{\boldsymbol{S}_{NT}}_{(j)})^{q-1}\boldsymbol{c}^{\boldsymbol{S}_{NT}}_{(j)} + \boldsymbol{\Xi}^{\boldsymbol{S}_{NT}}_{L}(\boldsymbol{\Lambda}^{\boldsymbol{S}_{NT}}_{L})^{q-1}\boldsymbol{c}^{\boldsymbol{S}_{NT}}_{L}. \tag{9}$$

Thus, the asymptotic property of the $q-$th PLS coefficient vector $\tilde{\boldsymbol{\alpha}}^{PLS}_q$ depends on those of the eigenvalues and eigenvectors of the matrix $\boldsymbol{S}_{NT}$, the vector $\boldsymbol{b}_{NT}$, and the vectors $\boldsymbol{c}^{\boldsymbol{S}_{NT}}_{(j)}$ and $\boldsymbol{c}^{\boldsymbol{S}_{NT}}_{L}$. 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) $\lambda_h(\boldsymbol{S}_{NT}) = \sigma^2_j + O_p(m^{-\gamma})$, for $h = ks(j-1)+1, ..., ks(j)$ and $j = 1, ..., J$.

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

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

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

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

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

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

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

Let $\boldsymbol{H}_{NT} = (NT)^{-1/2} \Xi_L^{\boldsymbol{S}_{NT}\prime} \boldsymbol{\mathcal{Q}}(\boldsymbol{\Phi}) \boldsymbol{E}' \boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})$ 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) $\left\| \boldsymbol{c}_L^{\boldsymbol{S}_{NT}} - T^{-1/2} \boldsymbol{H}_{NT} \boldsymbol{u} \right\|_2 = O_p(m^{-3/2})$;

(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 $\left\| \boldsymbol{c}_L^{\boldsymbol{S}_{NT}} \right\|_2 = O_p(T^{-1/2} + m^{-3/2})$. Second, the convergency speed of $\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}$ 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 $\boldsymbol{e}_{\cdot t}$ are uncorrelated with the error u_{t+1}, linear combinations of the N error terms in $\boldsymbol{e}_{\cdot t}$ 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 . Define

$$
\begin{aligned}
\mu_j^{\boldsymbol{S}_{NT}} &= \lambda_{ks(j-1)+1}^{\boldsymbol{S}_{NT}}, \text{for } J = 1, ..., R; \\
\boldsymbol{d}_0(q) &= ((\mu_1^{\boldsymbol{S}_{NT}})^{q-1}, (\mu_2^{\boldsymbol{S}_{NT}})^{q-1}, ..., (\mu_R^{\boldsymbol{S}_{NT}})^{q-1})'; \\
\boldsymbol{D}_0(q) &= (\boldsymbol{d}_0(1), \boldsymbol{d}_0(2), ..., \boldsymbol{d}_0(q)).
\end{aligned}
$$

Notice that $\mu_j^{\boldsymbol{S}_{NT}}$ is the largest one in the jth group of the eigenvalues, $\lambda_{ks(j-1)+1}^{\boldsymbol{S}_{NT}}, \ldots, \lambda_{ks(j)}^{\boldsymbol{S}_{NT}}$. 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 $\mu_j^{\boldsymbol{S}_{NT}}$ $(j = 1, ..., R)$ are all distinct even asymptotically. By Lemma 2.4.1, $\mu_j^{\boldsymbol{S}_{NT}} \to_p \sigma_j^2$ as $m \to \infty$.

With the terms defined above, we can easily show that

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

where

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

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

$$
\tilde{\boldsymbol{A}}_{1:q}^{PLS} = \boldsymbol{V}_0\boldsymbol{D}_0(q) + \boldsymbol{V}_{H1}(q) + \boldsymbol{V}_{H2}(q) + \boldsymbol{V}_L(q) \tag{10}
$$

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 $\boldsymbol{\alpha}_q^{\tilde{P}LS}$ 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

$$
\begin{aligned}
&\boldsymbol{\Pi}_{NT} = N^{-1/2}[\boldsymbol{\Phi}_{(1)}\boldsymbol{\beta}_{(1)}, \boldsymbol{\Phi}_{(2)}\boldsymbol{\beta}_{(2)}, ..., \boldsymbol{\Phi}_{(R)}\boldsymbol{\beta}_{(R)}];\\
&\boldsymbol{\Sigma}_R = \boldsymbol{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_R^2).
\end{aligned}
$$

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 $q = 1, \dots, R$,

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

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

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

$$\begin{aligned}&\boldsymbol{G}_0 = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{V}_0;\\&\boldsymbol{g}_{H1}(q) = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{v}_{H1}(q); \quad \boldsymbol{g}_{H2}(q) = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{v}_{H2}(q)\\&\boldsymbol{g}_L(q) = (NT)^{-1/2}\boldsymbol{X}\boldsymbol{v}_L(q)\end{aligned}$$

With this notation, we have

$$(NT)^{-1/2}\tilde{\boldsymbol{p}}_q^{PLS} = (NT)^{-1/2}\boldsymbol{X}\tilde{\boldsymbol{\alpha}}_q^{PLS} = \boldsymbol{G}_0\boldsymbol{d}_0(q) + \boldsymbol{g}_{H1}(q) + \boldsymbol{g}_{H2}(q) + \boldsymbol{g}_L(q) \tag{11}$$

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

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

where

$$\begin{aligned}&\hat{\boldsymbol{d}}_0(q) = \boldsymbol{d}_0(q) + (\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}\boldsymbol{G}_0'\boldsymbol{g}_{H1}(q);\\&\boldsymbol{g}_H^c(q) = (\boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q), \boldsymbol{g}_{H2}(q)).\end{aligned}$$

By eq:plsfq, we also have

$$(NT)^{-1/2}\tilde{\boldsymbol{P}}_{1:q}^{PLS} = \boldsymbol{G}_0\hat{\boldsymbol{D}}_0(q) + \boldsymbol{G}_H^c(q) + \boldsymbol{G}_L(q), \tag{12}$$

where

$$\begin{aligned}&\hat{\boldsymbol{D}}_0(q) = (\hat{\boldsymbol{d}}_0(1), ..., \hat{\boldsymbol{d}}_0(q)) = \boldsymbol{D}_0(q) + (\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}\boldsymbol{G}_0'\boldsymbol{G}_{H1}(q);\\&\boldsymbol{G}_H^c(q) = \boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\boldsymbol{G}_{H1}(q) + \boldsymbol{G}_{H2}(q),\end{aligned}$$

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

Two remarks follow on equation eq: PLSFQ. 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 $\boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\boldsymbol{G}_{H1}(q)$ 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 $\boldsymbol{f}_{\cdot t}$ have distinct asymptotic variances and are correlated with the target variable y_{T+2}. For the case, $\boldsymbol{G}_H^c(q) = \boldsymbol{0}_{T\times q}$ 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) $\|\boldsymbol{G}_H^c(q)\|_F = O_p(m^{-\gamma})$; $\|\boldsymbol{G}_L(q)\|_F = O_p(m^{-1/2}(T^{-1/2} + m^{-3/2}))$;

(ii) $\left\| \hat{\boldsymbol{D}}_0(q) - \boldsymbol{D}_0(q) \right\|_F = O_p(m^{-\gamma})$;

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

Lemma 2.4.3 and Corollary 2.4.3 indicate that the asymptotically dominant term in $\tilde{\boldsymbol{P}}_{1:q}^{PLS}$ is $\boldsymbol{G}_0$. For $q \leq R$, the asymptotic properties of the q PLS factors in $\tilde{\boldsymbol{P}}_{1:R}^{PLS}$ 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) $\left\| (NT)^{-1}\tilde{\boldsymbol{P}}_{1:R}^{PLS\prime}\tilde{\boldsymbol{P}}_{1:R}^{PLS} - \hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) \right\|_F = O_p(m^{-\gamma})$;

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

When $R = K$,

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

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

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

Theorem 2: Under (A.1) – (A.8),

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

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

(iii) $R^2_{1:R} \equiv \frac{\boldsymbol{y}'\boldsymbol{\mathcal{P}}\left(\tilde{\boldsymbol{P}}^{PLS}_{1:R}\right)\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} \to_p R^2_{\max} \equiv \frac{\Sigma_{j=1}^R \sigma_j^2 \boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)}}{\Sigma_{j=1}^R \sigma_j^2 \boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)} + \sigma_u^2}.$

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, $y^*_{T+2} = \Sigma_{j=1}^R \boldsymbol{f}'_{(j)T+1}\boldsymbol{\beta}_{(j)}$. We can show that the forecast by a fewer number of PLS factor is not consistent for y^*_{T+2}. Thus, the minimum number of the PLS factors that can produce a consistent estimator of y^*_{T+2} is R, the number of distinct asymptotic variances of the common factors in $\boldsymbol{f}_{\cdot t}$ 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 y^*_{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 $\boldsymbol{F}_{(1)}, \ldots, \boldsymbol{F}_{(R)}$. 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

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

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

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

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

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

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

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

When $R < K$, $m^{2\gamma}\mathcal{Y}_{1,NT}$ and $m^{2\gamma}\mathcal{Y}_{2,NT}$ are positive random variables. Consequently, $(\mathcal{Y}_{2,NT})^2/\mathcal{Y}_{1,NT} = O_p(m^{-2\gamma}) = o_p(1)$. 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) $\quad T^{-1}\boldsymbol{y}'\boldsymbol{\mathcal{P}}(\tilde{\boldsymbol{P}}^{PLS}_{1:R+1})\boldsymbol{y} = \Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)} + |O_p(1)| > \Sigma^R_{j=1}\sigma^2_j\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)}.$

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

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

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. By the inversion rule for partitioned matrix, we can show

$$\begin{aligned} N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R+1} &= \left((NT)^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R+1}\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}\right)^{-1} N^{-1/2}T^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R+1}\boldsymbol{y} \\ &= \left((NT)^{-1}\begin{pmatrix} \tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R} & \tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1} \\ \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{P}}^{PLS}_{1:R} & \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{p}}^{PLS}_{R+1} \end{pmatrix}\right)^{-1} N^{-1/2}T^{-1}\begin{pmatrix} \tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y} \\ \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\boldsymbol{y} \end{pmatrix}. \\ &= \begin{pmatrix} N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} \\ 0 \end{pmatrix} - \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}\mathcal{Y}_{NT}, \end{aligned}$$

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

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

(iii) $\left\| \tilde{y}^{PLS}_{T+2|R+1} - y^*_{T+2} \right\| = O_p(1)$, 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 $m^{\gamma}(\tilde{y}^{PLS}_{T+2|R+1} - y^*_{T+2})$ depends on $\mathcal{Y}_{NT}$, whose mean is not zero. That is, the forecast $\tilde{y}^{PLS}_{T+2|R+1}$ is an asymptotically biased estimator of the y^*_{T+2}. This result suggests that the finite-sample property of $\tilde{y}^{PLS}_{T+2|R+1}$ may not be as good as that of $\tilde{y}^{PLS}_{T+2|R}$, 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 y^*_{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.

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.

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 $\boldsymbol{f}_{\cdot t}^* = (f_{1t}^*, ..., f_{Kt}^*)'$ 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 $\boldsymbol{f}_{\cdot t}^*$, while a_y equals the probability limit of the R^2 from the regression of y_{t+1} on $\boldsymbol{f}_{\cdot t}^*$.

We use $\boldsymbol{\Omega}_{\boldsymbol{F}}^*$ to denote $\text{Var}(\boldsymbol{f}_{\cdot t}^*) = \boldsymbol{diag}(v_1, ..., v_K)$. 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:

$$
\begin{aligned}
f_{ht} &\approx a_x^{1/2} f_{ht}^* / \sqrt{\operatorname{var}(\Sigma_{h=1}^K f_{ht}^* \phi_{hi})}; \\
e_{it} &\approx (1-a_x) e_{it}^* / \sqrt{\operatorname{var}(e_{it}^*)}; \\
\beta_h &\approx \beta_h^* \frac{a_y^{1/2} \sqrt{\operatorname{var}(\Sigma_{h=1}^K f_{ht}^* \phi_{hi})}}{a_x^{1/2} \sqrt{\operatorname{var}(\Sigma_{k=1}^K f_{ht}^* \beta_h^*)}}; \\
\boldsymbol{\Omega_F} &\approx \frac{a_x}{\operatorname{var}(\Sigma_{h=1}^K f_{ht}^* \phi_{hi})} \boldsymbol{\Omega_F^*}.
\end{aligned}
$$

For each set of the parameter values chosen (T, N, K, $\boldsymbol{\Omega_F}$, 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 errors by different forecasts. Using the forecasting errors from the 1,000 samples, we compute the following out-of-sample R^2 of a forecast, $\hat{y}_{T+2}$:

$$
R_{OS}^2 \equiv 1 - \frac{\Sigma_{s=1}^{1000} (y_{T+2}^{[s]} - \hat{y}_{T+2}^{[s]})^2}{\Sigma_{s=1}^{1000} (y_{T+2}^{[s]} - \bar{y}^{[s]})^2}
$$

where $\bar{y}^{[s]} = T^{-1} \Sigma_{t=1}^T y_{t+1}^{[s]}$ and s indexes simulated samples. The second term of R_{OS}^2 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_{OS}^2 must be a positive number. In contrast, when the historical mean outperforms, the measure becomes negative. The R_{OS}^2 measure is also used in Kelly and Pruitt (2015).

Our benchmark case is the case in which data are generated with $N = T = 100$, $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)'$, $\boldsymbol{\Omega_F^*} = \boldsymbol{diag}(3,3,5,5)$, $a_x = 0.2$, $a_y = 0.7$, 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.

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.

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-sampleR^2's and the R_{OS}^2'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^2 does not. 1 depicts the

changes in average in-sample (adjusted) R^2 and R^2_{OS} as the number of PLS factors used (q) increases.

1 and 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 further from 2, the R^2_{OS} keeps falling. For example, as q increases to 10, the R^2_{OS} falls to 18 percent points while the in-sample R^2 increases to 90 percent points. 1 and 1 clearly show that a PLS regression with higher in-sample R^2 does 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 $y^*_{T+2} = \Sigma^K_{h=1} f_{h,T+1}\beta_h$. Interestingly, however, the simulation results reported in 1 and 1 indicate that using 3 or 4 PLS factors would rather produce less accurate forecasts. The results in 1 and 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^2_{OS}'s. The results reported in Table 1 and 1 are also consistent with this prediction.

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 2 and 3 report the out-of-sample forecasting performances of the PLS regression with the first PLS factor only (PLS1), the PC regression with first 5 PC factors (PC5), and the usual OLS regression with all predictor variables (OLS). Table 2 reports the results obtained from the data with $(N, T) = (80, 100)$, while Table 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 2 and 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 2 and 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 2 and 3 is that when the PLS1 regression outperforms the PC5 regression, it does so by a relatively greater margin. For example, in Table 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 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 R^2_{OS}'s from the PLS1 and PC5 regressions are 48.5 percent points and 33.2 percent points, respectively.

Tables 2 and 3 also report the number of common factors ($\hat{K}$) estimated by the Eigenvalue Ratio (ER) method of method of Ahn and Horenstein (2013). The tables show that when a_x is low, the ER method tends to underestimate the number of common factors in predictor variables. Not surprisingly, the PC regression with the estimated number of factors ($\hat{K}$) 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 2 and 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.

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 4 – 6 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 4 – 6 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 4 to 6, the optimal number of PLS factors for forecasting is three ($R = 3$).

Table 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 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 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 $N = T = 10,000$, 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 4 – 6 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 \ref{table1}, 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 4 – 6, the variances of the three factors, f_{ht} ($h = 1, 2, 3$) 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 ($\tilde{\boldsymbol{P}}^{PLS}_{1:R}$) 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 $\boldsymbol{y}$.

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 ratio N/T. 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 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 1. Figure 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 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 as more PLS factors are used.

Figures 3 and 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, theN/T ratios used for the two figures are the same as those which are used for Figure 2. The reported results in Figures 3 and 4 are not materially different from those in Figure 2. Overall, the results reported in Figures 2 – 4 are consistent with the notion that the severity of the spurious correlation problem and the N/T ratio are inversely related.

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 increase.

Figure 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 5 are the same as those that are used for Figure 1. Figure 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 of 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 6 – 7 report the results obtained replicating the simulation exercises used for Figure 5, but with greater values of a_x (0.5 and 0.7, respectively). The patterns of the PLS forecasting performance reported in Figures 6 and 7 are virtually identical to those that are reported in Figure 5.

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 1 and Figure 1. Figure 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 5-th 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 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 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 9 and 10 show, the results from Table 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 Figures 8 to 10.

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 11 and 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 1 and other figures that using more than the informative PLS factors decreases the accuracy of the PLS forecast.

Figure 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 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 12, we experiment the same simulations conducted for Figure 11, but with larger data. The data are generated with $N = T = 2000$. In Figure 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 11 and 12 would change if more uninformative factors are added to predictor variables, we conduct the same simulation exercises used for Figures 11 and 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 \ref{figure13} and Figure \ref{figure14}. For the factor model used for the figures, there are two informative and four uninformative PLS factors.

Figure 13 reports the results obtained using the data with $N = T = 100$ as in Figure 11. From Figure 13, we can see that the regression using the 6-th 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 14 reports the result from the data with $N = T = 2,000$ as in Figure 12. Three out of four uninformative factors perform more consistently with what our asymptotic results predict. However, the last uninformative factor (the 6-th PLS factor) behaves more like a spurious factor, especially when a_x is low. The main point from Figures 11 – 12 and 13 –

14 is that using uninformative PLS factors can significantly lower the accuracy of the PLS forecast, unless data are unusually large.

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 $N = T = 2,000$). 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, Barigozzi, and Capasso (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.

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 $\text{int}(\cdot)$ 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 8 to 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 8 to 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 Tables 8 – 10 are as follows. First, consistent with the results reported in Tables 4 and 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 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 forecasting 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 Kapetanio (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 $\boldsymbol{f}_{\cdot t}$. 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}:

$$x_{1t} = \Sigma_{h=1}^{5} \phi_{hi} f_{ht}^{*} + e_{1t}^{*}; \quad e_{1t}^{*} = \rho_{eu}^{1/2} u_{t+1}^{*} + (1 - \rho_{eu})^{1/2} v_{1t}^{*} \tag{13}$$

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 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 8 – 10, except that the first predictor variable is generated by (13). Table 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 8 – 11. However, one interesting observation from Table 10 is that the spurious correction problem by using the sixth PLS factor (which is a spurious factor) mitigates as ρ_{eu} increases. 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 10 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.

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 ER method of Ahn and Horenstein (2013).

The following forecasting equation is used for our data analysis:

$$\hat{y}_{t+12|t}^{(12m)} = \hat{a} + \hat{\boldsymbol{b}}' \hat{\boldsymbol{f}}_{\bullet t} + \Sigma_{h=1}^{p} \hat{c}_h y_{t-h+1}, \tag{14}$$

where $\hat{y}_{t+12:t}^{(12m)}$ denotes the 12 month ahead forecast of a target variable $y_{t+12}^{(12m)}$ made at time t,$\hat{\boldsymbol{f}}_{\bullet t}$ is a vector of PLS or PC factors, 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 $\hat{\boldsymbol{f}}_{\bullet t}$ are restricted to be 6 and 12, respectively.

The number of factors used matters for predictive power. We have conducted many experiments with different choices of the number of PLS or PC factors. First, for both the regressions with PLS and PC factors, we have tried 12 different numbers of factors: $q = 1, 2, ..., 12$. We denote the regression with q PLS (PC) factors by "PLSq" ("PCq"). Tables 11 and 12 display the forecasting results from the PLS1 – PLS4 and PC1 – PC4 regressions. Second, we estimate the the number of latent factors in predictor variables by the Bayesian Information Criterion (BIC) method of Stock and Watson (2002b) and the ER method of Ahn and Horenstein (2013). We denote the regressions with these two estimates of the number of latent factors by PC-BIC and PC-AH, respectively. Third and finally, we estimate the optimal number of PLS factors by the BIC method of Stock and Watson (2002b) and the Cross-Validation method used in subsection 3.9. The regressions with these two estimates are denoted by PLS-BIC and PLS-CV, respectively.

The target variables $y_{T+12}^{(12m)}$ 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). Under this assumption, to forecast a real variable such as industrial production (IP), we use the target variable

$$y_{t+12}^{(12m)} = (1200/12) \ln(\mathrm{IP}_{t+12}/\mathrm{IP}_t),$$

and the regressors

$$y_{t-h} = 1200 \ln(\mathrm{IP}_{t-h}/\mathrm{IP}_{t-h-1})$$

In contrast, to forecast a price variable such as Consumer Price Index (CPI), we use the target variable

$$y^{(12m)}_{T+12} = (1200/12) \ln(\mathrm{CPI}_{T+12}/\mathrm{CPI}_T) - 1200 \ln(\mathrm{CPI}_{T+12}/\mathrm{CPI}_T),$$

and the regressors

$$y_{t-h} = 1200\Delta \ln(\mathrm{CPI}_{t-h}/\mathrm{CPI}_{t-h-1}).$$

Since PLS factors computed with target variables, we should have enough time series observations for the target variables. Target variables with too many missing values may lead to inaccurate PLS factors making our empirical analysis unreliable. Therefore, we only use the variables whose time-series observations are more than 80% of the first-step estimation period. This variable selection rule leave us 144 different target variables.

Our empirical analysis is conducted by the following way. Using the data up to a given time T, we compute the target variables, lagged dependent variables, and PC and PLS factors. We also estimate the optimal numbers of the PC factors by the BIC method of Stock and Watson (2002b) and the ER method of Ahn and Horenstein (2013). We also estimate the number of PLS factors to be used by the BIC of Stock and Watson (2002b) and the CV method discussed in subsection 3.9. The number of the lagged dependent variables, p, is estimated by the BIC method of Stock and Watson (2002b) or the CV procedure discussed in subsection 3.9. With these estimates, the forecasting equation 14 is estimated by regressing $y^{(12m)}_{t+12}$ on the PC or PLS factor vectors $\hat{\boldsymbol{f}}_{\boldsymbol{\cdot}t}$ and the p lagged dependent variables y_{t-h}. This procedure yields the estimated parameters, $\hat{a}$, $\hat{\boldsymbol{b}}$ and the $\hat{c}_h$. With these results, we make the forecasts for $y^{(12m)}_{T+12}$, which are denoted by $\hat{y}^{(12m)}_{T+12|T}$ in 14. We repeat ths procedure using the data up to $T + 1$ and forecast $y^{(12m)}_{T+13}$. We continue this exercise until the whole data are exhausted.

More specifically, our first forecasting starts from $T = 1970{:}01$. at the first step, we use the predictors from 1959:03 to 1970:01 to estimate PLS and PC factors at time T, $\hat{\boldsymbol{f}}_{\boldsymbol{\cdot}T}$. The first two-month observations in the data are dropped due to possible second difference (in log) in the data transformation procedure. The optimal number of PC and PLS factors to be used are also estimated. Second, we estimate equation 14 using the data up to 1970:01. Third and finally, using the results from the first and second steps, we make the forecasts for $y^{(12m)}_{T+12}$, $\hat{y}^{(12m)}_{T+12|T}$. This procedure repeats until T becomes 2018:12.

For the cross-validation method used for regressions witj PLS factors, we use 70% and 30% of the available data as the training and test data sets, respectively. For instance, when we make the forecast for $y^{(12m)}_{1971:01}$ at 1970:01, we use the data $\{\boldsymbol{x}_{\cdot T}\}^{1966:04}_{T=1959:01}$and $\{y^{(12m)}_{T+12}\}^{1965:04}_{T=1959:01}$ as training data because they are roughly 70% of the data available up to 1970:01. Using the training data, we compute PLS factors and estimate the parameters necessary for forecasting for each of the possible combinations of the number of PLS factors (q) and the number of lagged dependent variables y_{t-h} (p). For each possible combination of q and p, we also compute the forecast, $\hat{y}^{(12m)}_{T+12|T=1966:4}$. We repeat this procedure for $T = 1966{:}5$ and calcuate $\hat{y}^{(12m)}_{T+12|T=1966:5}$. By continuing this procedure up to $T = 1969{:}01$, we can obtain the set of forecasts $\{\hat{y}^{(12m)}_{T+12|T}\}^{1969:01}_{T=1966.4}$. Comparing $\{\hat{y}^{(12m)}_{T+12|T}\}^{1969:01}_{T=1966.4}$ and $\{y^{(12m)}_{T+12}\}^{1969:01}_{T=1966.4}$, we compute the Mean Square Error (MSE) for each possible combination of q and p. We denote by the "PLS-CV" forecast for 1971:01 the forecast obtained using the combination of q and p that minimizes the MSE. By repeating this procedure for T = 1970:02 and other future months, we can obtain the PLS-CV forecasts for 1971:01 up to 2019:12.

For each forecasting method applied to each of the 144 economic variables, we compute the mean squared errors (MSE) by comparing $\{y^{(12m)}_{T+12}\}^{2018:12}_{T=1970:01}$ and $\{\hat{y}^{(12m)}_{T+12|T}\}^{2018:12}_{T=1970:01}$. The results are presented in Tables 11 and 12. The entries are the percentage R^2_{OS}, 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 a forecasting method is the method's MSE relative to that of a forecast based on a naïve historical mean of the target variable.

Table 11 displays the R^2_{OS}'s from different forecasting models for the eight variables on which the analysis of Stock and Watson (2002b) focuses. Table 12 reports the forecasting results for the whole 144 target variables. To save space, we categorize the 144 variables into eight different groups and report the median percentage R^2_{OS} for each group. Tables 11 and 12 reveal some interesting results. First, consistent with our simulation results, incorporating more than one 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; for example, Personal Income. Some target variables show improvement when we use more than PLS factors; for example, Producer Price Index. However, even for such cases, the predictive improvement is marginal and the regressions with only one PLS factor (PLS1) still often produce better forecasting results.

Second, the PLS-CV forecast does not dominate the PLS1 forecast. Rather, the PLS-CV forecasts are very often dominated by the PLS1 forecast. This result is again consistent with our simulation results. Third, the PLS-BIC forecast shows significantly worse performance. Even the forecasts based on a historical mean strictly dominates the PLS-BIC forecast in many cases. This is not surprising, because the BIC method chooses the number of PLS factors that maximizes the in-sample fit. Therefore, the number of PLS factors that explains the in-sample movements very well does not necessarily produce better forecasts. As the two tables confirm, the PLS forecasts obtained using the number of factors estimated by the BIC method actually poorly performs in the data. Finally, the PLS1 regression outperforms other alternative regression methods. Sometimes the PLS1 regression does not produce the best forecasting results. However, for such cases, the performances of the best forecasts and the PLS1 forecast are very similar.

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.

References

Ahn, Seung C. and Alex R. Horenstein (2013). "Eigenvalue ratio test for the number of factors". In: *Econometrica* 81.3, pp. 1203–1227.

Alessi, Lucia, Matteo Barigozzi, and Marco Capasso (2010). "Improved penalization for determining the number of factors in approximate factor models". In: *Statistics and Probability Letters* 80, pp. 1806–1813.

Anderson, Theodore W. (1963). "Asymptotic theory for principal component analysis". In: *Annals of Mathematical Statistics* 34.1, pp. 122–148.

Bai, Jushan (2003). "Inferential theory for factor models of large dimensions". In: *Econometrica* 71.1, pp. 135–171.

Bai, Jushan and Serena. Ng (2002). "Determining the number of factors in approximate factor models". In: *Econometrica* 70.1, pp. 1991–221.

Butler, N. A. and M. C. Denham (2000). "The peculiar shrinkage properties of Partial Least Squares Regression". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 62.3, pp. 585–593.

Carrasco, Marine and Barbara Rossi (2016). "In-sample inference and forecasting in misspecified factor models". In: *Journal of Business & Economic Statistics* 34.3, pp. 313–338.

Dax, A. (2017). "The numerical rank of Krylov matrix". In: *Linear Algebra and its Applications.*

De Jong, Sijmen (1993). "PLS fits closer than PCR". In: *Journal of Chemometrics* 7.6, pp. 551–557.

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008). "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" In: *Journal of Econometrics* 146.2, pp. 318–328.

Eaton, Morris L. and David E. Tyler (1991). "On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix". In: *The Annals of Statistics* 19.1, pp. 260–271.

Fang, C. and P. R. Krishnaiah (1982). "Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations". In: *Journal of Multivariate Analysis* 12.1, pp. 39–63.

Groen, Jan J. J. and George Kapetanio (2016). "Revisiting useful approaches to data-rich macroeconomic forecasting". In: *Computational Statistics & Data Analysis* 100.C, pp. 221–239.

Groen, Jan J. J. and George Kapetanios (2009). "Revisiting useful approaches to data-rich macroeconomic forecasting". In: *FRB of New York Staff Report* 327.

Helland, Inge S. (1988). "On the structure of partial least squares regression". In: *Communications in Statistics - Simulation and Computation* 17.2, pp. 581–607.

— (1990). "Partial least squares regression and statistical models". In: *Scandinavian Journal of Statistics* 17.2, pp. 97–114.

Huang, Dashian, Fuwei Jiang, Jun Tu, and Guofu Zhou (2015). "Investor sentiment aligned: A powerful predictor of stock returns." In: *The Review of Financial Studies* 28.3, pp. 791–837.

Kelly, Bryan and Seth Pruitt (2013). "Market expectations in the cross-section of present values". In: *Journal of Finance* 68.5, pp. 1721–1756.

— (2015). "The three-pass regression filter: A new approach to forecasting using many predictors". In: *Journal of Econometrics* 186.2, pp. 294–316.

Light, N., D. Maslov, and O. Rytchkov (2017). "Aggregation of information about the cross section of stock returns: A latent variable approach". In: *Review of Financial Studies* 30.4, pp. 1339–1381.

McCracken, Michael W. and Serena Ng (2016). "FRED-MD: A monthly database for macroeconomic research". In: *Journal of Business & Economic Statistics* 34.4, pp. 574–589.

Mehmood, T., K. H. Liland, L. Snipen, and S. Sæbø (2012). "A review of variable selection methods in Partial Least Squares regression". In: *Chemometrics and Intelligent Laboratory Systems* 118, pp. 62–69.

Moon, Hyunsik R. and Martin Weidner (2015). "Linear regression for panel with unknown number of factors as interactive fixed effects". In: *Econometrica* 83.4, pp. 11543–1579.

Onatski, Alexei (2010). "Determining the number of factors from empirical distribution of eigenvalues". In: *The Review of Economics and Statistics* 92.4, pp. 1004–1016.

Phatak, A. and F. De Hoog (2002). "Exploiting the connection between PLS, Lanczos methods and conjugate gradients: alternative proofs of some properties of PLS". In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 16.7, pp. 361–367.

Rytchkov, O. and X. Zhong (2021). "2019. Information aggregation and P-hacking". In: *Management Science* 66.4, pp. 1605–1626.

Stewart, Gilbert W. and Ji-guang Sun (1990). *Matrix Perturbation Theory*. San Diego: Academic Press.

Stock, James H. and Mark W. Watson (2002a). "Forecasting using principal components from a large number of predictors". In: *Journal of the American Statistical Association* 97.460, pp. 1167–1179.

Stock, James H. and Mark. W. Watson (2002b). "Macroeconomic forecasting using diffusion indexes". In: *Journal of Business & Economic Statistics* 20.2, pp. 147–162.

Stock, James H. and Mark W. Watson (2006). "Forecasting with many predictors". In: *Handbook of Economic Forecasting*. Ed. by G. Elliot, C. Granger, and A. Timmermann. Vol. 1. Elsevier, pp. 515–554.

Taniguchi, M. and P. R. Krishnaiah (1987). "Asymptotic distributions of functions of the eigenvalues of sample covariance matrix and canonical correlation matrix in multivariate time series". In: *Journal of Multivariate Analysis* 22.1, pp. 156–176.

Tu, Yundong and Tae-Hwy Lee (2019). "Forecasting using supervised factor models". In: *Journal of Management Science and Engineering* 4.1, pp. 12–27.

Wold, Herman (1966). "Estimation of principal components and related models by iterative least squares". In: *Multivariate Analysis*. Ed. by P. R. Krishnaiah. Academic Press, pp. 391–420.

— (1973). "Nonlinear iterative partial least squares modelling: Some current developments". In: *Proceedings of the 3rd International Symposium on Multivariate Analysis*. Ed. by P. R. Krishnaiah. Academic Press, pp. 383–407.

— (1982). "Soft modeling: the basic design and some extensions". In: *Systems under Indirect Observation: Causality-Structure-Prediction*. Ed. by K. G. Jorestog and H. Wold. Vol. 2. North-Holland.

Appendix

Appendix A: 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 $\boldsymbol{X}^{(1)} = \boldsymbol{X}$, $\hat{\boldsymbol{\alpha}}_1 = \boldsymbol{X}^{(1)\prime}\boldsymbol{y}$, and $\hat{\boldsymbol{p}}_1 = \boldsymbol{X}^{(1)}\hat{\boldsymbol{\alpha}}_1$. For $j = 2, ..., q$, we iteratively create

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

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 $\boldsymbol{x}_{\cdot T+1}^{(1)} = \boldsymbol{x}_{\cdot T+1}$, $\hat{p}_{1,T+1} = \boldsymbol{x}_{\cdot T+1}^{(1)}{}'\hat{\alpha}_1$, and

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

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

Appendix B: 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 $\boldsymbol{Diag}(\boldsymbol{A}_1, ... \boldsymbol{A}_l)$ is not a square matrix unless 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. Finally, 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

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

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:

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

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 $h = ks(j-1)+1, \ldots, ks(j)$ 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,

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

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

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

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

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

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

Now, observe that

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

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

Appendix C: Proofs of Theorems

Lemma C.1: Define the following orthonormal matrix:

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

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, $\boldsymbol{O}^{\boldsymbol{\Omega_F}} = \boldsymbol{\Xi}(\boldsymbol{\Omega_F}|1:K)$.

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

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

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

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

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

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

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

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

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

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

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

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

Proof: The part (i) holds by (A.6) because

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

Similarly, (ii) holds because

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

The part (iii) holds because

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

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

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

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

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

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

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

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

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

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

(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

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

For the second part of (ii), let

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

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

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

Part (ii) holds by the second part of (i) and Lemma B.3. Finally, let $\tilde{\boldsymbol{\Lambda}} = \boldsymbol{diag}(\tilde{\lambda}_1, ..., \tilde{\lambda}_K) = \boldsymbol{\Lambda}(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}|1:K)$; and $\tilde{\boldsymbol{\Xi}} = \boldsymbol{\Xi}(\tilde{\boldsymbol{\Omega}}_{\Phi}|1:K)$. By (ii), $\tilde{\lambda}_h^{1/2} - 1 = (\tilde{\lambda}_h - 1)/(\tilde{\lambda}_h^{1/2} + 1) = O_p(m^{-\gamma})$, which implies $\left\|\tilde{\boldsymbol{\Lambda}}^{1/2} - \boldsymbol{I}_K\right\|_F = O_p(m^{-\gamma})$. Thus, (iii) holds because

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

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

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

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

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

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

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

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

Proof: We can easily show

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

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 $j = 1,...,J$.

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

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

With this, we can obtain the first result:

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

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

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

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

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

Lemma C.8: Under (A.1) – (A.6),

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

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

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

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

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

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

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

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

For $q \geq K+1$, (iii) holds because $\lambda_{K+1}(\boldsymbol{Z}_{NT}) = 0$. *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 $\boldsymbol{O}_{jj}^{\boldsymbol{S}_{NT}}$ such that $\left\|\boldsymbol{\Xi}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{O}_{jj}^{\boldsymbol{S}_{NT}} - \boldsymbol{\Xi}_{(j)}^{\boldsymbol{Z}_{NT}}\right\|_F = O_p(m^{-1})$.

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

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

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

Proof: Observe that

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

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

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

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

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

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

Lemma C.12: Under (A.1) – (A.7),

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

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

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

In addition,

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

This completes the proof because the Frobenius norms of $\boldsymbol{E}'\boldsymbol{\mathcal{Q}}(\boldsymbol{F})\boldsymbol{u}$ and $\tilde{\boldsymbol{\Phi}}'\boldsymbol{E}'\boldsymbol{\mathcal{Q}}(\boldsymbol{F})\boldsymbol{u}$ are 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) $\left\| (NT)^{-1/2}\boldsymbol{E}'\boldsymbol{y} \right\|_2 = O_p(1)$.

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

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

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,

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

This completes the proof. *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

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

Thus, by Lemma C.13,

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

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

Lemma C.15: Under (A.1) – (A.7), with the matrix $\boldsymbol{O}_{jj}^*$ that is defined in Lemma C.11,

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

(ii) $\left\| \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}} \right\|_2 = O_p(m^{-\gamma})$, for $j \geq R+1$

Proof: For $j \le R$,

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

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

$$
\begin{aligned}
&\left\| \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}} - \boldsymbol{O}_{jj}^{*}{}' \frac{\boldsymbol{\Phi}_{(j)}'}{N^{1/2}} \left(\Sigma_{j'=1}^{R} \frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}} \sigma_{j'}^2 \boldsymbol{\beta}_{(j')} \right) \right\|_F \\
&\quad \le \left\| \boldsymbol{O}_{jj}^{*}{}' \frac{\boldsymbol{\Phi}_{(j)}'}{N^{1/2}} \right\|_F \left\| \boldsymbol{b}_{NT} - \Sigma_{j'=1}^{R} \sigma_{j'}^2 \frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}} \boldsymbol{\beta}_{(j')} \right\|_F \\
&\qquad + \left\| \boldsymbol{\Xi}_{(j)}^{\boldsymbol{S}_{NT}} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right\|_F \left\| \Sigma_{j'=1}^{R} \sigma_{j'}^2 \frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}} \boldsymbol{\beta}_{(j')} \right\|_F \\
&\qquad + \left\| \boldsymbol{\Xi}_{(j)}^{\boldsymbol{S}_{NT}} - \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right\|_F \left\| \boldsymbol{b}_{NT} - \Sigma_{j'=1}^{R} \sigma_{j'}^2 \frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}} \boldsymbol{\beta}_{(j')} \right\|_F \\
&\quad = O_p(T^{-\gamma}) + O_p(m^{-\gamma}) + O_p(T^{-\gamma} m^{-\gamma}) = O_p(m^{-\gamma}).
\end{aligned}
$$

This implies

$$
\left\| \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}} - \boldsymbol{O}_{jj}^{*}{}' \frac{\boldsymbol{\Phi}_{(j)}'}{N^{1/2}} \left(\Sigma_{j'=1}^{R} \frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}} \sigma_{j'}^2 \boldsymbol{\beta}_{(j')} \right) \right\|_F = O_p(m^{-\gamma}). \tag{C.1}
$$

In addition, we can easily show

$$
\left\| \frac{\boldsymbol{\Phi}_{(j)}'}{N^{1/2}} \left(\Sigma_{j'=1}^{R} \frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}} \sigma_{j'}^2 \boldsymbol{\beta}_{(j')} \right) - \frac{\boldsymbol{\Phi}_{(j)}' \boldsymbol{\Phi}_{(j)}}{N^{1/2}} \sigma_j^2 \boldsymbol{\beta}_{(j)} \right\|_F = O_p(m^{-\gamma}); \tag{C.2}
$$

$$
\left\| \frac{\boldsymbol{\Phi}_{(j)}' \boldsymbol{\Phi}_{(j)}}{N} \sigma_j^2 \boldsymbol{\beta}_{(j)} - \sigma_j^2 \boldsymbol{\beta}_{(j)} \right\|_F = O_p(m^{-\gamma}). \tag{C.3}
$$

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

$$
\begin{aligned}
&\left\|\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}-\boldsymbol{O}_{jj}^{*}\sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F \\
&\quad=\left\|\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}-\boldsymbol{O}_{jj}^{*}\sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F \\
&\quad\le\left\|\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}-\boldsymbol{O}_{jj}^{*}{}'\frac{\boldsymbol{\Phi}_{(j)}'}{N^{1/2}}\left(\Sigma_{j'=1}^{R}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma_{j'}^2\boldsymbol{\beta}_{(j')}\right)\right\|_F \\
&\qquad+\left\|\boldsymbol{O}_{jj}^{*}\right\|_F\left\|\frac{\boldsymbol{\Phi}_{(j)}'}{N^{1/2}}\left(\Sigma_{j'=1}^{R}\frac{\boldsymbol{\Phi}_{(j')}}{N^{1/2}}\sigma_{j'}^2\boldsymbol{\beta}_{(j')}\right)-\frac{\boldsymbol{\Phi}_{(j)}'\boldsymbol{\Phi}_{(j)}}{N^{1/2}}\sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F \\
&\qquad+\left\|\boldsymbol{O}_{jj}^{*}\right\|_F\left\|\frac{\boldsymbol{\Phi}_{(j)}'\boldsymbol{\Phi}_{(j)}}{N}\sigma_j^2\boldsymbol{\beta}_{(j)}-\sigma_j^2\boldsymbol{\beta}_{(j)}\right\|_F.
\end{aligned}
$$

Part (ii) can be shown similarly. *Q.E.D.*

Lemma C.16: Under 2.2 – 2.2, for $j=1,...,J$,

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

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

Proof: By Lemma C.11, we can show

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

In addition, we have

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

because

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

Finally, we have

$$\left\| \frac{\boldsymbol{X}}{(NT)^{1/2}} \Xi_{(j)}^{\boldsymbol{S}_{NT}} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \boldsymbol{O}_{jj}^{*} \right\|_F$$
$$\leq \left\| \frac{\boldsymbol{X}}{(NT)^{1/2}} \Xi_{(j)}^{\boldsymbol{S}_{NT}} - \frac{\boldsymbol{X}}{(NT)^{1/2}} \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} \boldsymbol{O}_{jj}^{*} \right\|_F + \left\| \frac{\boldsymbol{X}}{(NT)^{1/2}} \frac{\boldsymbol{\Phi}_{(j)}}{N^{1/2}} - \frac{\boldsymbol{F}_{(j)}}{T^{1/2}} \right\|_F \|\boldsymbol{O}_{jj}^{*}\|_F,$$

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

Lemma C.17: Let $\Xi_H^{\boldsymbol{S}_{NT}} = (\Xi_{(1)}^{\boldsymbol{S}_{NT}}, ..., \Xi_{(J)}^{\boldsymbol{S}_{NT}}) = \Xi(\boldsymbol{S}_{NT}|1:K)$; $\boldsymbol{S}_{NT}^{*} = \boldsymbol{X}'\boldsymbol{X}/(NT)$;$\Xi_H^{*} = \Xi(\boldsymbol{S}_{NT}^{*}|1:K)$; and $\tilde{\boldsymbol{F}} = \boldsymbol{F} + \boldsymbol{E}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}$. Under (A.1) – (A.8), the following holds.

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

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

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,

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

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 $\boldsymbol{O}^{\boldsymbol{S}_{NT}}$ such that

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

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

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

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

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

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

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

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

Lemma C.18: Let $\boldsymbol{H}_{NT} = (NT)^{-1/2}\Xi_L^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\mathcal{Q}}(\boldsymbol{\Phi})\boldsymbol{E}'\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})$, where $\tilde{\boldsymbol{F}}$ is defined in Lemma C.17. 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, under 2.2 – 2.2, the following holds.

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

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

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

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

$$
\begin{aligned}
&\boldsymbol{c}_L^{\boldsymbol{S}_{NT}} - T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u} \\
&= \Xi_L^{\boldsymbol{S}_{NT}\prime}\frac{\boldsymbol{X}'}{(NT)^{1/2}}\boldsymbol{\mathcal{Q}}(\Xi_H^*)\frac{\boldsymbol{y}}{T^{1/2}} + \Xi_L^{\boldsymbol{S}_{NT}\prime}\frac{\boldsymbol{X}'}{(NT)^{1/2}}\boldsymbol{\mathcal{P}}(\Xi_H^*)\frac{\boldsymbol{y}}{T^{1/2}} - T^{-1/2}\boldsymbol{H}_{NT}\boldsymbol{u} \\
&= \Xi_L^{\boldsymbol{S}_{NT}\prime}\frac{\boldsymbol{X}'}{(NT)^{1/2}}\boldsymbol{\mathcal{Q}}(\Xi_H^*)\frac{\boldsymbol{y}}{T^{1/2}} - \frac{\boldsymbol{H}_{NT}\boldsymbol{u}}{T^{-1/2}} \\
&= \Xi_L^{\boldsymbol{S}_{NT}\prime}\frac{\boldsymbol{X}'}{(NT)^{1/2}}\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})\frac{\boldsymbol{y}}{T^{1/2}} + \Xi_L^{\boldsymbol{S}_{NT}}\frac{\boldsymbol{X}'}{(NT)^{1/2}}\left(\boldsymbol{\mathcal{Q}}(\Xi_H^*) - \boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})\right)\frac{\boldsymbol{y}}{T^{1/2}} - \frac{\boldsymbol{H}_{NT}\boldsymbol{u}}{T^{-1/2}} \\
&= \frac{1}{T^{1/2}}\Xi_L^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\mathcal{Q}}(\boldsymbol{\Phi})\frac{\boldsymbol{E}'}{(NT)^{1/2}}\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})\boldsymbol{u} - \frac{\boldsymbol{H}_{NT}\boldsymbol{u}}{T^{-1/2}} \\
&\quad - \frac{1}{N^{1/2}}\Xi_L^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\mathcal{Q}}(\boldsymbol{\Phi})\frac{\boldsymbol{E}'}{(NT)^{1/2}}\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})\frac{\boldsymbol{E}\boldsymbol{\Phi}}{(NT)^{1/2}}\left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}\right)^{-1}\boldsymbol{\beta} \\
&\quad + \Xi_L^{\boldsymbol{S}_{NT}\prime}\frac{\boldsymbol{X}'}{(NT)^{1/2}}\left(\boldsymbol{\mathcal{Q}}(\Xi_H^*) - \boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})\right)\frac{\boldsymbol{y}}{T^{1/2}} \\
&= -\frac{1}{N^{1/2}}\Xi_L^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\mathcal{Q}}(\boldsymbol{\Phi})\frac{\boldsymbol{E}'}{(NT)^{1/2}}\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})\frac{\boldsymbol{E}\boldsymbol{\Phi}}{(NT)^{1/2}}\left(\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}\right)^{-1}\boldsymbol{\beta} \\
&\quad + \Xi_L^{\boldsymbol{S}_{NT}\prime}\frac{\boldsymbol{X}'}{(NT)^{1/2}}\left(\boldsymbol{\mathcal{Q}}(\Xi_H^*) - \boldsymbol{\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 $\boldsymbol{E}'\boldsymbol{E} - \boldsymbol{E}'\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{F}})\boldsymbol{E}$ is positive semi-definite and $\boldsymbol{\mathcal{Q}}(\boldsymbol{\Phi})$ is idempotent, we have

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

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

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

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

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

With these results, we can show that

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

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

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

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

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

By these results, we have

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

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

Proof of Lemma 2.4.1: The parts (i) and (ii) hold by Lemma C.9. The part (iii) holds by Lemma C.14. The part (iv) holds by Lemma C.11, while the parts (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 , $\lambda^{\boldsymbol{S}_{NT}}_{ks(j-1)+h} = \sigma^2_j + O_p(m^{-\gamma})$, for $j = 1, 2, ..., J$ and $h = 1, 2, ..., k(j)$. Cnosequently, $\lambda^{\boldsymbol{S}_{NT}}_{ks(j-1)+h} - \lambda^{\boldsymbol{S}_{NT}}_{ks(j-1)+1} = O_p(m^{-\gamma})$. Observe that

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

Therefore,

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

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

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

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

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

Finally, by Lemma C.18, we have

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

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

Corollary C.19: Under (A.1) – (A.8), 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

$$\|\boldsymbol{V}_{H1}(q)\|_F \le \Sigma_{j=1}^R \|\boldsymbol{v}_{H1}(j)\|_2\,;$$
$$\|\boldsymbol{V}_{H2}(q)\|_F \le \Sigma_{j=1}^R \|\boldsymbol{v}_{H2}(j)\|_2\,;$$
$$\|\boldsymbol{V}_{L}(q)\|_F \le q \times \|\boldsymbol{v}_{L}(1)\|_2 \qquad Q.E.D.$$

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

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

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

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

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

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

For (ii), observe that

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

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

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

Lemma C.20: The following equalities hold:

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

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

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

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

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

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

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

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

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

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

Finally, we can show

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

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

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

Lemma C.21: Under (A.1) – (A.8),

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

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

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

This implies (i) because

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

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

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

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

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

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

(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 $q = 1$, $\boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(1) = 0_{T\times 1}$. For $q \geq 2$,

$$\|\boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q)\|_2 \leq \|\boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\|_2 \left\|\frac{\boldsymbol{X}}{(NT)^{1/2}}\right\|_2 \|\boldsymbol{v}_{H1}(q)\|_2$$
$$= \left\|\frac{\boldsymbol{X}}{(NT)^{1/2}}\right\|_2 \|\boldsymbol{v}_{H1}(q)\|_2 = O_p(1) \times O_p(m^{-\gamma}),$$

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) for $q \geq 2$. Part (ii) holds by Lemma 2.4.2 because

$$\left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0(q)\right\|_2 \leq \left\|\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}\boldsymbol{G}_0\right\|_2 \|\boldsymbol{g}_{H1}(q)\|_2 = O_p(m^{-\gamma}).$$

Part (ii) holds by Lemma 2.4.1 because

$$\|\boldsymbol{g}_L(q)\|_2 \leq \left\|\frac{\boldsymbol{X}}{(NT)^{1/2}}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}}\right\|_2 \left\|\boldsymbol{\Lambda}_L^{\boldsymbol{S}_{NT}}\right\|_2^{q-1} \left\|\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}\right\|_2$$
$$= O_p(m^{-(q-1/2)}(T^{-1/2} + m^{-3/2})). \qquad (Q.E.D.)$$

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

$$\rho_{j,q} = \frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\left(\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\left((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1}\right)\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\right)}{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}$$

Under (A.1) – (A.8), for $j = 1, 2, ..., R$ and $q = 1, 2, ...,$

$$\rho_{j,q} = \frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\left((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1}\right)\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\mu_j^{\boldsymbol{S}_{NT}}} + O_p(m^{-2\gamma}).$$

Proof: Observe that by Lemma 2.4.1, $\left\|\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}} - \bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}}\right\|_F = O_p(m^{-\gamma})$ for $j = 1, ..., R$. With this result, we can show

$$\begin{aligned}
&\rho_{j,q} - \frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\left((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1}\right)\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\mu_j^{\boldsymbol{S}_{NT}}} \\
&= \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\left(\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - \bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\right) \\
&\quad \times \left(\frac{1}{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}} - \frac{1}{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}\right) \\
&= \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}(\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}} - \boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\left(\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\right)\left(\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\right)} \\
&\quad \times \left(\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\right) \\
&= O_p(m^{-2\gamma})
\end{aligned}$$

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

$$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q)\right\|_2 = O_p(m^{-2\gamma}); \tag{C.8}$$

$$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_{H2}(q)\right\|_2 = O_p(m^{-2\gamma}). \tag{C.9}$$

Consider τ_q defined in Lemma C.20. By Lemma C.23 and Lemma 2.4.1, for $q = 1, 2, ..., R$,

$$\begin{aligned}\tau_q &= \Sigma_{j=1}^{R} \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}} - \Sigma_{j=1}^{R}\rho_{j,q} \\ &= \Sigma_{j=1}^{R} \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}} \\ &\quad - \Sigma_{j=1}^{R} \frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1})\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\mu_j^{\boldsymbol{S}_{NT}}} + O_p(m^{-2\gamma}) \\ &= \Sigma_{j=1}^{R} \frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}(\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}} - \boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})\left((\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1} - (\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1}\right)\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\mu_j^{\boldsymbol{S}_{NT}}} \\ &\quad + O_p(m^{-2\gamma}) = O_p(m^{-2\gamma}).\end{aligned}$$

Thus, (C.8) holds because, by Lemma C.20, $T^{-1/2}\boldsymbol{y}'\boldsymbol{\mathcal{Q}}(\boldsymbol{G}_0)\boldsymbol{g}_{H1}(q) = \tau_q$. Finally, by Lemma C.20 and Lemma 2.4.1, (C.9) also holds because

$$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_{H2}(q)\right\|_2 = \Sigma_{j=R+1}^{J}\left\|(\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}})^{q-1}\right\|_2 \left\|\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\right\|_2^2 = O_p(m^{-2\gamma}).$$

We can obtain (ii) because

$$\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(q)\right\|_2 = \left\|\boldsymbol{c}_L^{\boldsymbol{S}_{NT}\prime}(\boldsymbol{\Lambda}_L^{\boldsymbol{S}_{NT}})^{q-1}\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}\right\|_2 \leq O_p(m^{-(q-1)})\left\|\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}\right\|_2^2 \qquad Q.E.D.$$

Corollary C.24: Under (A.1) – (A.8), for $q \geq 1$,

(i) $\left\|T^{-1/2}\boldsymbol{y}_H^{\boldsymbol{G}c}(q)\right\|_2 = O_p(m^{-2\gamma})$;

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

Proof: Observe that

$$\begin{aligned}&\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_H^c(q)\right\|_2 \leq \Sigma_{j=1}^{q}\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_H^c(j)\right\|_2; \\ &\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{G}_L(q)\right\|_2 \leq \Sigma_{j=1}^{R}\left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(q)\right\|_2 \leq q \times \left\|T^{-1/2}\boldsymbol{y}'\boldsymbol{g}_L(1)\right\|_2.\end{aligned}$$

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

$$\begin{aligned}\left\|(\hat{\boldsymbol{D}}_0(R))^{-1} - (\boldsymbol{D}_0(R))^{-1}\right\|_F &= \left\|(\hat{\boldsymbol{D}}_0(R))^{-1}\right\|_F \left\|\hat{\boldsymbol{D}}_0(R) - \boldsymbol{D}_0(R)\right\|_F \left\|(\boldsymbol{D}_0(R))^{-1}\right\|_F \\ &= O_p(m^{-\gamma}). \qquad (Q.E.D.)\end{aligned}$$

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 $\hat{\boldsymbol{\Sigma}}_R=\boldsymbol{diag}(\mu_1^{\boldsymbol{S}_{NT}},...,\mu_R^{\boldsymbol{S}_{NT}})$. Observe that for $j=1,...,R$,

$$\left|\frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}}}-\frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\mu_j^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}\right|=O_p(m^{-\gamma}) \tag{C.10}$$

because

$$\begin{aligned}&\left|\frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}-\frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\mu_j^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}\right|\\
&\quad\le\frac{\left|\mu_j^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}-\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}}\right|}{(\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}})\mu_j^{\boldsymbol{S}_{NT}}}\\
&\quad\le\left\|\frac{1}{(\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}})\mu_j^{\boldsymbol{S}_{NT}}}\right\|_2\left\|\bar{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}}-\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\right\|_F\left\|\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}\right\|_2^2\\
&\quad=O_p(1)\times O_p(m^{-\gamma})\times O_p(1)=O_p(m^{-\gamma}).\end{aligned}$$

With (C.10), we can have

$$\left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}-\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R\right\|_2=O_p(m^{-\gamma}), \tag{C.11}$$

because

$$\begin{aligned}&\left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}\frac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}}-\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R\right\|_2\\
&\quad=\left\|\begin{pmatrix}\frac{\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}}}{\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(1)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}}}\\ \vdots\\ \frac{\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}}}{\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(R)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}}}\end{pmatrix}-\begin{pmatrix}\frac{\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}}}{\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}}}\\ \vdots\\ \frac{\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}}}{\mu_R^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}}}\end{pmatrix}\right\|_2\\
&\quad\le\Sigma_{j=1}^R\left|\frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(j)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}}}-\frac{\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}{\mu_j^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}}}\right|.\end{aligned}$$

By Lemma 2.4.1, we have $1/\mu_j^{\boldsymbol{S}_{NT}}-1/\sigma_j^2=(\sigma_j^2-\mu_j^{\boldsymbol{S}_{NT}})/(\mu_j^{\boldsymbol{S}_{NT}}\sigma_j^2)=O_p(m^{-\gamma})$. Thus,

$$\left\|\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R-\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2=O_p(m^{-\gamma}). \tag{C.12}$$

By (C.11) and (C.12), we can obtain the desired result because

$$\begin{aligned}&\left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}-\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2\\
&\quad\le\left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}-\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R\right\|_2+\left\|\hat{\boldsymbol{\Sigma}}_R^{-1}\mathbf{1}_R-\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R\right\|_2 \qquad (Q.E.D.)\end{aligned}$$

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:

$$\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\tilde{\boldsymbol{P}}_{1:R}^{PLS}}{NT} - \hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\boldsymbol{D}_0(R) = \boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c\hat{}(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R);$$
$$\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\tilde{\boldsymbol{p}}_{R+1}^{PLS}}{NT} - \hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) = \boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1);$$
$$\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\boldsymbol{y}}{N^{1/2}T} - \hat{\boldsymbol{D}}_0(R)'T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} = T^{-1/2}\boldsymbol{G}_H^c(R)'\boldsymbol{y} + T^{-1/2}\boldsymbol{G}_L(R)'\boldsymbol{y};$$
$$\frac{\tilde{\boldsymbol{p}}_R^{PLS}{}'\boldsymbol{y}}{N^{1/2}T} - \hat{\boldsymbol{d}}_0(R)'\frac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}} = \frac{\boldsymbol{g}_H^c(R)'\boldsymbol{y}}{T^{1/2}} + \frac{\boldsymbol{g}_L(R)'\boldsymbol{y}}{T^{1/2}};$$
$$\frac{\tilde{\boldsymbol{p}}_{R+1}^{PLS}{}'\boldsymbol{y}}{N^{1/2}T} - \hat{\boldsymbol{d}}_0(R+1)'\frac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}} = \frac{\boldsymbol{g}_H^c(R+1)'\boldsymbol{y}}{T^{1/2}} + \frac{\boldsymbol{g}_L(R+1)'\boldsymbol{y}}{T^{1/2}}.$$

Lemma C.27: Under (A.1) – (A.8),

$$\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)\|_F = O_p(m^{-2\gamma});$$
$$\left\|T^{-1}\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1)\right\|_2 = O_p(m^{-2\gamma}); \quad \left\|T^{-1}\boldsymbol{G}_H^c(R)'\boldsymbol{y}\right\|_2 = O_p(m^{-2\gamma});$$
$$\left\|T^{-1/2}\boldsymbol{g}_H^c(R)'\boldsymbol{y}\right\|_2 = O_p(m^{-2\gamma}); \quad \left\|T^{-1/2}\boldsymbol{g}_H^c(R+1)'\boldsymbol{y}\right\|_2 = O_p(m^{-2\gamma}).$$

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),

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

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 $\boldsymbol{d}_0^*(q) = \boldsymbol{\Sigma}_R^{q-1}\boldsymbol{1}_R$ and $\boldsymbol{D}_0^*(q) = (\boldsymbol{d}_0^*(1), ..., \boldsymbol{d}_0^*(q))$. Under (A.1) – (A.8), as $m \to \infty$,

(i) $\hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) \ \to_p \ \boldsymbol{\Psi}^* \equiv \boldsymbol{D}_0^*(R)'\boldsymbol{Diag}(\sigma_1^6\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)}, ..., \sigma_R^6\boldsymbol{\beta}_{(R)}'\boldsymbol{\beta}_{(R)})\boldsymbol{D}_0^*(R);$

(ii) $\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) \ \to_p \ \boldsymbol{\psi}^* \equiv \boldsymbol{D}_0^*(R)'\boldsymbol{diag}(\sigma_1^6\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)}, ..., \sigma_R^6\boldsymbol{\beta}_{(R)}'\boldsymbol{\beta}_{(1)})\boldsymbol{d}_0^*(R+1)$

(iii) $T^{-1/2}\hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{y} \ \to_p \ \boldsymbol{\pi}^* \equiv \boldsymbol{D}_0^*(R)'\left(\sigma_1^4\boldsymbol{\beta}_{(1)}'\boldsymbol{\beta}_{(1)},...,\sigma_R^4\boldsymbol{\beta}_{(R)}'\boldsymbol{\beta}_{(R)}\right)'.$

Proof: By Lemma 2.4.1, for $j = 1,...,R$, we have

$$\hat{\boldsymbol{\Lambda}}_{(j)}^{\boldsymbol{S}_{NT}} \ \to_p \ \sigma_j^2\boldsymbol{I}_{k(j)}; \quad \boldsymbol{c}_{(j)}^{\boldsymbol{S}_{NT}} \ \to_p \ \boldsymbol{O}_{jj}^*\sigma_j^2\boldsymbol{\beta}_{(j)}. \tag{C.13}$$

In addition, it is straightforward to show

$$\boldsymbol{G}_0'\boldsymbol{G}_0 = \boldsymbol{V}_0'\boldsymbol{\Lambda}_{H1}^{\boldsymbol{S}_{NT}}\boldsymbol{V}_0 = \boldsymbol{diag}(\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(1)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}},...,\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Lambda}_{(R)}^{\boldsymbol{S}_{NT}}\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}}); \tag{C.14}$$

$$\frac{\boldsymbol{G}_0'\boldsymbol{y}}{T^{1/2}} = \begin{pmatrix} \boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Xi}_{(1)}^{\boldsymbol{S}_{NT}\prime} \\ \vdots \\ \boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{\Xi}_{(R)}^{\boldsymbol{S}_{NT}\prime} \end{pmatrix}; \quad \frac{\boldsymbol{X}'\boldsymbol{y}}{N^{1/2}T} = \begin{pmatrix} \boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(1)}^{\boldsymbol{S}_{NT}} \\ \vdots \\ \boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}\prime}\boldsymbol{c}_{(R)}^{\boldsymbol{S}_{NT}} \end{pmatrix}. \tag{C.15}$$

By Lemmas 2.4.1 and 2.4.3,

$$\left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0^*(q)\right\| \le \left\|\hat{\boldsymbol{d}}_0(q) - \boldsymbol{d}_0(q)\right\| + \|\boldsymbol{d}_0(q) - \boldsymbol{d}_0^*(q)\| = O_p(m^{-\gamma}). \tag{C.16}$$

The desired results can be obtained by (C.13) – (C.16) and Lemma 2.4.1. *Q.E.D.*

Proof of Theorem 2: We begin by considering the cases in which $R < K$. Observe that

$$(\hat{\boldsymbol{D}}_0(R))^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} = \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\hat{\boldsymbol{D}}_0(R)'T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}. \tag{C.17}$$

With this result and Lemma 2.4.4, we can show

$$\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\hat{\boldsymbol{D}}_0(R))^{-1}\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 = O_p(m^{-2\gamma}). \tag{C.18}$$

Lemma C.26 and (C.17) imply

$$\begin{aligned}
&\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\hat{\boldsymbol{D}}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\quad \le \left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\hat{\boldsymbol{D}}_0^(R))^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y}\right\|_2 \\
&\qquad + \left\|(\hat{\boldsymbol{D}}_0(R))^{-1}\right\|_F \left\|(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} - \boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\quad = O_p(m^{-2\gamma}) + O_p(m^{-\gamma}) = O_p(m^{-\gamma}).
\end{aligned}$$

By this result and Lemma C.25, we can obtain (i) because

$$\begin{aligned}
&\left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\quad = \left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\hat{\boldsymbol{D}}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R + (\hat{\boldsymbol{D}}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\quad \le \left\|N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\hat{\boldsymbol{D}}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 + \left\|(\hat{\boldsymbol{D}}_0(R))^{-1} - (\boldsymbol{D}_0(R))^{-1}\right\|_F \left\|\boldsymbol{\Sigma}_R^{-1}\boldsymbol{1}_R\right\|_2 \\
&\quad = O_p(m^{-\gamma}) + O_p(m^{-\gamma}) = O_p(m^{-\gamma}).
\end{aligned}$$

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

$$\begin{aligned}
&\left\| N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R \right\|_2 \\
&\quad \leq \left\| N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} - (\boldsymbol{D}_0(R))^{-1}(\boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} \right\|_2 \\
&\qquad + \left\|(\boldsymbol{D}_0(R))^{-1}\right\|_F \left\| \boldsymbol{G}_0'\boldsymbol{G}_0)^{-1}T^{-1/2}\boldsymbol{G}_0'\boldsymbol{y} - \boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R \right\|_2 \\
&\quad = O_p\left(m^{-R-1}(T^{-1/2}+m^{-3/2})^2\right) + O_p(m^{-\gamma}) = O_p(m^{-\gamma}).
\end{aligned}$$

Part (ii) is obtained by Theorem 1 and (i) because

$$\begin{aligned}
\left\| \tilde{y}^{PLS}_{T+2|R} - \hat{y}^*_{T+2} \right\|_2 &= \left\| \frac{\boldsymbol{x}'_{\cdot T+1}}{N^{1/2}} \tilde{\boldsymbol{A}}^{PLS}_{1:R} N^{1/2}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R} - \hat{y}^*_{T+2} \right\|_2 \\
&= \left\| \boldsymbol{g}'_{T+1}\boldsymbol{\Sigma}_R\boldsymbol{D}_0(R)(\boldsymbol{D}_0(R))^{-1}\boldsymbol{\Sigma}_R^{-1}\mathbf{1}_R - \hat{y}^*_{T+2} \right\|_2 + O_p(m^{-\gamma}) \\
&= 0 + O_p(m^{-\gamma}).
\end{aligned}$$

Finally, for (iii), observe that by Lemmas 2.4.4 and C.29, we have

$$\begin{aligned}
\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}}{N^{1/2}T} &= T^{-1/2}(\hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{y} + \boldsymbol{G}^c_H(R)'\boldsymbol{y} + \boldsymbol{G}_L(R)'\boldsymbol{y}) \\
&\rightarrow_p \ \boldsymbol{D}^*_0(R)\left(\sigma_1^4\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}, ..., \sigma_1^4\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}\right)';
\end{aligned}$$

$$\begin{aligned}
\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R})}{NT} &= \hat{\boldsymbol{D}}_0(R)'\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) + \boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R) \\
&\rightarrow_p \ \boldsymbol{D}^*_0(R)'\boldsymbol{diag}(\sigma_1^6\boldsymbol{\beta}'_{(1)}\boldsymbol{\beta}_{(1)}, ..., \sigma_R^6\boldsymbol{\beta}'_{(R)}\boldsymbol{\beta}_{(R)})\boldsymbol{D}^*_0(R).
\end{aligned}$$

By these two results, we can obtain

$$\frac{\boldsymbol{y}'\boldsymbol{\mathcal{P}}|(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}}{T} = \frac{\boldsymbol{y}'\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{N^{1/2}T}\left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}}{N^{1/2}T} \rightarrow_p \Sigma_{j=1}^R \sigma_j^2\boldsymbol{\beta}'_{(j)}\boldsymbol{\beta}_{(j)},$$

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

Lemma C.30: Let $\tilde{\boldsymbol{\theta}} \equiv (\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R})^{-1}\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1}$. 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

$$\left(\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\tilde{\boldsymbol{P}}_{1:R}^{PLS}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}_{1:R}^{PLS}{}'\tilde{\boldsymbol{p}}_{R+1}^{PLS}}{NT}$$
$$= (\hat{\boldsymbol{D}}_0(R))^{-1}\hat{\boldsymbol{d}}_0(R+1) + \boldsymbol{A}_1\left(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right)$$
$$+ \boldsymbol{A}_1\left(\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)\boldsymbol{a}_2,$$

where

$$\boldsymbol{A}_1 = \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1};$$
$$\boldsymbol{a}_2 = \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\boldsymbol{D}_0(R) + \boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right)^{-1}$$
$$\times\left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) + \boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\right).$$

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

$$\left\|\tilde{\boldsymbol{\theta}} - (\hat{\boldsymbol{D}}_0(R))^{-1}\hat{\boldsymbol{d}}_0(R+1)\right\|_2 = O_p\left(\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_F\right). \tag{C.19}$$

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 $\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R) = \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)$. By substituting these results into (C.21) and applying Lemma C.28, we can obtain (ii).

For (i), observe that $\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_F = O_p(m^{-2\gamma})$ because $\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)$ asymptotically dominates $\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)$. Thus, from (C.19), we have

$$\left\|\tilde{\boldsymbol{\theta}} - (\hat{\boldsymbol{D}}_0(R))^{-1}\hat{\boldsymbol{d}}_0(R+1)\right\|_2 = O_p(m^{-2\gamma}). \tag{C.20}$$

Now, by Lemma C.25 and Corollary 2.4.3, we can show

$$\left\|(\hat{\boldsymbol{D}}_0(R))^{-1}\hat{\boldsymbol{d}}_0(R+1) - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1)\right\|_2$$
$$\leq \left\|(\hat{\boldsymbol{D}}_0(R))^{-1}\hat{\boldsymbol{d}}_0(R+1) - (\hat{\boldsymbol{D}}_0(R))^{-1}\boldsymbol{d}_0(R+1)\right\|_2$$
$$+ \left\|(\hat{\boldsymbol{D}}_0(R))^{-1}\boldsymbol{d}_0(R+1) - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1)\right\|_2$$
$$= \left\|(\hat{\boldsymbol{D}}_0(R))^{-1}\right\|_F\left\|\hat{\boldsymbol{d}}_0(R+1) - \boldsymbol{d}_0(R+1)\right\|_2$$
$$+ \left\|(\hat{\boldsymbol{D}}_0(R))^{-1} - (\boldsymbol{D}_0(R))^{-1}\right\|_F\left\|\boldsymbol{d}_0^{(}R+1)\right\|_2$$
$$= O_p(m^{-\gamma}).$$

which, with (C.19), implies

$$\left\|\tilde{\boldsymbol{\theta}} - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1)\right\|_2$$
$$\leq \left\|\tilde{\boldsymbol{\theta}} - (\hat{\boldsymbol{D}}_0(R))^{-1}\hat{\boldsymbol{d}}_0(R+1)\right\|_2 + \left\|(\hat{\boldsymbol{D}}_0(R))^{-1}\hat{\boldsymbol{d}}_0(R+1) - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1)\right\|_2$$
$$= O_p(m^{-2\gamma}) + O_p(m^{-\gamma}) = O_p(m^{-\gamma}).$$

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

Lemma C.31: Define $\mathcal{Y}_{1,NT} = (NT)^{-1}\tilde{\boldsymbol{p}}_{R+1}^{PLS\prime}\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{P}}_{1:R}^{PLS})\tilde{\boldsymbol{p}}_{R+1}^{PLS}$. Under (A.1) – (A.8),

$$\begin{aligned}\mathcal{Y}_{1,NT} = &\, \boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\\ &- (\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R))\,\boldsymbol{a}_3\\ &- \boldsymbol{a}_4'\,(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1))\\ &+ \boldsymbol{a}_4'\,(\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R))\,\boldsymbol{a}_2,\end{aligned}$$

where $\boldsymbol{a}_2$ 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 - (\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R))\left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\\
&\quad\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\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(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\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}(\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R))\\
&\quad\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\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 - (\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R))\,\boldsymbol{a}_3\\
&\quad - \boldsymbol{a}_4'\,(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1))\\
&\quad + \boldsymbol{a}_4'\,(\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R))\,\boldsymbol{a}_2,\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{a}_3 &= \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1}\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{d}}_0(R+1) \to_p (\boldsymbol{\Psi}^*)^{-1}\boldsymbol{\psi}^*;\\
\boldsymbol{a}_4' &= \left(\hat{\boldsymbol{d}}_0(R+1)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R) + \boldsymbol{g}_H^*(R+1)'\boldsymbol{G}_H^*(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\right)\\
&\quad \times \left(\hat{\boldsymbol{D}}_0(R)\boldsymbol{G}_0'\boldsymbol{G}_0\hat{\boldsymbol{D}}_0(R)\right)^{-1} \ \to_p \ \boldsymbol{\psi}^{*\prime}(\boldsymbol{\Psi}^*)^{-1},\end{aligned}$$

by Lemmas C.29 and 2.4.3. *Q.E.D.*

Corollary C.31: Under (A.1) – (A.8),

(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 $R < K$,

$$
\begin{aligned}
&\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\|_2 \\
&\quad \leq \|\boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1)\|_2 + \|\boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\|_2 \\
&\quad = O_p(m^{-2\gamma}) + O_p(m^{-2(R+1/2)})O_p\left((T^{-1/2} + m^{-3/2})^2\right) = O_p(m^{-2\gamma}); \\
&\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\|_2 \\
&\quad \leq \|\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R)\|_2 + \|\boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\|_2 \\
&\quad = O_p(m^{-2\gamma}) + O_p(m^{-R-1})O_p\left((T^{-1/2} + m^{-3/2})^2\right) = O_p(m^{-2\gamma}); \\
&\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_F \\
&\quad \leq \|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R)\|_F + \|\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_F \\
&\quad = O_p(m^{-2\gamma}) + O_p\left(m^{-1}(T^{-1/2} + m^{-3/2})^2\right) = O_p(m^{-2\gamma})
\end{aligned}
$$

which imply (i). When $R = K$,

$$
\begin{aligned}
&\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{g}_H^c(R+1) + \boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\|_2 \\
&\quad = \|\boldsymbol{g}_L(R+1)'\boldsymbol{g}_L(R+1)\|_2 = O_p(m^{-2(R+1/2)})O_p\left((T^{-1/2} + m^{-3/2})^2\right); \\
&\|\boldsymbol{g}_H^c(R+1)'\boldsymbol{G}_H^c(R) + \boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\|_2 \\
&\quad \leq \|\boldsymbol{g}_L(R+1)'\boldsymbol{G}_L(R)\|_2 = O_p(m^{-R-1})O_p\left((T^{-1/2} + m^{-3/2})^2\right); \\
&\|\boldsymbol{G}_H^c(R)'\boldsymbol{G}_H^c(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_2 \\
&\quad = \|\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_2 = O_p(m^{-1})O_p\left((T^{-1/2} + m^{-3/2})^2\right)
\end{aligned}
$$

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

Lemma C.32: Define $\mathcal{Y}_{2,NT} = \boldsymbol{y}'\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{P}}_{1:R}^{PLS})\tilde{\boldsymbol{p}}_{R+1}^{PLS}\boldsymbol{y}/(N^{1/2}T)$. 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}}_{R+1}^{PLS}}{N^{1/2}T} - \frac{\boldsymbol{y}'\tilde{\boldsymbol{P}}_{1:R}^{PLS}}{N^{1/2}T}\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}\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}_H^c(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}_H^c(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\quad \times\left(\boldsymbol{G}_H^c(R)'\boldsymbol{g}_H^c(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}_H^c(R)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}}\right) \\
&\quad\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^*(R)'\boldsymbol{G}_H^*(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\right) \\
&\quad\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\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}_H^c(R+1)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}} - \left(\frac{\boldsymbol{y}'\boldsymbol{G}_H^c(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) $\mathcal{Y}_{2,NT} = O_p(m^{-2\gamma})$, when $R < K$;

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

Proof: When $R < K$,

$$\begin{aligned}
&\left\| \frac{\boldsymbol{y}'\boldsymbol{g}^c_H(R+1)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}} \right\|_2 \\
&\quad \leq \left\| \frac{\boldsymbol{y}'\boldsymbol{g}^c_H(R+1)}{T^{1/2}} \right\|_2 + \left\| \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}} \right\|_2 \\
&\quad = O_p(m^{-2\gamma}) + O_p(m^{-R})O_p((T^{-1} + m^{-3/2})^2) = O_p(m^{-2\gamma}); \\
&\left\| \frac{\boldsymbol{y}'\boldsymbol{G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}} \right\|_2 \\
&\quad \leq \left\| \frac{\boldsymbol{y}'\boldsymbol{G}^c_H(R)}{T^{1/2}} \right\|_2 + \left\| \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}} \right\|_2 \\
&\quad = O_p(m^{-2\gamma}) + O_p((T^{-1} + m^{-3/2})^2)O_p(m^{-2\gamma}); \\
&\|\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\|_2 \\
&\quad \leq \|\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1)\|_2 + \|\boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\|_2 \\
&\quad = O_p(m^{-2\gamma}) + O_p(m^{-R-1})O_p((T^{-1} + m^{-3/2})^2) = O_p(m^{-2\gamma}); \\
&\|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_F \\
&\quad \leq \|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R)\|_F + \|\boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_F \\
&\quad = O_p(m^{-2\gamma}) + O_p(m^{-1})O_p((T^{-1} + m^{-3/2})^2) = O_p(m^{-2\gamma}),
\end{aligned}$$

which imply (i). When $R = K$,

$$\begin{aligned}
&\left\| \frac{\boldsymbol{y}'\boldsymbol{g}^c_H(R+1)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}} \right\|_2 \\
&\quad = \left\| \frac{\boldsymbol{y}'\boldsymbol{g}_L(R+1)}{T^{1/2}} \right\|_2 = O_p(m^{-R})O_p\left((T^{-1/2} + m^{-3/2})^2\right) \\
&\left\| \frac{\boldsymbol{y}'\boldsymbol{G}^c_H(R)}{T^{1/2}} + \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}} \right\|_2 = \left\| \frac{\boldsymbol{y}'\boldsymbol{G}_L(R)}{T^{1/2}} \right\|_2 = O_p\left((T^{-1/2} + m^{-3/2})^2\right); \\
&\|\boldsymbol{G}^c_H(R)'\boldsymbol{g}^c_H(R+1) + \boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\|_2 \\
&\quad = \|\boldsymbol{G}_L(R)'\boldsymbol{g}_L(R+1)\|_2 = O_p(m^{-R-1})O_p\left((T^{-1/2} + m^{-3/2})^2\right); \\
&\|\boldsymbol{G}^c_H(R)'\boldsymbol{G}^c_H(R) + \boldsymbol{G}_L(R)'\boldsymbol{G}_L(R)\|_F = O_p(m^{-1})O_p\left((T^{-1/2} + m^{-3/2})^2\right)
\end{aligned}$$

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

Proof of Lemma 2.4.5: Parts (i) and (iii) hold by Lemma C.30. Parts (ii) and (vi) hold by Corollaries C.31 and C.32. *Q.E.D.*

Proof of Theorem 3: Observe that

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

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

$$N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R+1} = \left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R+1}\tilde{\boldsymbol{P}}^{PLS}_{1:R+1}}{NT}\right)^{-1}\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R+1}\boldsymbol{y}}{N^{1/2}T}$$
$$= \begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \end{pmatrix}^{-1}\begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}}{N^{1/2}T} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\boldsymbol{y}}{N^{1/2}T} \end{pmatrix}. \tag{C.21}$$

By the inversion rule for partitioned matrix,

$$\begin{pmatrix} \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \\ \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT} & \frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT} \end{pmatrix}^{-1}$$
$$\begin{pmatrix} \left(\frac{\tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\tilde{\boldsymbol{P}}^{PLS}_{1:R}}{NT}\right)^{-1} & \boldsymbol{0} \\ \boldsymbol{0}' & 0 \end{pmatrix} + \begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}\left(\frac{\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\tilde{\boldsymbol{p}}^{PLS}_{R+1}}{NT}\right)^{-1}\begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}' \tag{C.22}$$

where $\tilde{\boldsymbol{\theta}}$ is defined in Lemma 2.4.5. In addition,

$$\begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix}'\begin{pmatrix} \tilde{\boldsymbol{P}}^{PLS\prime}_{1:R}\boldsymbol{y}/(N^{1/2}T) \\ \tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\boldsymbol{y}/(N^{1/2}T) \end{pmatrix} = -\tilde{\boldsymbol{p}}^{PLS\prime}_{R+1}\boldsymbol{\mathcal{Q}}(\tilde{\boldsymbol{P}}^{PLS}_{1:R})\boldsymbol{y}/(N^{1/2}T). \tag{C.23}$$

We can obtain the desired result by substituting (C.22) and (C.23) into (C.21). *Q.E.D.*

Proof of Theorem 4: Part (i) holds by Lemma C.34. For (ii), observe that

$$\begin{pmatrix} (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix}' \begin{pmatrix} \boldsymbol{D}_0(R)' \\ \boldsymbol{d}_0(R+1)' \end{pmatrix} = \boldsymbol{0}_{1\times R}.$$

Using this fact and (i), we can easily show

$$\begin{aligned}
&\tilde{y}^{PLS}_{T+2|R+1} - \tilde{y}^{PLS}_{T+2|R} \\
&= \boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\tilde{\boldsymbol{A}}^{PLS}_{1:R+1}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R+1} - \boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\tilde{\boldsymbol{A}}^{PLS}_{1:R}\tilde{\boldsymbol{\delta}}^{PLS}_{1:R} \\
&= \frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}(\tilde{\boldsymbol{A}}^{PLS}_{1:R}, \tilde{\boldsymbol{\alpha}}^{PLS}_{R+1}) \begin{pmatrix} N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} \\ 0 \end{pmatrix} + \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R+1}\begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ -1 \end{pmatrix} \\
&\quad - \frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R} N^{1/2}\tilde{\boldsymbol{\delta}}_{1:R} \\
&= \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R+1}\begin{pmatrix} (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix} \\
&\quad + \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R}(\tilde{\boldsymbol{\theta}} - (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1)) \\
&= \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}(\boldsymbol{V}_{H1}(R+1) + \boldsymbol{V}_{H2}(R+1) + \boldsymbol{V}_L(R+1)) \\
&\qquad \times \begin{pmatrix} (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix} \\
&\quad + \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\tilde{\boldsymbol{A}}^{PLS}_{1:R}(\tilde{\boldsymbol{\theta}} - ([\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1)).
\end{aligned}$$

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 $\left\|\tilde{y}^{PLS}_{T+2|R+1} - \tilde{y}^{PLS}_{T+2|R}\right\|_2 = O_p(m^{-\gamma})$, which implies (i) because

$$\left\|\tilde{y}^{PLS}_{T+2|R+1} - y^*_{T+2}\right\|_2 \leq \left\|\tilde{y}^{PLS}_{T+2|R+1} - \tilde{y}^{PLS}_{T+2|R}\right\|_2 + \left\|\tilde{y}^{PLS}_{T+2|R} - y^*_{T+2}\right\|_2 = O_p(m^{-\gamma}).$$

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 can show

$$\begin{aligned}
&\left\|\tilde{y}^{PLS}_{T+2|R+1} - \tilde{y}^{PLS}_{T+2|R} - \mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{V}_L(R+1)\begin{pmatrix} (\boldsymbol{D}_0(R))^{-1}\boldsymbol{d}_0(R+1) \\ -1 \end{pmatrix}\right\|_2 \\
&= O_p((T^{-1/2} + m^{-3/2})^2). \qquad \text{(C.24)}
\end{aligned}$$

Notice that by Lemma C.18,

$$\left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{v}_L(1)\right\|_2 = \left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}}\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}\right\|_2 = O_p((mT)^{-1/2} + m^{-3/2}) \tag{C.25}$$

because

$$\begin{aligned}\left\|N^{-1/2}\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}}\right\|_2 &= (N^{-1}\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}\prime}\boldsymbol{x}_{\boldsymbol{\cdot}T+1})^{1/2} \\ &\leq (N^{-1}\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}\boldsymbol{x}_{\boldsymbol{\cdot}T+1})^{1/2} = O_p(1).\end{aligned}$$

For $q \geq 2$,

$$\begin{aligned}\left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{v}_L(q)\right\|_2 &= \left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}}(\Lambda_L^{\boldsymbol{S}_{NT}})^{q-1}\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}\right\|_2 \\ &\leq \left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}}\right\|_2 \left\|(\Lambda_L^{\boldsymbol{S}_{NT}})^{q-1}\right\|_2 \left\|\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}\right\|_2 \\ &\leq O_p(1)O_p(m^{-(q-1)})O_p((T^{-1/2} + m^{-3/2})) \leq O_p(m^{-3/2}).\end{aligned}$$

Thus, we have

$$\begin{aligned}\left\|\mathcal{Y}_{NT}\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{V}_L(R+1)\right\|_2 &\leq (R+1) \times \left\|\frac{\boldsymbol{x}'_{\boldsymbol{\cdot}T+1}}{N^{1/2}}\boldsymbol{\Xi}_L^{\boldsymbol{S}_{NT}}\boldsymbol{c}_L^{\boldsymbol{S}_{NT}}\right\|_2 |\mathcal{Y}_{NT}| \\ &\leq O_p(m^{1/2}T^{-1/2} + m^{-1/2})\end{aligned}$$

By (C.24) and (C.26), we can have

$$\left\|\tilde{y}^{PLS}_{T+2|R+1} - \tilde{y}^{PLS}_{T+2|R}\right\|_2 = O_p((m/T)^{1/2} + m^{-1/2})$$

which implies parts (ii) and (iii). *Q.E.D.*

Tables and Figures

Number of PLS factors used (q)	Mean in-sample R^2	Standard Dev. of in-sample R^2	Out-of-sample R^2
1	62.55	4.58	50.64
2	72.88	3.74	53.03
3	78.68	3.24	49.32
4	82.02	3.00	46.36
5	84.35	2.78	41.10
6	86.08	2.61	36.91
7	87.40	2.50	33.10
8	88.48	2.40	28.87
9	89.39	2.33	24.45
10	90.17	2.25	18.77

Table 1: In-Sample and Out-of-Sample Percentage R^2 of PLS Regressions ($R = 2$, $K = 4$)

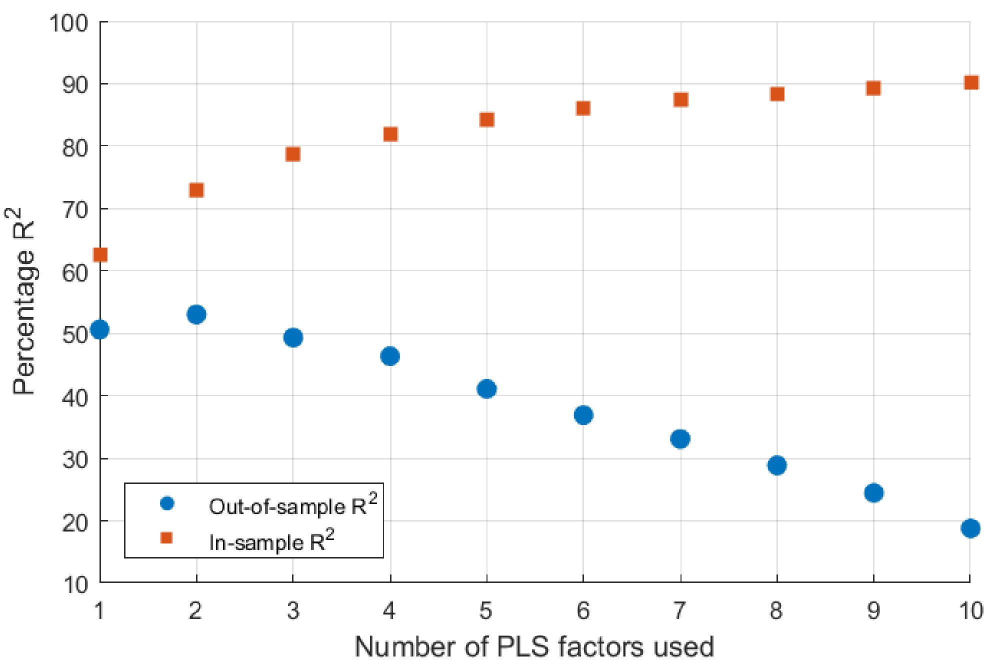

Figure 1: Graphical Representation of 1

Notes: Spurious correlation ooccurs as more PLS factors are used. This grapf is for the case in which there are four common latent factors in predictors and two informative PLS factors. Thus, forecasting power is maximized when the first 2 PLS factors are used. The R^2_{OS} of PLS factors significantly decreases as more than two PLS factors are used while the in-sample adjusted R^2 always increases.

$T=100$, $N=80$								
a_x	a_y	ρ_c	ρ_e	ρ_f	PLS1	PC5	OLS	$\hat{K}$
0.1	0.3	0.0	0.0	0.0	0.137	**0.162**	-2.851	2.015
0.1	0.3	0.3	0.3	0.3	**0.112**	0.08	-3.043	2.629
0.1	0.3	0.5	0.5	0.5	**0.112**	0.072	-2.871	3.191
0.1	0.5	0.0	0.0	0.0	0.241	**0.245**	-2.268	1.999
0.1	0.5	0.3	0.3	0.3	**0.292**	0.169	-1.671	2.679
0.1	0.5	0.5	0.5	0.5	**0.287**	0.151	-2.034	3.139
0.1	0.7	0.0	0.0	0.0	**0.350**	0.321	-1.790	2.001
0.1	0.7	0.3	0.3	0.3	**0.394**	0.250	-1.230	2.687
0.1	0.7	0.5	0.5	0.5	**0.399**	0.205	-0.815	3.012
0.3	0.3	0.0	0.0	0.0	0.230	**0.239**	-2.706	4.748
0.3	0.3	0.3	0.3	0.3	0.205	**0.218**	-3.044	3.690
0.3	0.3	0.5	0.5	0.5	**0.254**	0.250	-3.158	2.921
0.3	0.5	0.0	0.0	0.0	0.404	**0.405**	-1.559	4.737
0.3	0.5	0.3	0.3	0.3	0.364	**0.368**	-1.892	3.665
0.3	0.5	0.5	0.5	0.5	**0.396**	0.389	-1.914	2.934
0.3	0.7	0.0	0.0	0.0	0.599	**0.611**	-0.889	4.743
0.3	0.7	0.3	0.3	0.3	**0.595**	0.589	-0.692	3.678
0.3	0.7	0.5	0.5	0.5	**0.586**	0.578	-0.551	2.895
0.5	0.3	0.0	0.0	0.0	**0.282**	0.278	-2.553	5.000
0.5	0.3	0.3	0.3	0.3	0.266	**0.275**	-2.174	4.987
0.5	0.3	0.5	0.5	0.5	**0.245**	0.243	-3.096	4.781
0.5	0.5	0.0	0.0	0.0	0.454	**0.466**	-1.531	5.000
0.5	0.5	0.3	0.3	0.3	0.446	**0.447**	-1.646	4.973
0.5	0.5	0.5	0.5	0.5	0.489	**0.500**	-1.459	4.825
0.5	0.7	0.0	0.0	0.0	0.622	**0.639**	-0.575	4.999
0.5	0.7	0.3	0.3	0.3	0.627	**0.640**	-0.602	4.963
0.5	0.7	0.5	0.5	0.5	0.628	**0.645**	-0.864	4.806

Table 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 with a combination of a_x, a_y, ρ_c, ρ_e,and ρ_f, the highest R^2_{OS} 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 $\hat{K}$ denotes the estimated number of the latent common factors in predictor variables by the Eigenvalue Ratio (ER) method of Ahn and Horenstein (2013).

$T = 200$, $N = 160$								
a_x	a_y	ρ_c	ρ_e	ρ_f	PLS1	PC5	OLS	$\hat{K}$
0.1	0.3	0.0	0.0	0.0	0.175	**0.207**	-2.989	4.168
0.1	0.3	0.3	0.3	0.3	**0.197**	0.181	-2.664	2.150
0.1	0.3	0.5	0.5	0.5	**0.185**	0.132	-2.608	2.203
0.1	0.5	0.0	0.0	0.0	0.354	**0.364**	-1.847	4.228
0.1	0.5	0.3	0.3	0.3	**0.369**	0.306	-1.676	2.181
0.1	0.5	0.5	0.5	0.5	**0.287**	0.206	-2.252	2.211
0.1	0.7	0.0	0.0	0.0	**0.524**	0.514	-1.212	4.291
0.1	0.7	0.3	0.3	0.3	**0.496**	0.406	-1.170	2.160
0.1	0.7	0.5	0.5	0.5	**0.485**	0.332	-0.752	2.192
0.3	0.3	0.0	0.0	0.0	0.262	**0.264**	-2.221	5.000
0.3	0.3	0.3	0.3	0.3	0.282	**0.283**	-2.490	5.000
0.3	0.3	0.5	0.5	0.5	0.272	**0.279**	-2.941	4.988
0.3	0.5	0.0	0.0	0.0	**0.473**	0.472	-1.495	5.000
0.3	0.5	0.3	0.3	0.3	0.433	**0.456**	-1.680	5.000
0.3	0.5	0.5	0.5	0.5	0.441	**0.458**	-1.542	4.982
0.3	0.0	0.0	0.0	0.0	0.636	**0.656**	-0.817	5.000
0.3	0.7	0.3	0.3	0.3	0.626	**0.647**	-0.617	5.000
0.3	0.7	0.5	0.5	0.5	0.605	**0.620**	-0.700	4.959
0.5	0.3	0.0	0.0	0.0	0.270	**0.278**	-2.359	5.000
0.5	0.3	0.3	0.3	0.3	**0.285**	0.282	-2.689	5.000
0.5	0.3	0.5	0.5	0.5	0.268	**0.279**	-2.978	5.000
0.5	0.5	0.0	0.0	0.0	0.429	**0.439**	-1.447	5.000
0.5	0.5	0.3	0.3	0.3	0.479	**0.496**	-1.525	5.000
0.5	0.5	0.5	0.5	0.5	0.501	**0.516**	-1.756	5.000
0.5	0.7	0.0	0.0	0.0	0.641	**0.669**	-0.738	5.000
0.5	0.7	0.3	0.3	0.3	0.644	**0.665**	-0.632	5.000
0.5	0.7	0.5	0.5	0.5	0.650	**0.666**	-0.788	5.000

Table 3: 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 $N = 160$ and $T = 200$. For each case with a combination of a_x, a_y, ρ_c, ρ_e,and ρ_f, the highest R^2_{OS} 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 $\hat{K}$ denotes the estimated number of the latent common factors in predictor variables by the Eigenvalue Ratio (ER) method of Ahn and Horenstein (2013).

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	0.0	**0.462**	0.397	0.290	**0.612**	0.549	0.500
0.1	0.0	0.0	0.3	**0.445**	0.374	0.243	**0.600**	0.511	0.457
0.1	0.0	0.0	0.5	**0.484**	0.421	0.300	**0.593**	0.517	0.455
0.1	0.3	0.3	00	**0.431**	0.421	0.359	**0.612**	0.569	0.519
0.1	0.3	0.3	0.3	**0.458**	0.450	0.376	**0.598**	0.561	0.520
0.1	0.3	0.3	0.5	**0.424**	0.395	0.333	**0.603**	0.560	0.510
0.1	0.5	0.5	0.0	0.426	**0.455**	0.446	**0.604**	0.590	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	0.0	0.567	**0.569**	0.448	**0.643**	0.624	0.556
0.2	0.0	0.0	0.3	0.573	**0.574**	0.485	**0.664**	0.639	0.577
0.2	0.0	0.0	0.5	**0.562**	0.561	0.471	**0.646**	0.612	0.568
0.2	0.3	0.3	0.0	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**	0.500	**0.664**	0.644	0.599
0.2	0.5	0.5	0.0	**0.554**	0.550	0.506	**0.632**	0.595	0.573
0.2	0.5	0.5	0.3	0.550	**0.551**	0.513	**0.640**	0.613	0.582
0.2	0.5	0.5	0.5	0.559	**0.580**	0.530	**0.634**	0.623	0.582
0.3	0.0	0.0	0.0	**0.616**	0.612	0.525	**0.665**	0.638	0.584
0.3	0.0	0.0	0.3	0.632	**0.640**	0.560	**0.696**	0.668	0.609
0.3	0.0	0.0	0.5	0.629	**0.636**	0.540	**0.677**	0.671	0.611
0.3	0.3	0.3	0.0	0.598	**0.613**	0.545	**0.654**	0.627	0.585
0.3	0.3	0.3	0.3	0.574	**0.605**	0.540	**0.670**	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	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	0.610	**0.626**	0.600	**0.662**	0.643	0.594

Table 4: Forecasting by Regressions with Different Numbers of PLS Factors ($R = K = 3$)
Notes: This table reports the forecasting performances of the regressions with three different numbers of informative PLS factors: one (PLS1), two (PLS2), and three (PLS3). For each data specification, the highest R^2_{OS} square is marked by 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	0.0	0.651	**0.669**	0.614	0.703	**0.719**	0.692
0.1	0.0	0.0	0.3	0.667	**0.684**	0.648	0.657	**0.672**	0.624
0.1	0.0	0.0	0.5	0.670	**0.700**	0.664	0.696	**0.704**	0.666
0.1	0.3	0.3	00	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	0.635	**0.652**	0.618	0.680	**0.695**	0.670
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	0.0	0.674	**0.694**	0.678	0.667	**0.680**	0.665
0.2	0.0	0.0	0.3	0.701	**0.719**	0.700	0.683	**0.691**	0.676
0.2	0.0	0.0	0.5	0.655	**0.685**	0.666	0.657	**0.666**	0.649
0.2	0.3	0.3	0.0	0.700	**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	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	**0.730**	0.724
0.3	0.0	0.0	0.0	0.626	**0.656**	0.648	0.686	**0.699**	0.696
0.3	0.0	0.0	0.3	0.651	**0.687**	0.677	0.701	**0.715**	0.714
0.3	0.0	0.0	0.5	0.650	**0.683**	0.678	0.681	**0.690**	0.689
0.3	0.3	0.3	0.0	0.664	**0.700**	0.686	0.695	**0.712**	0.710
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	0.670	**0.695**	0.687	0.689	0.710	**0.710**
0.3	0.5	0.5	0.3	0.653	**0.678**	0.666	0.690	**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 5: Forecasting by Regressions with Different Numbers of PLS Factors ($R = K = 3$)
Notes: This table reports the forecasting performances of the regressions with three different numbers of informative PLS factors: one (PLS1), two (PLS2), and three (PLS3). For each data specification, the highest R^2_{OS} square is marked by 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				$T = N = 7000$			$T = N = 10,000$		
a_x	ρ_d	ρ_e	ρ_f	PLS1	PLS2	PLS3	0.731	0.734	**0.736**
0.1	0.0	0.0	0.0	0.612	**0.617**	0.613	0.731	0.734	**0.736**
0.1	0.0	0.0	0.3	0.580	**0.590**	0.586	0.731	0.735	**0.738**
0.1	0.0	0.0	0.5	0.579	**0.591**	0.589	0.739	0.744	**0.747**
0.1	0.3	0.3	00	0.610	**0.616**	0.613	0.729	0.733	**0.734**
0.1	0.3	0.3	0.3	0.579	**0.589**	0.586	0.729	0.734	**0.736**
0.1	0.3	0.3	0.5	0.578	**0.591**	0.59	0.738	0.742	**0.745**
0.1	0.5	0.5	0.0	0.611	**0.617**	0.615	0.729	0.733	**0.734**
0.1	0.5	0.5	0.3	0.579	**0.591**	0.588	0.729	0.734	**0.736**
0.1	0.5	0.5	0.5	0.578	**0.593**	0.592	0.737	0.743	**0.745**
0.2	0.0	0.0	0.0	0.611	**0.615**	0.614	0.732	0.736	**0.738**
0.2	0.0	0.0	0.3	0.579	**0.588**	0.587	0.732	0.737	**0.740**
0.2	0.0	0.0	0.5	0.578	0.590	**0.590**	0.741	0.746	**0.748**
0.2	0.3	0.3	0.0	0.610	**0.615**	0.614	0.731	0.736	**0.739**
0.2	0.3	0.3	0.3	0.578	**0.588**	0.587	0.74	0.745	**0.747**
0.2	0.3	0.3	0.5	0.577	**0.590**	0.590	0.731	0.735	**0.737**
0.2	0.5	0.5	0.0	0.610	**0.616**	0.616	0.731	0.736	**0.739**
0.2	0.5	0.5	0.3	0.578	**0.589**	0.589	0.739	0.745	**0.747**
0.2	0.5	0.5	0.5	0.577	0.591	**0.592**	0.739	0.745	**0.747**
0.3	0.0	0.0	0.0	0.610	**0.615**	0.614	0.733	0.737	**0.738**
0.3	0.0	0.0	0.3	0.578	**0.587**	0.587	0.733	0.738	**0.740**
0.3	0.0	0.0	0.5	0.577	0.589	**0.59**	0.741	0.746	**0.749**
0.3	0.3	0.3	0.0	0.609	**0.615**	0.614	0.732	0.736	**0.738**
0.3	0.3	0.3	0.3	0.577	**0.587**	0.587	0.732	0.737	**0.739**
0.3	0.3	0.3	0.5	0.576	0.589	**0.590**	0.740	0.746	**0.748**
0.3	0.5	0.5	0.0	0.610	**0.615**	0.615	0.732	0.736	**0.738**
0.3	0.5	0.5	0.3	0.577	**0.588**	0.588	0.732	0.737	**0.739**
0.3	0.5	0.5	0.5	0.576	0.59	**0.591**	0.740	0.746	**0.748**

Table 6: Forecasting by Regressions with Different Numbers of PLS Factors ($R = K = 3$)
Notes: This table reports the forecasting performances of the regressions with three different numbers of informative PLS factors: one (PLS1), two (PLS2), and three (PLS3). For each data specification, the highest R^2_{OS} square is marked by bold. The other parameters used to generate the data are set at $a_y = 0.7$ and $\boldsymbol{\Omega}^*_{\boldsymbol{F}} = \boldsymbol{diag}(3, 5, 7)$. For the cases with $N = T = 10,000$, only 100 different samples are generated for each data generating processes.

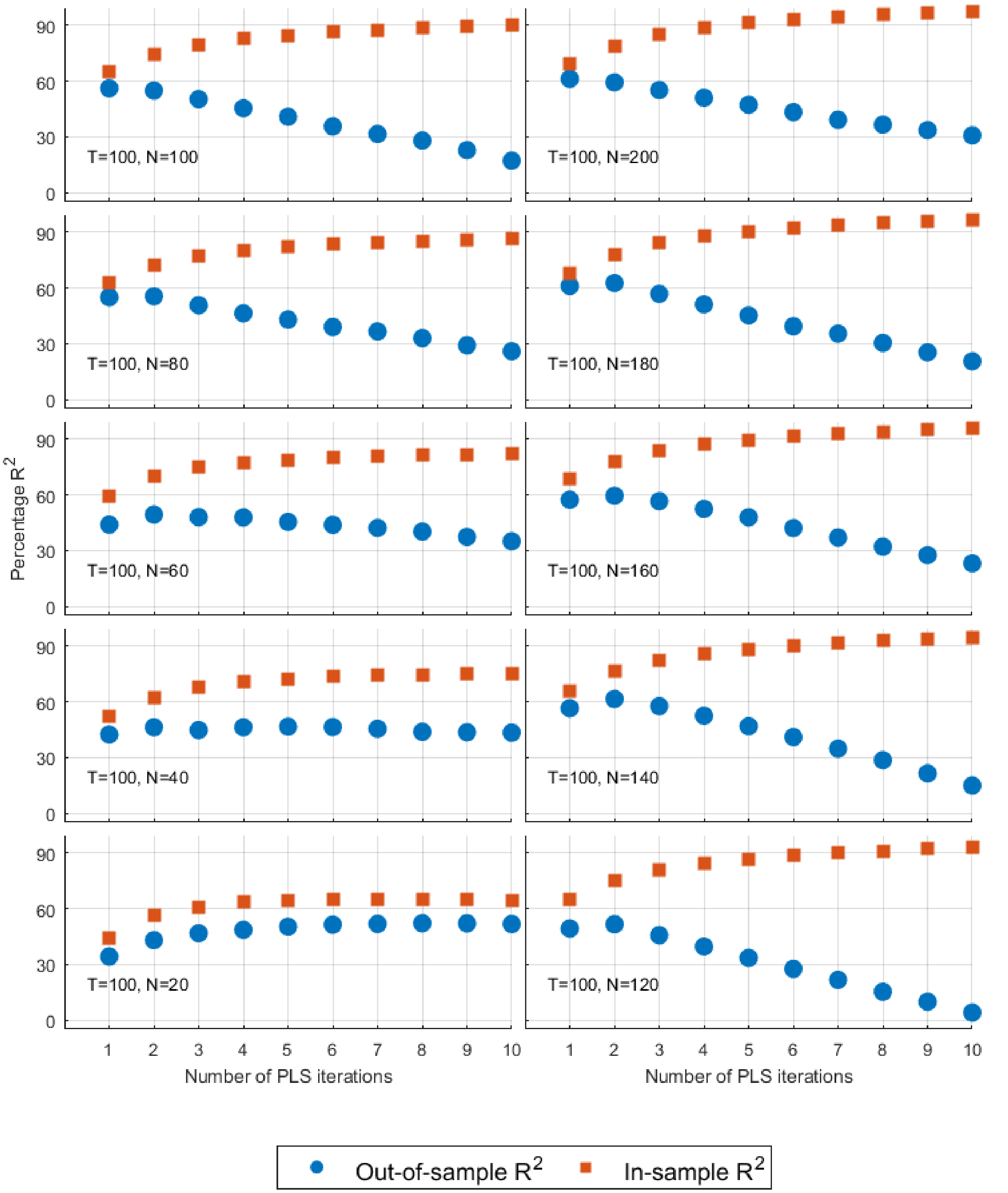

Figure 2: Performances of PLS Regression and Spurious Correlation ($T = 100$)

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 $\rho_f = \rho_e = \rho_c = 0.5$.

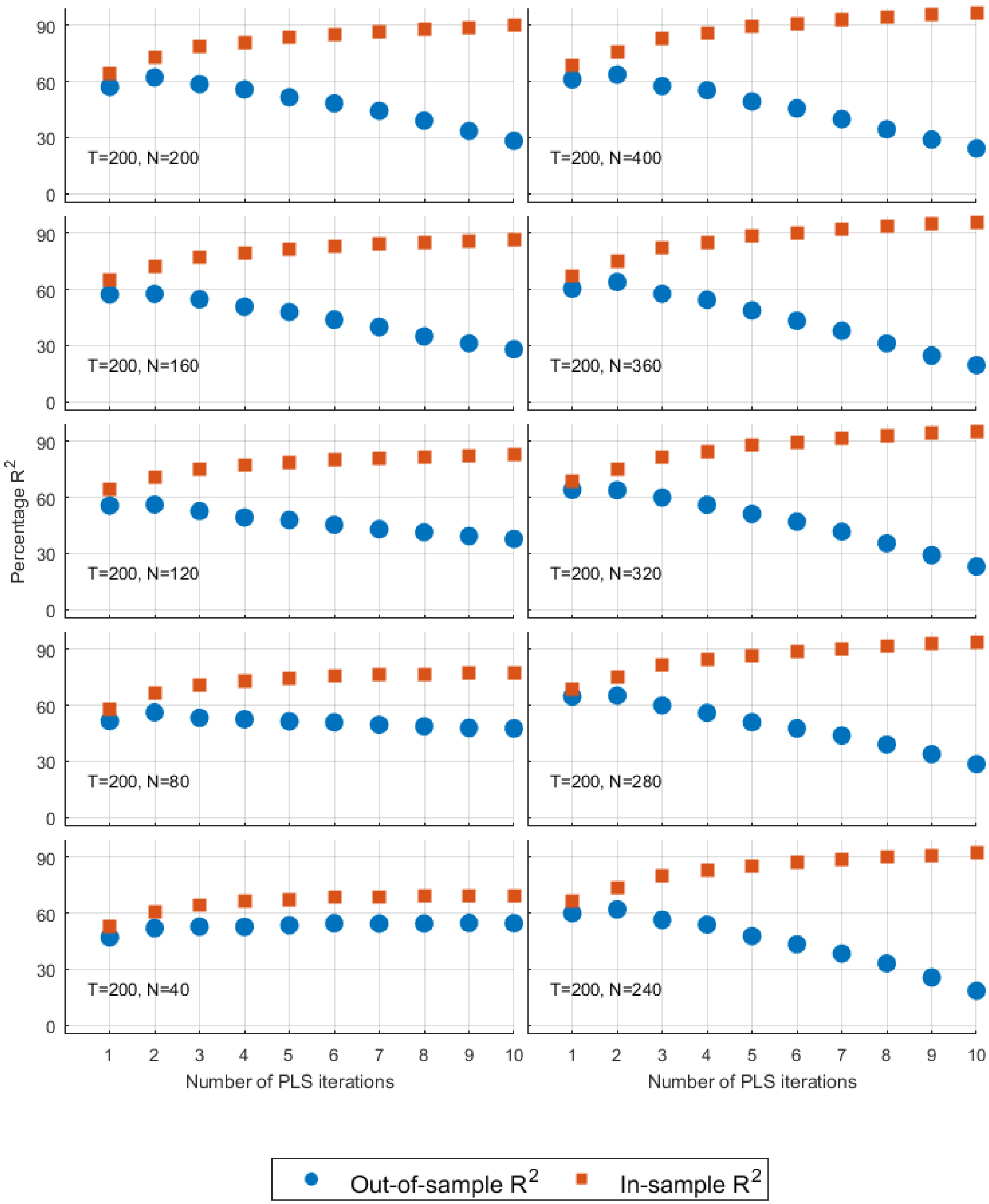

Figure 3: Performances of PLS Regression and Spurious Correlation ($T = 200$)

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 $\rho_f = \rho_e = \rho_c = 0.5$.

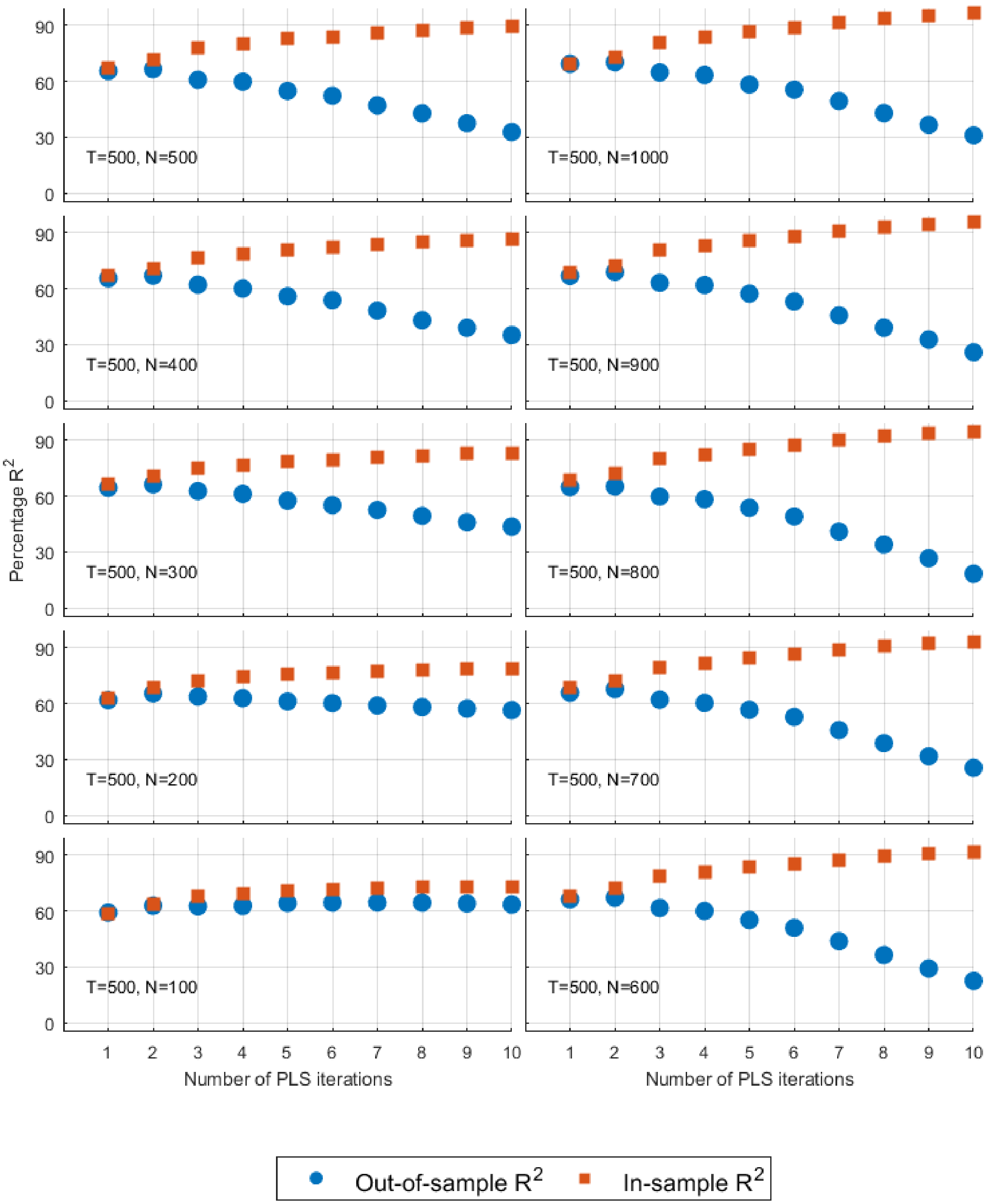

Figure 4: Performances of PLS Regression and Spurious Correlation ($T = 500$)
Notes: The other parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 5, 5)$, $R = 2$, $K = 4$, $a_x = 0.2$, $a_y = 0.7$, and $\rho_f = \rho_e = \rho_c = 0.5$.

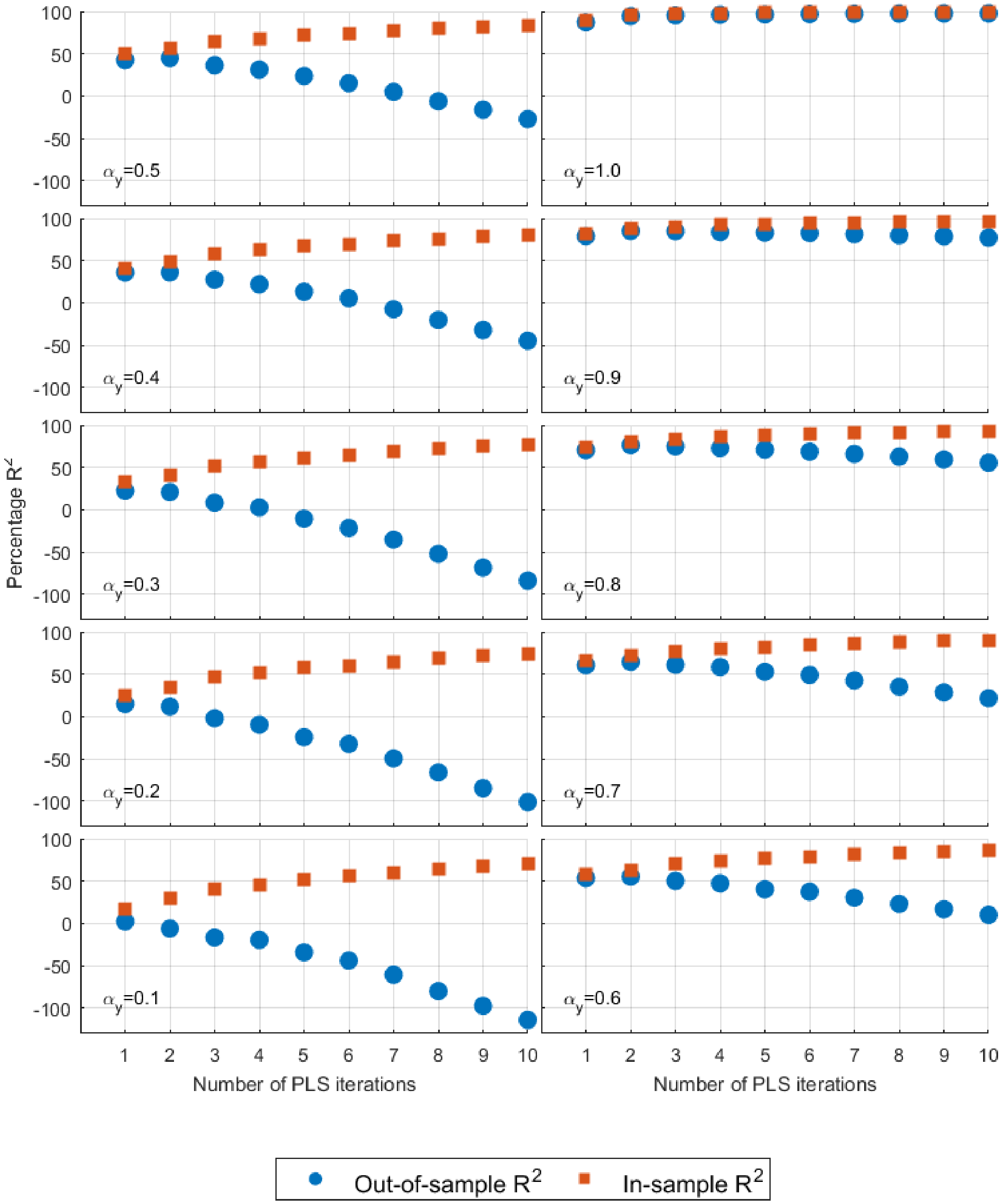

Figure 5: Performance of PLS Regression and Spurious Correlation ($a_x = 0.2$)

Notes: The data generating parameters other than a_y 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 $N = T = 100$.

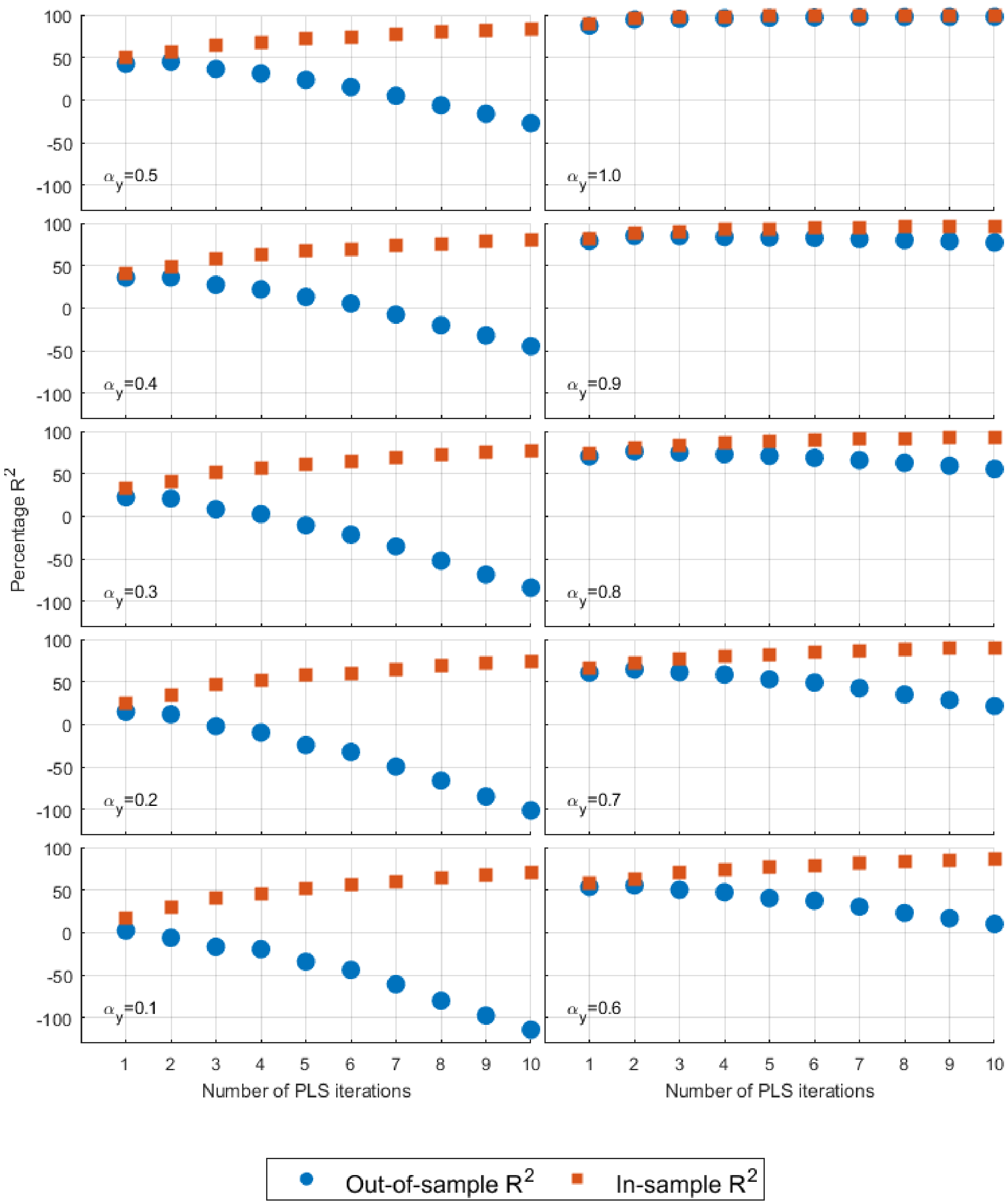

Figure 6: Performance of PLS Regression and Spurious Correlation ($a_x = 0.5$)

Notes: The data generating parameters other than a_y are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^* = \boldsymbol{diag}(3, 3, 5, 5)$, $R = 2$, $K = 4$, $a_x = 0.5$, $\rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

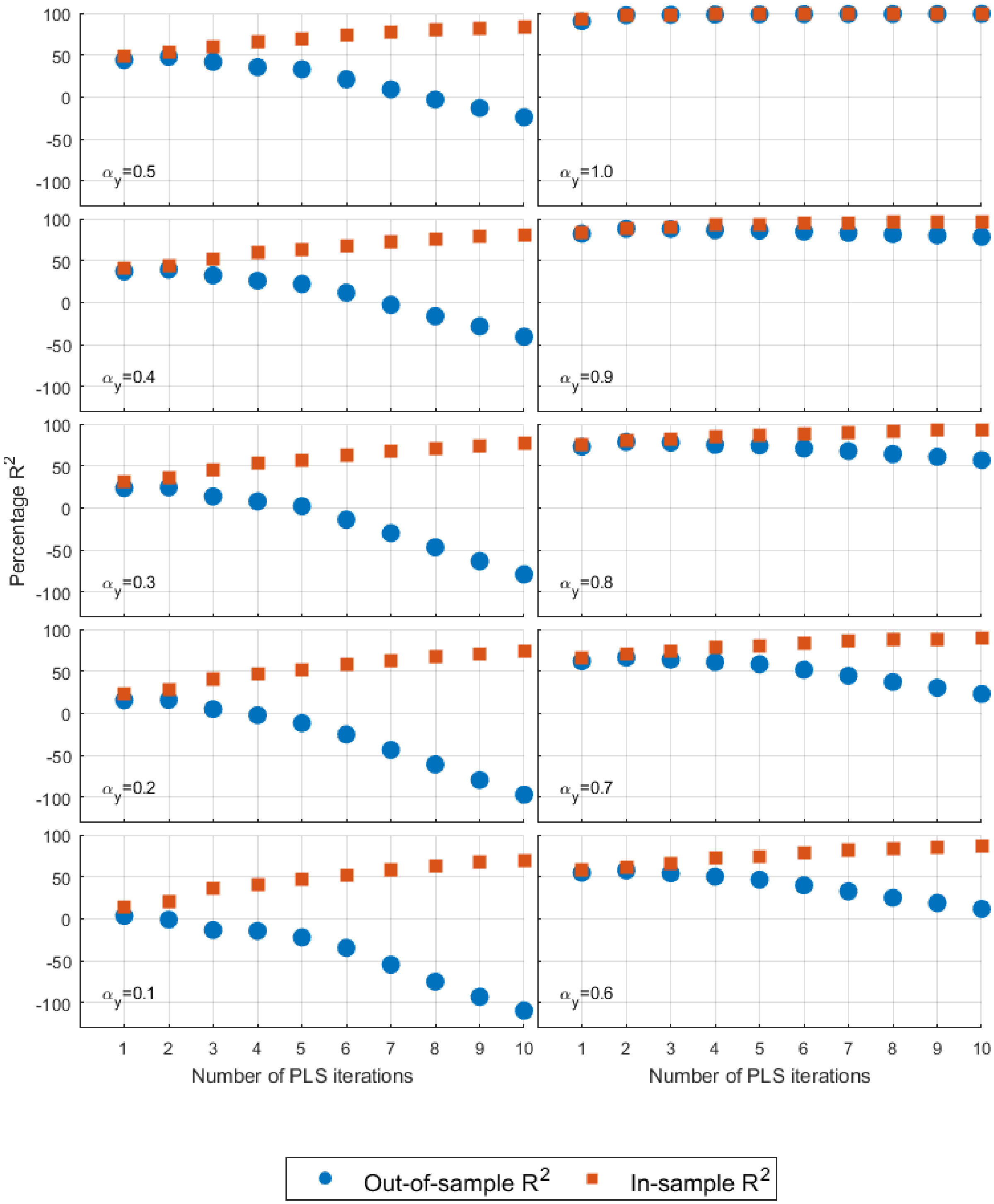

Figure 7: Performance of PLS Regression and Spurious Correlation ($a_x = 0.7$)

Notes: The data generating parameters other than a_y 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 $N = T = 100$.

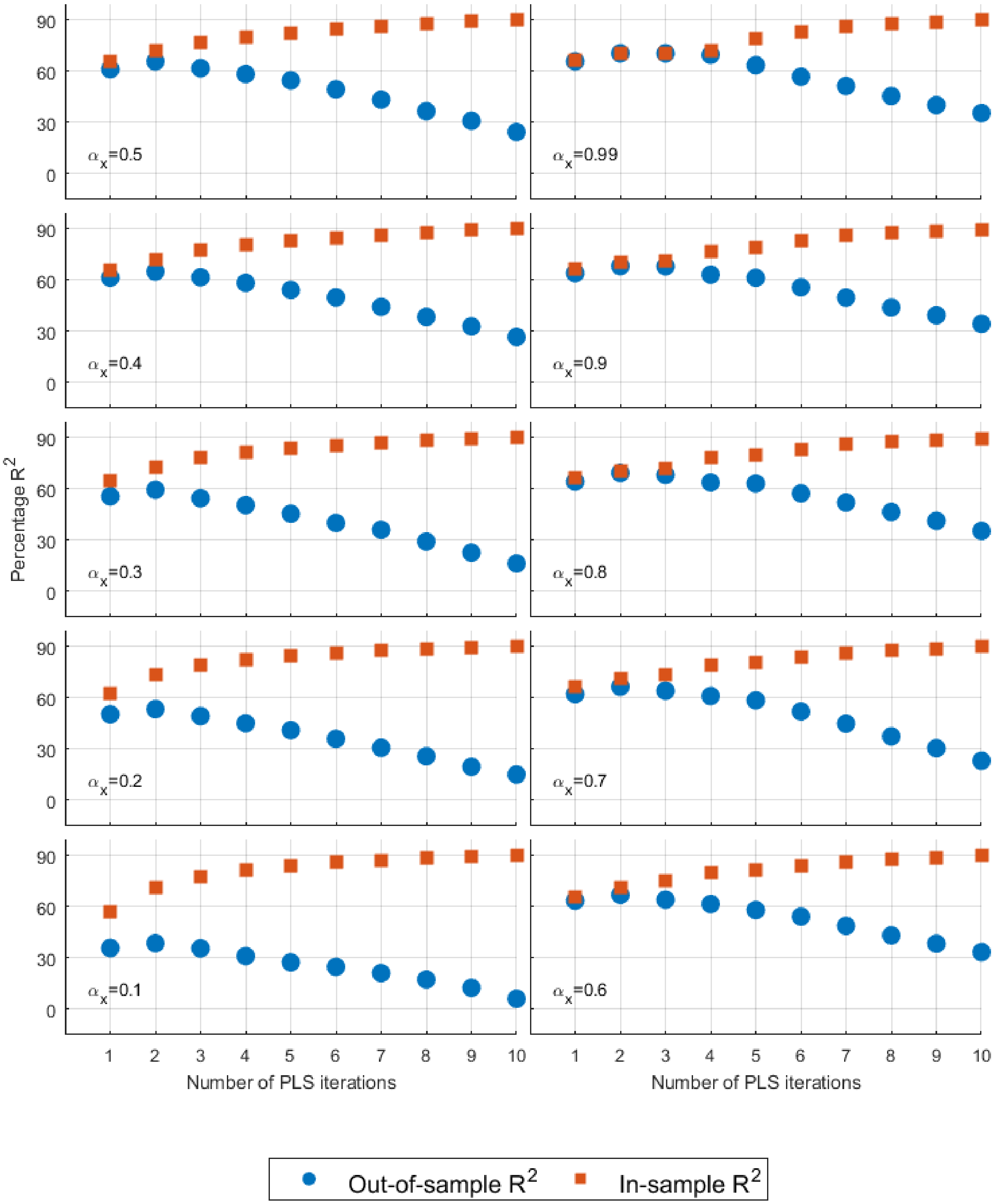

Figure 8: Performance of PLS Regression and Spurious Correlation ($a_y = 0.7$)

Notes: The data generating parameters other than a_x 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 $N = T = 100$.

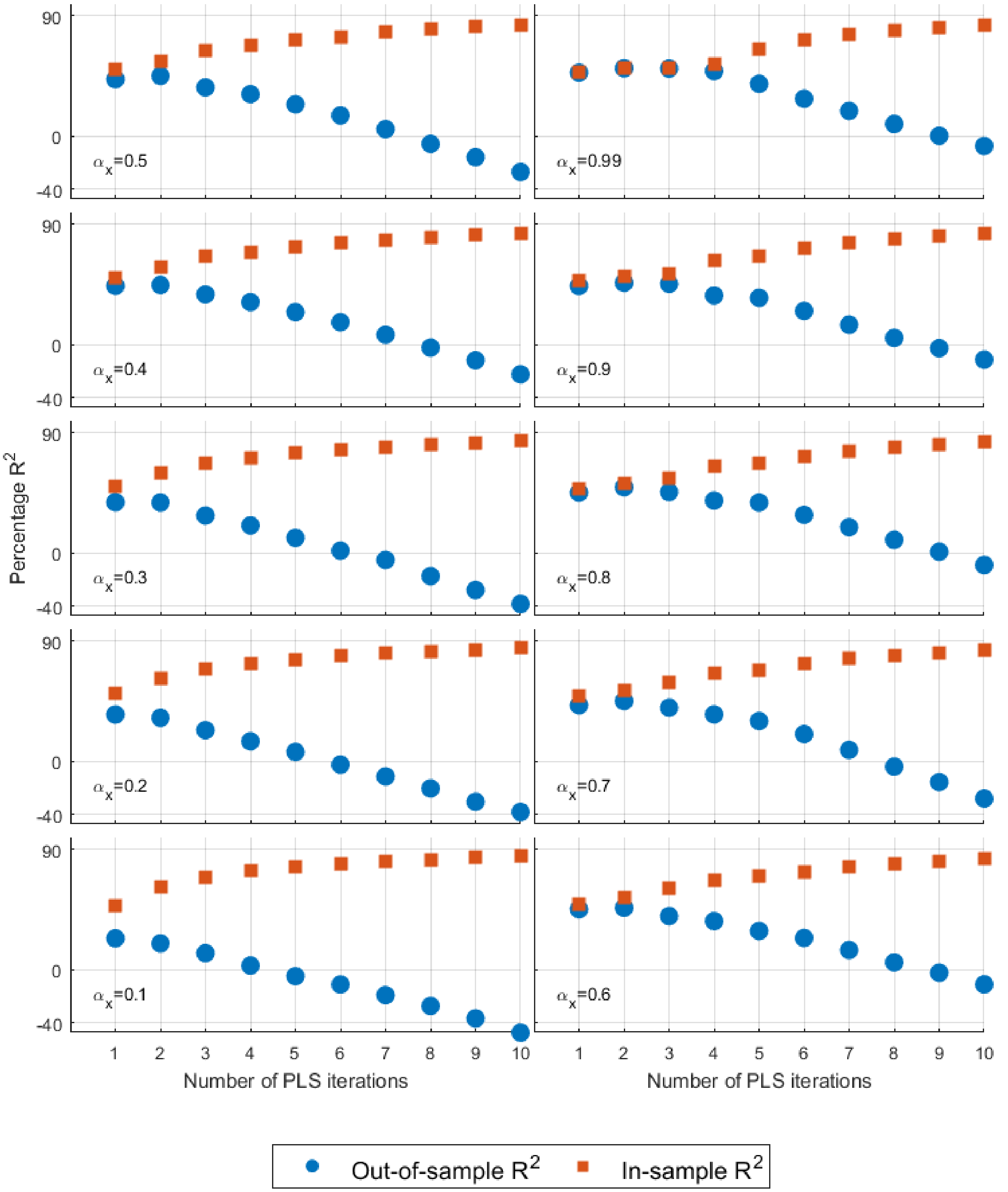

Figure 9: Performance of PLS Regression and Spurious Correlation ($a_y = 0.5$)

Notes: The data generating parameters other than a_x 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 $N = T = 100$.

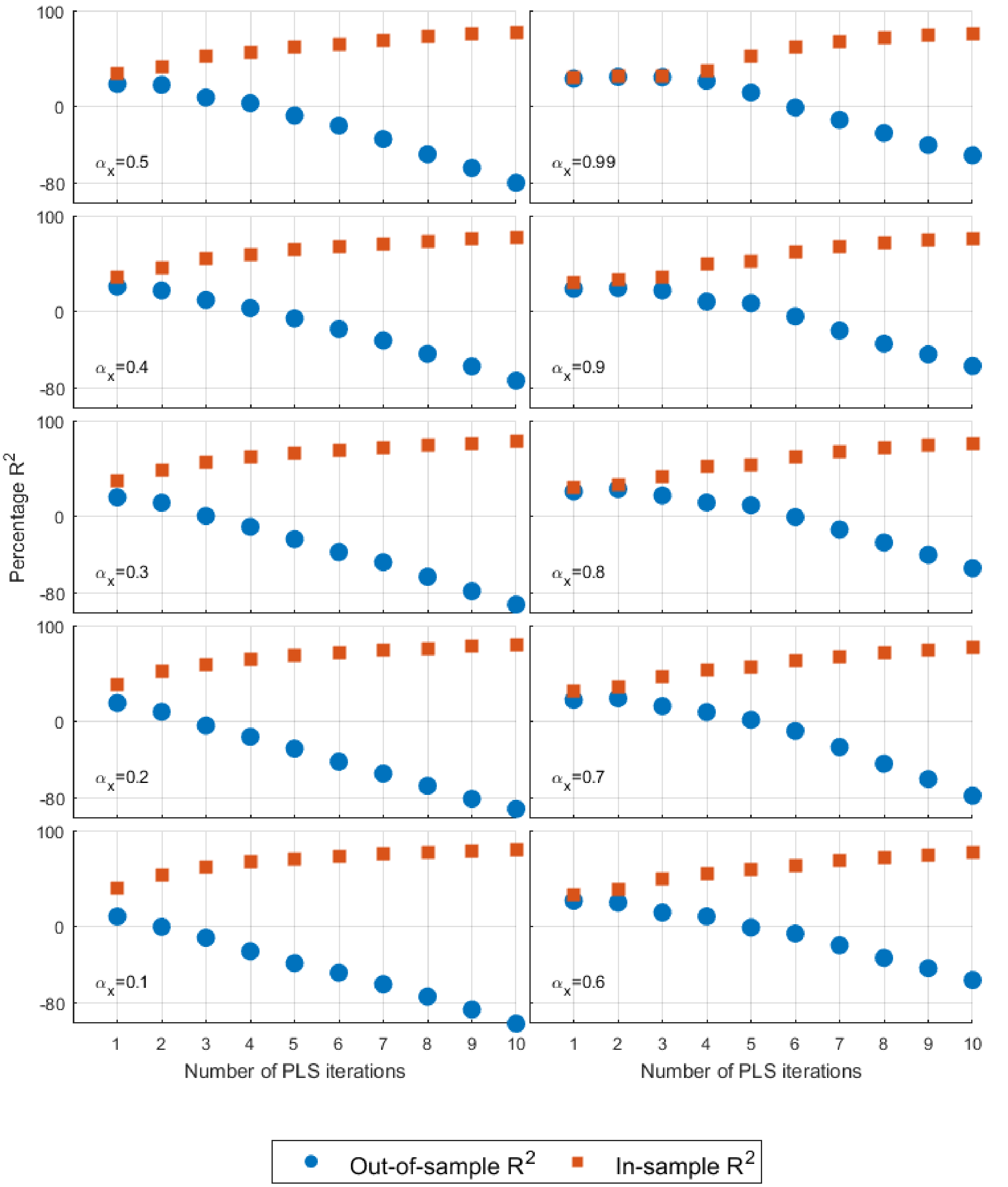

Figure 10: Performance of PLS Regression and Spurious Correlation ($a_y = 0.3$)

Notes: The data generating parameters other than a_x 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 $N = T = 100$.

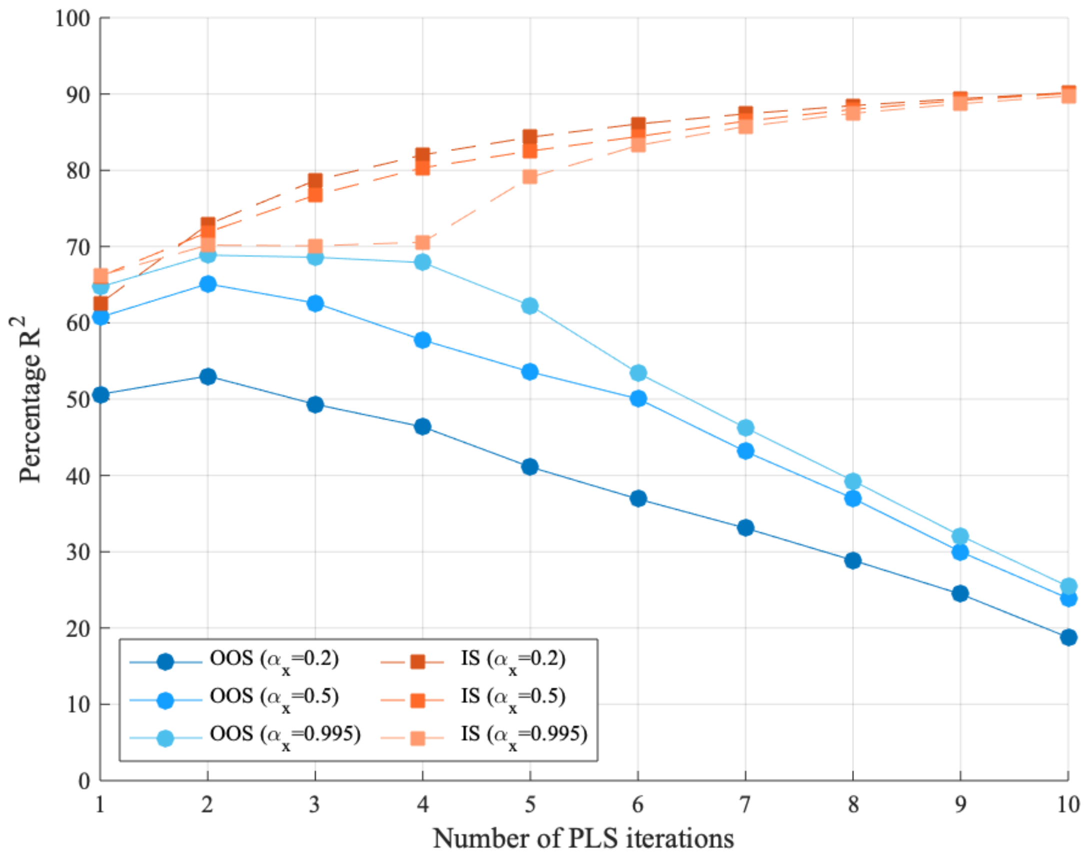

Figure 11: Forecasting with Uninformative and Spurious Factors ($N = T = 100$)

Notes: The parameters for data generating processes other than a_x 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 $N = T = 100$.

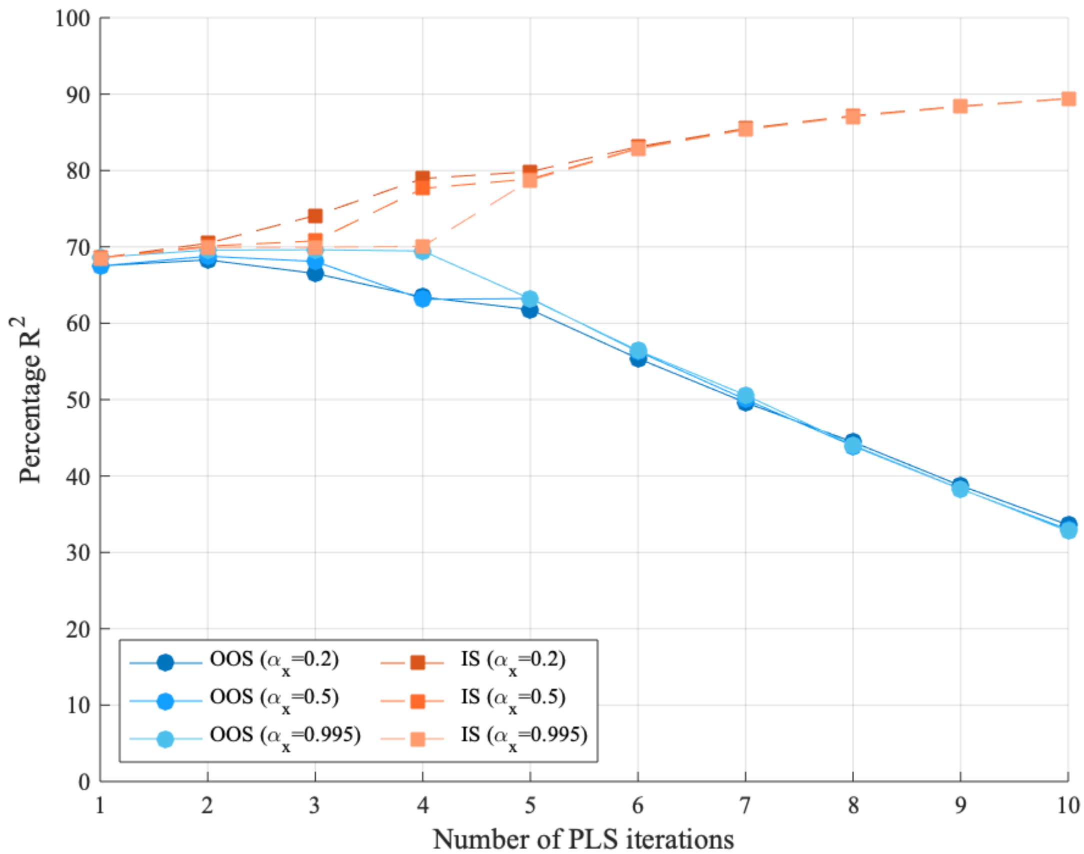

Figure 12: Forecasting with Uninformative and Spurious Factors ($N = T = 2,000$)

Notes: The parameters for data generating processes other than a_x 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 $N = T = 100$.

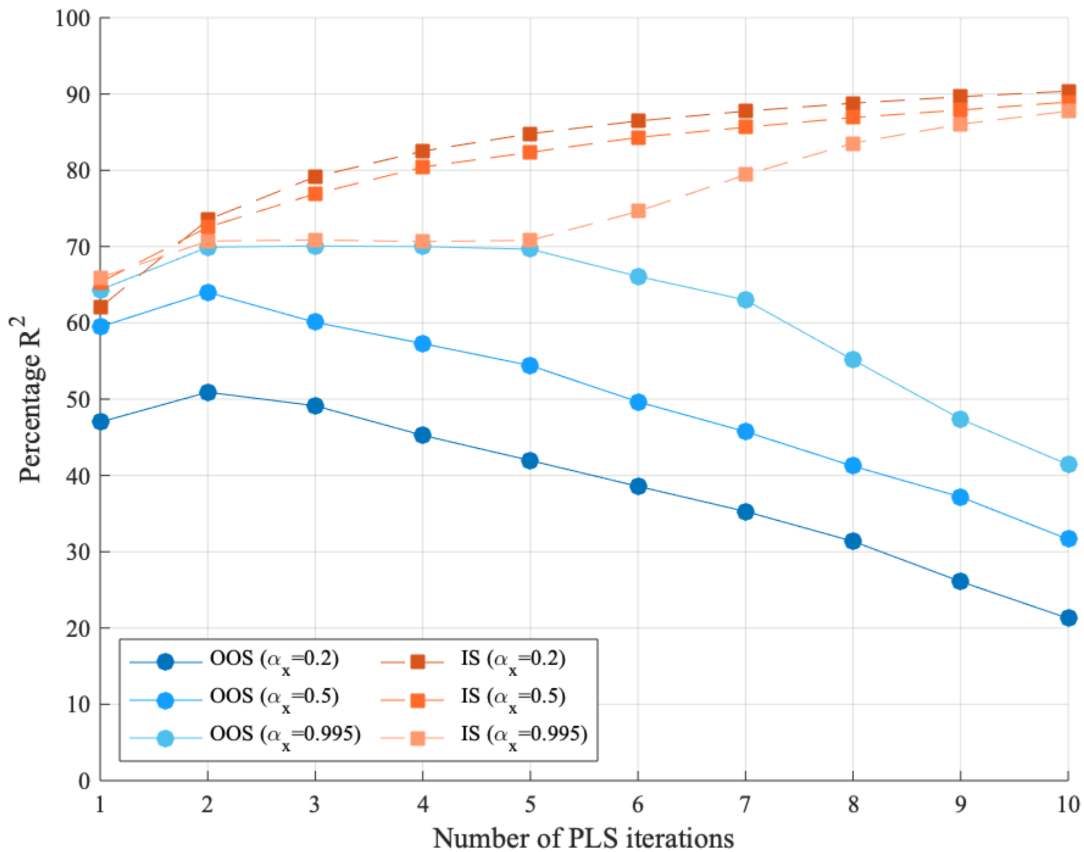

Figure 13: Forecasting with Uninformative and Spurious Factors $(K = 6, R = 2, N = T = 100)$
Notes: The parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 3, 3, 5, 5)$, $\boldsymbol{\beta}^{*} = (1, 0, 0, 1, 0, 0)'$, $R = 2, K = 6, a_y = 0.7$, $\rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

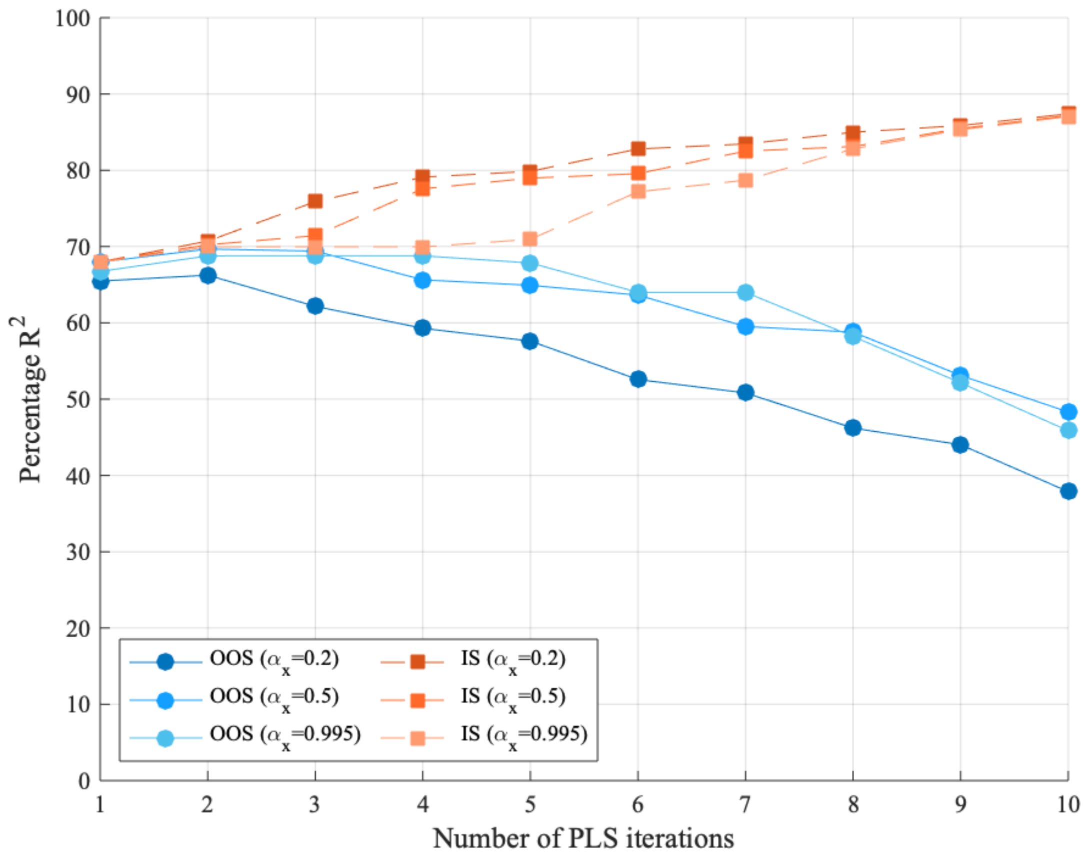

Figure 14: Forecasting with Uninformative and Spurious Factors $(K = 6, R = 2, N = T = 2,000)$

Notes: The parameters for data generating processes are set at $\boldsymbol{\Omega}_{\boldsymbol{F}}^{*} = \boldsymbol{diag}(3, 3, 3, 3, 5, 5)$, $\boldsymbol{\beta}^{*} = (1, 0, 0, 1, 0, 0)'$, $R = 2, K = 6, a_y = 0.7$, $\rho_f = \rho_e = \rho_c = 0.5$, and $N = T = 100$.

Sample Size			Cross-validation	Forecasting with PLS factors						CV-estimate of the optimal number of PLS factors ($\hat{R}_{CV}$)	
T	N	N/T	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	Mean	Std
100	20	0.0	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	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	0.62	0.62	**0.64**	0.59	0.52	0.47	0.42	1.87	1.01
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	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	0.65	0.62	**0.66**	0.62	0.59	0.55	0.51	1.85	0.56
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	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	0.68	0.65	**0.68**	0.65	0.63	0.58	0.54	1.84	0.41

Table 7: 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 size			Cross-validation	Forecasting with PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	a_x	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	Mean	Std.
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
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 8: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression Across Different a_x's

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 size			Cross-validation	Forecasting with PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	a_y	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	Mean	Std.
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
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 9: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression Across Different a_y's

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 size			Cross-validation	Forecasting with PLS factors						Statistics of $\hat{R}_{CV}$	
T	N	ρ_{eu}	CV	PLS1	PLS2	PLS3	PLS4	PLS5	PLS6	Mean	Std.
100	100	0.1	0.57	0.58	**0.59**	0.56	0.51	0.48	0.43	1.99	1.09
100	100	0.3	0.55	0.53	**0.56**	0.54	0.51	0.47	0.42	2.25	1.33
100	100	0.5	0.57	0.54	**0.57**	0.57	0.57	0.56	0.55	2.72	1.68
100	100	0.7	0.55	0.52	0.55	0.55	**0.56**	0.54	0.54	3.70	2.37
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	**0.88**	0.60	0.68	0.71	0.75	0.78	0.81	9.83	0.73
200	200	0.1	0.65	0.64	**0.66**	0.62	0.59	0.56	0.53	1.79	0.73
200	200	0.3	0.62	0.62	**0.64**	0.61	0.58	0.54	0.50	1.99	0.79
200	200	0.5	0.63	0.61	**0.64**	0.62	0.60	0.58	0.57	2.23	1.16
200	200	0.7	0.58	0.56	0.60	0.61	0.60	**0.61**	0.60	3.32	2.02
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	**0.85**	0.62	0.68	0.69	0.70	0.73	0.76	9.98	0.28

Table 10: Relative Forecasting Power of the Cross-Validation Augmented PLS Regression When Some Predictor Has Direct Forecasting Power

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 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. When $\rho_{eu} = 0$, the data generating process used for this table are identical to those which is used for Table 1.

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**	26.0	-4.6	-54.5	2.7	30.6	26.9	29.1	-559.1	29.6	26	23.7
Nonagg. 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 except 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 11: Forecasting Results for Eight Macroeconomic Variables

Notes: The $100 \times R^2_{OS}$'s from the regressions with eight economic variables are reported. The highest value of $100 \times R^2_{OS}$ obtained for each variable is marked by bold.

Categories of Variables	PLS1	PLS2	PLS3	PLS4	PC1	PC2	PC3	PC4	PLS BIC	PLS CV	PC BIC	PC AH
Overall	**35.0**	34.3	15.6	0.5	24.1	33.8	32.4	33.0	-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	28.0	41.5	39.9	38.7	-82.6	40.2	**43.5**	38.2
Housing	45.5	46.8	27.4	44.0	51.6	52.0	52.4	53.2	35.5	33.6	52.0	**54.6**
Consumption	**13.6**	2.8	-46.0	-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	27.0	42.0	44.3
Interest and Exchange Rates	11.0	-1.8	-16.6	-27.5	11.6	11.1	11.0	8.9	-122.8	5.3	-0.9	**12.1**
Prices	**60.0**	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	**8.0**	3.2	2.1	-0.8	-150.7	-2.6	1.0	1.5

Table 12: Forecasting Results for 144 Macroeconomic Variables

Notes: The whole 144 target variables are forecasted. The variables are categorized into eight groups. The median value of $100 \times R^2_{OS}$'s is reported for each category. The highest median value of $100 \times R^2_{OS}$ for each category is marked by bold. The category "Consumption" includes consumption, orders, and inventory variables.