Forecasting Using Cross-Section Average-Augmented Time Series Regressions

Hande Karabiyik
VU University

Joakim Westerlund*
Lund University
and
Centre for Financial Econometrics
Deakin University

February 19, 2019

Abstract

There is a large and growing literature concerned with forecasting time series variables using factor-augmented regression models. The workhorse of this literature is a two-step approach in which the factors are first estimated by applying the principal components method to a large panel of variables, and the forecast regression is estimated conditional on the first-step factor estimates. Another stream of research that has attracted much attention is that concerned with the use of cross-section averages as common factor estimates in interactive effects panel regression models. The main justification for this second development is the simplicity and good performance of the cross-section averages when compared to estimated principal component factors. In view of this, it is quite surprising that no one has yet considered the use of cross-section averages for forecasting. Indeed, given the purpose to forecast the conditional mean, the use of the cross-section average to estimate the factors is only natural. The present paper can be seen as a reaction to this. The purpose is to investigate the asymptotic and small-sample properties of forecasts based on cross-section average-augmented regressions. In contrast to existing studies, the investigation is carried out while allowing the number of factors to be known.

JEL Classification: C12; C13; C33.

Keywords: Forecasting; Factor-augmented regressions; Cross-section average.

*Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: `joakim.westerlund@nek.lu.se`.

1 Introduction

Consider the scalar variable y_t, observable for $t = 1, ..., T$ time periods. The data generating process (DGP) of this variable is the same as in Bai and Ng (2006), and is given by

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h} = \delta' z_t + \varepsilon_{t+h}, \tag{1}$$

where $z_t = [F_t', W_t']'$, $\delta = [\alpha', \beta']'$, $h \geq 0$ is the forecast horizon, F_t is a $r \times 1$ vector of unobserved common factors, or "diffusion indices", W_t is a $n \times 1$ vector of known variables, and ε_t is an error term. While we do not get to observe F_t, we assume that there is another variable available that is informative regarding F_t. Specifically, we assume the existence of an "external" $m \times 1$ panel data variable, $x_{i,t}$, which is observable not only across time but also across $i = 1, ..., N$ cross-sectional units. Similarly to Bai and Ng (2006), the DGP of this variable is assumed to have the following common factor representation:

$$x_{i,t} = \lambda_i' F_t + e_{i,t}, \tag{2}$$

where λ_i is a $r \times m$ matrix of factor loadings and $e_{i,t}$ is a $m \times 1$ vector of errors that are "largely idiosyncratic".

As is well known, if F_t and δ were known, and $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, ...) = 0$, the mean square optimal forecast of y_t is given by

$$y_{T+h|T} = E(y_{T+h}|z_T, z_{T-1}, ...) = \delta' z_T. \tag{3}$$

Of course, F_t and δ are not known, and we therefore use $\widehat{y}_{T+h|T} = \widehat{\delta}'\widehat{z}_T$ in place of $y_{T+h|T}$. Here $\widehat{z}_t = [\widehat{F}_t', W_t']'$, where $\widehat{F}_t$ is an estimator of the space spanned by F_t, and $\widehat{\delta} = [\widehat{\alpha}', \widehat{\beta}']'$ is the least squares (LS) slope estimator in a regression of y_{t+h} onto $\widehat{z}_t$. An important question here is: How to construct $\widehat{F}_t$? The previous literature has focused almost exclusively on the case when $\widehat{F}_t$ is obtained using the principal components (PC) method (see Bai and Ng, 2006; Corradi and Swanson, 2014; Djogbenou et al., 2015, 2017; Stock and Watson, 2002a, 2002b, to mention a few).[1] The idea here is to chose $\widehat{F}_t$ (and $\widehat{\lambda}_i$) so as to minimize the variance of the resulting idiosyncratic errors, a problem that can be solved by performing an eigenvalue decomposition of the sample covariance matrix of $x_{i,t}$. The estimated PC factors can then be

[1] Bai and Ng (2009) provide a review of some alternative computationally more demanding approaches, and give additional references.

seen as weighted cross-section averages of $x_{i,t}$ with weights set equal to the eigenvectors of the sample covariance matrix of $x_{i,t}$.

The results obtained by using the PC method have been very encouraging (see, for example, Eickmeier and Ziegler, 2008). One reason for this is the "averaging effect" that occurs when pooling from across the cross-sectional dimension of $x_{i,t}$, and that works by effectively smoothing any structural instabilities that might exist (see, for example, De Mol et al., 2008; Banerjee et al., 2008; Stock and Watson, 2009). A drawback is that PC is by construction sensitive to both the level and variation of the variance of $e_{i,t}$ (see, for example, Breitung and Tenhofen, 2011; Choi, 2012; Boivin and Ng, 2006). Hence, while the averaging makes it robust to certain features, PC is sensitive to the weights. This begs the question: Why not consider the equal weighted cross-section average (CA) of $x_{i,t}$ as an estimator of the space spanned by F_t? The fact that this question has not yet received an answer is particularly surprising given the good performance of the equal weighted average in combining forecasts from different models. In fact, the performance of the equal weighted forecast combination has been so good that it has given rise to what has become known as the "forecast combination puzzle" (see, for example, Stock and Watson, 2004; Timmermann, 2006). As a partial explanation, Smith and Wallis (2009) point out that, in analogy to the usual comparison of the LS and weighted LS estimators, relatively sophisticated combinations based on estimated weights suffer from an additional source of small-sample error that is not there when setting the weights equal, and that this might well account the difference in performance. Similar results have been documented also in the literature on interactive effects panel data regression models, in which CA estimation of the common factors tends to lead to better performance than when said estimation is carried out using PC (see, for example, Chudik et al., 2011; Westerlund and Urbain, 2015).

The current paper is motivated by the discussion in the last paragraph. The purpose is to study the asymptotic and finite-sample properties of $\widehat{\delta}$ and $\widehat{y}_{T+h|T}$ when F_t is estimated using the CAs of $x_{i,t}$. The use of these CAs in the current context not only simplifies considerably the implementation of the forecasting exercise, but is in fact quite natural in the sense that it uses the (sample) mean of $x_{i,t}$ to estimate the (conditional) mean of y_{T+h}. An important issue is, as it turns out, how many factors there are. The existing theory based on PC estimation assume that r is known (see, for example, Bai and Ng, 2006; Djogbenou et al., 2015, 2017; Gonçalves and Perron, 2014; Stock and Watson, 2002a), which is not realistic. One may of course argue, as is indeed commonly done, that r can be consistently estimated and therefore that the known r assumption is without loss of generality. As the bulk of the Monte Carlo evidence shows, however, r is a difficult object to estimate, which is reflected also in the empirical literature (see, for example, Breitung and Eickmeier, 2011; Breitung and Pigorsch, 2013; Moon and Weidner, 2015; Stock and Watson, 2005, 2009). Typically, one ends up with too many factors, and as a result many researches have chosen to work with a fixed number (see, for example, Cheng and Hansen, 2015; De Mol et al., 2008; Moon and Weidner, 2015; Stock and Watson, 2002a, 2002b, 2009). It is therefore important to consider the case when the number of estimated factors is larger than r, which closely related to the "oversampling problem" discussed by Boivin and Ng (2006). One novelty of the present paper is therefore to relax the otherwise so common known r assumption.

Clearly, $\widehat{y}_{T+h|T}$ depends not only on $\widehat{\delta}$ but also on $\widehat{F}_t$ through $\widehat{z}_t$. Thus, to study the behavior of $\widehat{y}_{T+h|T}$, we must examine the asymptotic properties of $\widehat{\delta}$ and $\widehat{F}_t$. The number of factors is, as already pointed out, treated as unknown and estimated using the $m \geq r$ CAs of $x_{i,t}$.[2] Karabiyik et al. (2017) considered the pooled LS estimator of a factor-augmented panel data regression with r factors that are estimated using $m \geq r$ CAs. According to their results, while consistent, inference is impaired by the presence of a bias in the asymptotic distribution of the estimator. The exact form of the bias depends on whether $r = m$ or $r < m$; however, this is as far as the dependence on m and r goes. Moreover, the bias is decreasing (increasing) in N (T), which means that if $T/N \to 0$ there is no dependence on m and r at all. The results reported in the present paper are quite different. In particular, while $\widehat{\beta}$ is consistent and asymptotically normal, unless $m = r$, $\widehat{\alpha}$ is generally inconsistent (for the space spanned by α). Interestingly, in spite of this inconsistency, $\widehat{y}_{T+h|T}$ is still consistent for $y_{T+h|T}$ and asymptotically normal. Hence, the inconsistency of $\widehat{\alpha}$ does not interfere with the asymptotic properties of $\widehat{y}_{T+h|T}$. It does, however, affect inference, as the asymptotic variance of $\widehat{y}_{T+h|T}$ is inestimable when $m > r$. This means that confidence intervals for $y_{T+h|T}$ will not have correct asymptotic coverage. However, the coverage error goes in the "right" direction in the sense that the asymptotic coverage is at least as large as the nominal coverage. Hence, while the asymptotic coverage is not correct, we can control the type I error rate.

The balance of the paper is organized as follows. In Section 2, we present and discuss the

[2]When $r < m$, the LS estimator will be inconsistent, since then there are unattended factors in (1) that may be correlated with W_t. We therefore restrict attention to the case $m \geq r$.

assumptions, which are used in Section 3 to derive our asymptotic results. Section 4 presents the results of a brief Monte Carlo study. Section 5 is concerned with an empirical application aimed at forecasting industrial production (IP) of the US. Section 6 concludes. All proofs are given in the Appendix.

2 Assumptions

The conditions under which we will be working are summarized in Assumptions A–D. Here and throughout this paper $\text{tr}\, A$, $\text{rk}\, A$, A^+ and $\|A\| = \sqrt{\text{tr}(A'A)}$ denote the trace, the rank, the generalized Moore–Penrose inverse, and the Frobenius (Euclidean) norm, respectively, of the matrix A. For any two matrices A and B, $\text{diag}(A, B)$ denotes the block-diagonal matrix that takes A (B) as the upper left (lower right) block. For any matrix A_i, we use $\overline{A} = N^{-1} \sum_{i=1}^N A_i$ to denote its CA. Moreover, $\to_d$, $\to_p$ and $\to_{a.s.}$ signify convergence in distribution, convergence in probability and convergence almost surely, respectively.

Assumption A. λ_i is either random such that $E(\|\lambda_i\|^4) < \infty$, or non-random such that $\|\lambda_i\| < \infty$. In either case, $\text{rk}\, \overline{\lambda} = r \leq m$ for all N, including $N \to \infty$.

Assumption B.

1. $E(e_{i,t}) = 0_{m \times 1}$ and $E(\|e_{i,t}\|^8) < \infty$ for all i and t.
2. $(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \|\Sigma_{e,ijts}\|$, $N^{-1} \sum_{i=1}^N \sum_{j=1}^N \|\Sigma_{e,ijts}\|$ and $T^{-1} \sum_{t=1}^T \sum_{s=1}^T \|\Sigma_{e,ijts}\|$ are all finite, where $\Sigma_{e,ijts} = E(e_{i,t}e'_{j,s})$ and $\|\Sigma_{e,ijts}\| < \infty$ for all i, j, t and s.
3. $E(\|N^{-1/2} \sum_{i=1}^N (e_{i,t}e'_{i,s} - \Sigma_{e,iits})\|^4) < \infty$ for all t and s, and $E(\|(NT)^{-1/2} \sum_{i=1}^N \sum_{t=1}^T (e_{i,t}e'_{i,t} - \overline{\Sigma}_e)\|^4) < \infty$, where $\overline{\Sigma}_e = \lim_{N,T\to\infty} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \Sigma_{e,iitt}$.
4. $N^{-1/2} \sum_{i=1}^N e_{i,t} \to_d N(0_{m\times 1}, \Sigma_{e,t})$ as $N \to \infty$, where $\Sigma_{e,t} = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \Sigma_{e,ijtt}$.

Assumption C. λ_i, F_t and $e_{i,t}$ are mutually independent groups. Dependence within each group is allowed.

Assumption D.

1. $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, ...) = 0$ for $h > 0$.

2. z_t and ε_t are independent of $e_{i,s}$ for all i, t and s.

3. $E(\varepsilon_t^2) = \sigma_{\varepsilon,t}^2 < \infty$ and $T^{-1} \sum_{t=1}^{T} z_t z_t' \to_p \Sigma_z$ as $T \to \infty$, where the $(n+r) \times (n+r)$ matrix Σ_z is positive definite and $E(\|z_t\|^4) < \infty$

4. $T^{-1/2} \sum_{t=1}^{T-h} \varepsilon_{t+h} z_t \to_d N(0_{(r+n)\times 1}, \Sigma_{z\varepsilon})$ as $T \to \infty$, where the $(n+r) \times (n+r)$ matrix $\Sigma_{z\varepsilon} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T-h} E(\varepsilon_{t+h}^2 z_t z_t')$ is positive definite.

Assumptions A–D are almost the same as in Bai and Ng (2006). The only difference is Assumption A, which supposes that $\overline{\lambda}$ has rank $r \leq m$. This should be compared to Assumption B of Bai and Ng (2006), which requires that $N^{-1} \sum_{i=1}^{N} \lambda_i \lambda_i'$ is positive definite. Hence, while in PC each factor has to have a nontrivial contribution to the variance of $x_{i,t}$, in the CA approach considered here it is the contribution to the mean that matters. Assumption B allows $e_{i,t}$ to be both weakly dependent and heteroskedastic across time and cross-section units. Assumption C allows for serially correlated factors and cross-sectionally correlated factor loadings. Assumption D allows F_t and W_t to be correlated both with each other and over time.

3 Asymptotic results

We want to use $\widehat{y}_{T+h|T} = \widehat{\delta}'\widehat{z}_T$ to infer $y_{T+h|T} = \delta' z_T$. As already mentioned, the previous research has focused on the case in which r is known. One of the contributions of the current paper is to relax this assumption. As we will now show, not requiring $m = r$ has important consequences. We begin by noting how

$$\widehat{F}_t = \overline{x}_t = \overline{\lambda}' F_t + \overline{e}_t. \tag{4}$$

Remark 1. The use of equal weights makes CA similar to the block-factor approach of Moench et al. (2013), in which the data are divided up into a number of categories, and one PC factor is calculated for each category.

It is important to realize that the $r \times m$ matrix $\overline{\lambda}$ is not necessarily a square matrix. Let us therefore assume without loss of generality that the columns of $\overline{\lambda}$ can be organized such that $\overline{\lambda} = [\overline{\lambda}_r, \overline{\lambda}_{-r}]$, where $\overline{\lambda}_r$ is an $r \times r$ matrix that is asymptotically of full rank, and $\overline{\lambda}_{-r}$ is $r \times (m-r)$. Let us similarly partition $\overline{e}_t = [\overline{e}_{r,t}', \overline{e}_{-r,t}']'$, where $\overline{e}_{r,t}$ and $\overline{e}_{-r,t}$ are $r \times 1$ and $(m-r) \times 1$, respectively. If $r = m$, then we define $\overline{\lambda} = \overline{\lambda}_r$ and $\overline{e}_t = \overline{e}_{r,t}$. The rationale for considering $\widehat{F}_t = \overline{x}_t$ is that under Assumption B, we have $\|\overline{e}_t\| = O_p(N^{-1/2})$ uniformly in t, which means that $\|\widehat{F}_t - \overline{\lambda}' F_t\| = O_p(N^{-1/2})$. The fact that $\widehat{F}_t$ is consistent for $\overline{\lambda}' F_t$ and not for F_t itself is worthy of some discussion.

Suppose first that $r = m$. The fact that in this case $\overline{\lambda}$ is of full rank means that it can be absorbed into $\widehat{\alpha}$. In order to appreciate this, note how $\alpha' F_t = \alpha'(\overline{\lambda}^{-1})'\overline{\lambda}' F_t$, which in turn implies that $\widehat{\alpha}'\widehat{F}_t = \alpha' F_t + (\widehat{\alpha} - \overline{\lambda}^{-1}\alpha)'\overline{\lambda}' F_t + \widehat{\alpha}'(\widehat{F}_t - \overline{\lambda}' F_t)$. In Theorem 1 below we show that $\|\widehat{\alpha} - \overline{\lambda}^{-1}\alpha\| = o_p(1)$. Hence, since $\|\widehat{F}_t - \overline{\lambda}' F_t\| = o_p(1)$, we have that

$$\widehat{\alpha}'\widehat{F}_t = \alpha' F_t + o_p(1) \tag{5}$$

uniformly in t as $N \to \infty$. In other words, while we cannot estimate α and F_t separately, we can still consistently estimate their product, which is enough for our purposes. The problem is that when $r < m$, $\overline{\lambda}^{-1}$ is no longer defined, which means that $\widehat{\alpha}'\widehat{F}_t$ is not necessarily a good estimator of $\alpha' F_t$.

As the above discussion suggests, tackling the case when $r < m$ requires a more elaborate approach than when $r = m$. We begin by rotating the coordinate system, as in Karabiyik et al. (2017). Let us therefore define the $m \times m$ rotation matrix Λ, which is such if $r < m$, then

$$\Lambda = \begin{bmatrix} \overline{\lambda}_r^{-1} & -\overline{\lambda}_r^{-1}\overline{\lambda}_{-r} \\ 0_{(m-r)\times r} & I_{m-r} \end{bmatrix} = [\Lambda_r, \Lambda_{-r}],$$

where $\Lambda_r = [\overline{\lambda}_r^{-1\prime}, 0'_{(m-r)\times r}]'$ is $m \times r$ and $\Lambda_{-r} = [-\overline{\lambda}'_{-r}\overline{\lambda}_r^{-1\prime}, I_{m-r}]'$ is $m \times (m-r)$. If, on the other hand, $r = m$, then we define $\Lambda = \Lambda_r = \overline{\lambda}_r^{-1} = \overline{\lambda}^{-1}$. Consider $\Lambda'\widehat{F}_t$. If $r = m$, then

$$\Lambda'\widehat{F}_t = \Lambda'\overline{\lambda}' F_t + \Lambda'\overline{e}_t = F_t + \overline{\lambda}^{-1\prime}\overline{e}_t. \tag{6}$$

Pre-multiplication by Λ' therefore takes care of the scaling by $\overline{\lambda}$. If, however, $r < m$, then $\overline{\lambda}\Lambda = [I_r, 0_{r\times(m-r)}]$, and so

$$\Lambda'\widehat{F}_t = \Lambda'\overline{\lambda}' F_t + \Lambda'\overline{e}_t = \begin{bmatrix} F_t \\ 0_{(m-r)\times 1} \end{bmatrix} + \begin{bmatrix} \overline{\lambda}_r^{-1\prime}\overline{e}_{r,t} \\ \overline{e}_{-r,t} - \overline{\lambda}_r^{-1\prime}\overline{\lambda}'_{-r}\overline{e}_{r,t} \end{bmatrix}. \tag{7}$$

Hence, since $\|\overline{e}_t\| = O_p(N^{-1/2})$, we have that the last $m - r$ rows of $\Lambda'\widehat{F}_t$ are degenerate, which in turn means that the associated second order moment matrix is almost surely singular. This situation is similar to the one that occurs when fitting regressions involving regressors that are of different orders of integration (see, for example, Chang and Phillips, 1995), and the solution is the same. Specifically, in order to account for the limiting singularity, we introduce the $m \times m$

normalization matrix D_N, which is $D_N = I_m$ if $r = m$ and $D_N = \text{diag}(I_r, \sqrt{N} I_{m-r})$ if $r < m$. Define $F_t^0 = D_N \Lambda' \overline{\lambda}' F_t$ and $\overline{e}_t^0 = D_N \Lambda' \overline{e}_t$. If $r = m$, then $F_t^0 = F_t$ and $\overline{e}_t^0 = \overline{\lambda}^{-1\prime} \overline{e}_t$, whereas if $r < m$, then $F_t^0 = [F_t', 0_{(m-r)\times 1}']'$ and $\overline{e}_t^0 = [\overline{e}_{r,t}^{0\prime}, \overline{e}_{-r,t}^{0\prime}]' = [\overline{e}_{r,t}' \overline{\lambda}_r^{-1}, \sqrt{N}(\overline{e}_{-r,t}' - \overline{e}_{r,t}' \overline{\lambda}_{-r} \overline{\lambda}_r^{-1})]'$. The normalization by D_N ensures that the second order moment matrix of

$$\widehat{F}_t^0 = D_N \Lambda' \widehat{F}_t = F_t^0 + \overline{e}_t^0 \tag{8}$$

is almost surely positive definite regardless of whether $r = m$ or $r < m$. In the asymptotic analysis, $\widehat{F}_t^0$ is the estimator of interest. The fact that this estimator is in practice infeasible is not a problem as D_N and Λ are both invertible, which means expressions involving $\widehat{F}_t$ can be stated in terms of $\widehat{F}_t^0$, and vice versa. Note in particular that if we let $Q_N = \text{diag}(\Lambda D_N, I_n)$, $\widehat{z}_t^0 = Q_N' \widehat{z}_t = [\widehat{F}_t^{0\prime}, W_t']'$ and $\delta^0 = [\alpha'(D_N \Lambda' \overline{\lambda}')^+, \beta']'$, where $(D_N \Lambda' \overline{\lambda}') = I_m$ if $r = m$ and $(D_N \Lambda' \overline{\lambda}') = [I_r, 0_{r\times(m-r)}]'$ if $r < m$, then

$$\begin{aligned}
y_{T+h|T} &= \alpha' F_T + \beta' W_T \\
&= \alpha'(D_N \Lambda' \overline{\lambda}')^+ D_N \Lambda' \overline{\lambda}' F_T + \beta' W_T \\
&= \alpha'(D_N \Lambda' \overline{\lambda}')^+ \widehat{F}_T^0 + \beta' W_T - \alpha'(D_N \Lambda' \overline{\lambda}')^+ (\widehat{F}_T^0 - F_T^0) \\
&= \delta^{0\prime} \widehat{z}_T^0 - \alpha'(D_N \Lambda' \overline{\lambda}')^+ (\widehat{F}_T^0 - F_T^0).
\end{aligned}$$

Hence, since $\widehat{y}_{T+h|T} = \widehat{\delta}'(Q_N^{-1})' \widehat{z}_T^0$, we can show that

$$\widehat{y}_{T+h|T} - y_{T+h|T} = T^{-1/2} \sqrt{T}(\widehat{\delta} - Q_N \delta^0)'(Q_N^{-1})' \widehat{z}_T^0 + N^{-1/2} (D_N \Lambda' \overline{\lambda}')^+ \sqrt{N}(\widehat{F}_T^0 - F_T^0). \tag{9}$$

The asymptotic distribution of $\widehat{y}_{T+h|T} - y_{T+h|T}$ therefore has two sources; $\sqrt{T}(\widehat{\delta} - Q_N \delta^0)$ and $\sqrt{N}(\widehat{F}_T^0 - F_T^0)$. Theorem 1 reports the asymptotic distribution of the first term. The theorem is stated in terms of B, which is such that

$$B = \begin{bmatrix} 0_{r\times 1} \\ (\Lambda_{-r}' \overline{\Sigma}_e \Lambda_{-r})^{-1} \Lambda_{-r}' \overline{\Sigma}_e \Lambda_r \alpha \\ 0_{n\times 1} \end{bmatrix}$$

if $r < m$ and $B = 0_{(m+n)\times 1}$ if $r = m$.

Theorem 1. *Suppose that Assumptions A–D hold. Then, as N, $T \to \infty$ with $\sqrt{T}/N \to 0$,*

$$Q_N^{-1} \sqrt{T}(\widehat{\delta} - Q_N \delta^0) + \sqrt{T} N^{-1/2} B \to_d N(0_{(m+n)\times 1}, \Sigma_{g^0}^{-1} \Sigma_{g^0 \varepsilon} \Sigma_{g^0}^{-1}),$$

where

$$\Sigma_{g^0} = \operatorname*{plim}_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^{T-h} g_t^0 g_t^{0\prime},$$

$$\Sigma_{g^0\varepsilon} = \lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^{T-h} E(\varepsilon_{t+h}^2 g_t^0 g_t^{0\prime}),$$

with $g_t^0 = [F_t', \bar{e}_{-r,t}^{0\prime}, W_t]'$.

Suppose first that $r < m$. Let $\widehat{\alpha} = [\widehat{\alpha}_r', \widehat{\alpha}_{-r}']'$, where $\widehat{\alpha}_r$ and $\widehat{\alpha}_{-r}$ are $r \times 1$ and $(m-r)\times 1$, respectively. In this notation, we have

$$Q_N^{-1}\sqrt{T}(\widehat{\delta} - Q_N\delta^0) + \sqrt{T}N^{-1/2}B = \sqrt{T}\left[\begin{array}{c} \overline{\lambda}\widehat{\alpha} - \alpha \\ N^{-1/2}[\widehat{\alpha}_{-r} + (\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha] \\ \widehat{\beta} - \beta \end{array}\right],$$

which according to Theorem 1 is asymptotically normal. We can therefore show that $\widehat{\beta}$ is $\sqrt{T}$-consistent for β and asymptotically normal, which is in accordance with our expectations. The most striking observations are, however, related to $\widehat{\alpha}$. The first thing to note is that under the assumptions of Theorem 1, $\widehat{\alpha}_{-r}$ is not necessarily convergent, and that it is only when $\sqrt{T}N^{-1/2} \to \infty$ that $[\widehat{\alpha}_{-r} + (\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha] \to_p 0_{(m-r)\times 1}$.[3] But even if $\widehat{\alpha}_{-r}$ is in fact convergent, the rate at which this happens is slower than the usual $\sqrt{T}$ rate. In fact, under the Theorem 1 requirement that $\sqrt{T}/N \to 0$, we have $(\sqrt{T}N^{-1/2})T^{-1/4} \to 0$, which means that the rate of convergence is slower than $T^{1/4}$. Hence, since $\widehat{\delta}$ is dominated by the component that converges slower, the overall convergence rate is given by $\sqrt{T}N^{-1/2}$. The limit of $\widehat{\alpha}_{-r}$ is also interesting. To the extent that $\widehat{\delta}$ can be viewed as an estimator of $Q_N\delta^0$, the true value of $\widehat{\alpha}_{-r}$ is given by the zero vector. The fact that $(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha \neq 0_{(m-r)\times 1}$ whenever $\alpha \neq 0_{(m-r)\times 1}$ means that $\widehat{\alpha}_{-r}$ is generally inconsistent. Interestingly enough, in spite of all these problems, the linear combination $\overline{\lambda}\widehat{\alpha} = \overline{\lambda}_r\widehat{\alpha}_r + \overline{\lambda}_{-r}\widehat{\alpha}_{-r}$ is still $\sqrt{T}$-consistent for α. Of course, since $\overline{\lambda}$ is unknown, in practice this rate of convergence is not attainable. This finding is similar in spirit to the results reported by Chang and Phillips (1995), where the rate of convergence depends on whether or not the non-stationarity characteristics of the regressors are known.

Let us now consider the case when $r = m$, in which

$$Q_N^{-1}\sqrt{T}(\widehat{\delta} - Q_N\delta^0) = \sqrt{T}\left[\begin{array}{c} \overline{\lambda}\widehat{\alpha} - \alpha \\ \widehat{\beta} - \beta \end{array}\right].$$

[3]By using the results provided in the proof of Theorem 1, we can further show that $\sqrt{T}(\overline{\lambda}\widehat{\alpha} - \alpha)$ and $\sqrt{T}N^{-1/2}[\widehat{\alpha}_{-r} + (\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha]$ are asymptotically independent.

Hence, since $\overline{\lambda}$ is invertible in this case, we have that $\widehat{\alpha}$ is $\sqrt{T}$-consistent for $\overline{\lambda}^{-1}\alpha$, and asymptotically normal. This is consistent with the heuristic discussion provided in the beginning of this section.

The above results have implications for $\widehat{y}_{T+h|T} - y_{T+h|T}$. Note in particular how $\widehat{\delta}$ enters (9) through $Q_N^{-1}\sqrt{T}(\widehat{\delta} - Q_N\delta^0)$. This suggests treating $\sqrt{T}N^{-1/2}B$ as an asymptotic bias whenever $r < m$. We might therefore expect the asymptotic distribution of $\widehat{y}_{T+h|T} - y_{T+h|T}$ to depend on whether $r < m$ or $r = m$. Theorem 2 confirms this. Before stating the theorem, however, it is useful to introduce $P_{\Lambda_{-r}}$, which is such that $P_{\Lambda_{-r}} = \Lambda_{-r}(\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}$ if $r < m$ and $P_{\Lambda_{-r}} = 0_{m\times m}$ if $r = m$.

Theorem 2. *Under the conditions of Theorem 1,*

$$t(y_{T+h|T}) = \frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{T^{-1}\phi + N^{-1}\Phi^{0\prime}\Sigma_e\Phi^0}} \to_d N(0,1),$$

where

$$\begin{aligned}
\Sigma_e &= \lim_{T\to\infty} \Sigma_{e,T}, \\
\phi &= \phi^0 + \operatorname{tr}(P_{\Lambda_{-r}}\Sigma_{\bar{e}\varepsilon}P_{\Lambda_{-r}}\Sigma_e), \\
\phi^0 &= \lim_{T\to\infty} z_T'\Sigma_z^{-1}\Sigma_{z\varepsilon}\Sigma_z^{-1}z_T, \\
\Phi^0 &= (I_m - P_{\Lambda_{-r}}\overline{\Sigma}_e)\Lambda_r\alpha, \\
\Sigma_{\bar{e}\varepsilon} &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-h}\sigma^2_{\varepsilon,t+h}\Sigma_{e,t}.
\end{aligned}$$

Because the two variance terms vanish at different rates, the rate of consistency of $\widehat{y}_{T+h|T}$ is $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. To see this, note that

$$t(y_{T+h|T}) = \frac{\delta_{NT}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{\delta^2_{NT}T^{-1}\phi + \delta^2_{NT}N^{-1}\Phi^{0\prime}\Sigma_e\Phi^0}}, \tag{10}$$

where the denominator is clearly bounded. The rate of consistency is therefore given by δ_{NT}, which is unexpected given the relatively slow convergence rate of $\widehat{\alpha}_{-r}$. Corollary 1 summarizes this.

Corollary 1. *Under the conditions of Theorem 1,*

$$\delta_{NT}(\widehat{y}_{T+h|T} - y_{T+h|T}) = O_p(1).$$

Theorem 2 (and hence also Corollary 1) does not impose any restrictions on N/T. However, Theorem 2 do simplify if either $N/T \to 0$ or $T/N \to 0$. On the one hand, if $N/T \to 0$, then $\delta_{NT} = \sqrt{N}$ and so (10) reduces to

$$t(y_{T+h|T}) = \frac{\sqrt{N}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{NT^{-1}\phi + \Phi^{0\prime}\Sigma_e\Phi^0}} = \frac{\sqrt{N}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{\Phi^{0\prime}\Sigma_e\Phi^0}} + o_p(1) \to_d N(0,1) \tag{11}$$

as N, $T \to \infty$ with $\sqrt{T}/N \to 0$ and $N/T \to 0$. If, on the other hand, $T/N \to 0$, then $\delta_{NT} = \sqrt{T}$ and so

$$t(y_{T+h|T}) = \frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{\phi}} + o_p(1) \to_d N(0,1) \tag{12}$$

as N, $T \to \infty$. These results apply regardless of whether $r < m$ or $r = m$. The only difference is that if $r = m$, then ϕ and Φ^0 reduce to $\phi = \phi^0$ and $\Phi^0 = \overline{\lambda}^{-1}\alpha$, respectively.

Since the variance terms are unknown, the asymptotic $N(0,1)$ distribution theory reported in Theorem 2 is not very useful to us. What is missing here is consistent estimators of these variances. Analogous to Bai and Ng (2006), a natural candidate for an estimator of the first of the two variance terms is given by

$$\widehat{\phi} = \widehat{z}_T'\widehat{\Sigma}_z^+\widehat{\Sigma}_{z\varepsilon}\widehat{\Sigma}_z^+\widehat{z}_T, \tag{13}$$

where

$$\widehat{\Sigma}_{z\varepsilon} = \frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t\widehat{z}_t',$$
$$\widehat{\Sigma}_z = \frac{1}{T}\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t',$$

with $\widehat{\varepsilon}_{t+h} = y_{T+h} - \widehat{\delta}'\widehat{z}_t$. The estimator of the second term is given by $\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}$, where

$$\widehat{\Sigma}_e = \frac{1}{N}\sum_{i=1}^{N}\widehat{e}_{i,T}\widehat{e}_{i,T}' \tag{14}$$

with $\widehat{e}_{i,t} = x_{i,t} - \widehat{\lambda}_i'\widehat{F}_t$ and $\widehat{\lambda}_i$ being the LS slope estimator in a time series regression of $x_{i,t}$ onto $\widehat{F}_t$. The resulting feasible version of $t(y_{T+h|T})$ is given by

$$\widehat{t}(y_{T+h|T}) = \frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{T^{-1}\widehat{\phi} + N^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}}.$$

Remark 2. The above formula for $\widehat{\Sigma}_e$ is appropriate if $e_{i,t}$ is uncorrelated across the cross-section. If $e_{i,t}$ are weakly cross-section correlated and/or cross-section heteroskedastic, then we follow Bai and Ng (2006), and use

$$\widehat{\Sigma}_e = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\widehat{e}_{i,t}\widehat{e}_{j,t}', \tag{15}$$

where n is a cross-section truncation parameter satisfying $n/\delta_{NT} \to 0$.

Theorem 3. *Suppose that Assumptions A–D hold and $r < m$. Then, as $N,\ T \to \infty$ with $\sqrt{T}/N \to 0$ and $N/T \to \tau < \infty$,*

$$\widehat{\phi} \to_d \phi^0 + Z_m'\Sigma_e^{1/2\prime}P_{\Lambda_{-r}}\Sigma_{\bar{e}\varepsilon}P_{\Lambda_{-r}}\Sigma_e^{1/2}Z_m,$$
$$\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha} \to_d \Phi'\Sigma_e\Phi,$$

where $Z_k \sim N(0_{k\times 1}, I_k)$ for any $k \geq 1$, and

$$\Phi = \begin{bmatrix} \overline{\lambda}_r^{-1}\alpha \\ -(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha + \sqrt{\tau}\Omega^{1/2}Z_{m-r} \end{bmatrix},$$
$$\Omega = (\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\Sigma_{\bar{e}\varepsilon}\Lambda_{-r}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}.$$

Theorem 3 shows that $\widehat{\phi}$ and $\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}$ are inconsistent, and that they actually converge to random variables rather than to constants. This means that if $r < m$, we cannot use $\widehat{\phi}$ and $\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}$ to estimate the asymptotic variance of $\widehat{y}_{T+h|T} - y_{T+h|T}$.

Remark 3. The randomness of Φ disappears if $\tau = 0$. However, $\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}$ is still not consistent for $\Phi^{0\prime}\Sigma_e\Phi^0$. This is clear from noting that under $\tau = 0$,

$$\Phi = \begin{bmatrix} \overline{\lambda}_r^{-1}\alpha \\ -(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha \end{bmatrix} = \begin{bmatrix} \overline{\lambda}_r^{-1} \\ -(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r \end{bmatrix}\alpha$$
$$= \left(I_m - \begin{bmatrix} 0_{r\times(m-r)} \\ I_{m-r} \end{bmatrix}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\right)\Lambda_r\alpha \neq \Phi^0.$$

Corollary 2 is the $r = m$ counterpart of Theorem 3.

Corollary 2. *Suppose that Assumptions A–D hold and $r = m$. Then, as $N,\ T \to \infty$ with $\sqrt{T}/N \to 0$,*

$$\widehat{\phi} \to_p \phi^0,$$
$$\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha} \to_p \alpha'\overline{\lambda}^{-1\prime}\Sigma_e\overline{\lambda}^{-1}\alpha.$$

Corollary 2 and Theorem 2 imply that if $r = m$, then

$$\widehat{t}(y_{T+h|T}) = \frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{T^{-1}\phi^0 + N^{-1}\alpha'\overline{\lambda}^{-1\prime}\Sigma_e\overline{\lambda}^{-1}\alpha}} + o_p(1) \to_d N(0,1) \tag{16}$$

as $N, T \to \infty$ with $\sqrt{T}/N \to 0$. Hence, under these conditions, a $100(1-\alpha)\%$ confidence interval for $y_{T+h|T}$ can be easily constructed as

$$CI_\alpha(y_{T+h|T}) = [\,\widehat{y}_{T+h|T} - z_{\alpha/2} \cdot \sqrt{T^{-1}\widehat{\phi} + N^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}, \\ \widehat{y}_{T+h|T} + z_{\alpha/2} \cdot \sqrt{T^{-1}\widehat{\phi} + N^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}\,], \tag{17}$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$ is the $(1-\alpha)$-th quantile of the standard normal cumulative distribution function, here denoted is denoted $\Phi(x)$. The asymptotic coverage of this confidence interval can be easily deduced from (16) and is given by

$$\lim_{N,T\to\infty} P(y_{T+h|T} \in CI_\alpha(y_{T+h|T})) = \lim_{N,T\to\infty} P(|\widehat{t}(y_{T+h|T})| \le z_{\alpha/2}) = 1 - \alpha. \tag{18}$$

This confidence interval is for the conditional mean of y_{T+h}. If we want a confidence interval for y_{T+h} itself, then we have to assume that $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Under this restrictions, a $100(1-\alpha)\%$ confidence interval for y_{T+h} is given by

$$CI_\alpha(y_{T+h}) = [\,\widehat{y}_{T+h|T} - z_{\alpha/2} \cdot \sqrt{\widehat{\sigma}_\varepsilon^2 + T^{-1}\widehat{\phi} + N^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}, \\ \widehat{y}_{T+h|T} + z_{\alpha/2} \cdot \sqrt{\widehat{\sigma}_\varepsilon^2 + T^{-1}\widehat{\phi} + N^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}\,], \tag{19}$$

where $\widehat{\sigma}_\varepsilon^2 = T^{-1}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2$, whose asymptotic coverage is again given by $1-\alpha$.

The problem with the above results is of course that they only apply when $r = m$, which is unlikely to be the case in practice. Interestingly, if we accept that the confidence intervals are conservative, then asymptotically valid inference is possible also when $r < m$. This is formalized in Theorem 4, which holds for all $r \le m$.

Theorem 4. *Suppose that Assumptions A–D hold. Then, as $N, T \to \infty$ with $\sqrt{T}/N \to 0$ and $N/T \to 0$,*

$$\lim_{N,T\to\infty} P(|\widehat{t}(y_{T+h|T})| > c_{\alpha/2}) \le \alpha.$$

Theorem 4 implies that

$$\lim_{N,T\to\infty} P(y_{T+h|T} \in CI_\alpha(y_{T+h|T})) \ge 1 - \alpha. \tag{20}$$

This holds for all $r \leq m$. If, however, $r = m$, then the asymptotic coverage is exactly $1 - \alpha$ and the $N/T \to 0$ requirement is no longer needed, as is evident from (16). Hence, while Theorem 4 only requires that $r \leq m$, the results do depend on whether $r = m$ or $r < m$. In the next section, we use Monte Carlo simulations to investigate the effect of the conservativeness when $r < m$ in small samples. According to the results, the effect of using the conservative critical values is almost nonexistent. Finally, note that while stated in terms of $CI_\alpha(y_{T+h|T})$, the result in (20) applies also to $CI_\alpha(y_{T+h})$.

4 Monte Carlo simulations

In this section, we evaluate the small-sample properties of the CA-based forecasting method proposed in this paper. The results are compared to those obtained when using the PC-based method of Bai and Ng (2006) and to those obtained by using the true factors as a benchmark case. The DGP used for this purpose is similar to the DGP used in the Monte Carlo simulations of Bai and Ng (2006) and it can be seen a restricted version of (1) and (2) that sets $h = 4$, $W_1 = \cdots = W_T = 1$, and $\varepsilon_t \sim N(0,1)$. Serial correlation in F_t is permitted through

$$F_t = \rho F_{t-1} + \sqrt{1-\rho^2} u_t, \tag{21}$$

where $\rho = 0.5$ and $u_t \sim N(0,1)$. Bai and Ng (2006) assume that $m = r = 1$. According to our results, however, it matters whether $r = m$ or $r < m$, and in this section we therefore set $m \in \{1,2\}$. The DGP of $e_{i,t}$ is similar to the one considered by Bai and Ng (2006), and is given by

$$e_t = \Omega(b)^{1/2} v_t, \tag{22}$$

where $v_t = [v_{1,t}, ..., v_{N,t}]'$ is $N \times m$ with $v_{i,t} \sim N(0_{m \times 1}, \sigma^2_{v,i} I_m)$ being $m \times 1$. Also, $\Omega(b)^{1/2}$ is the lower triangular Cholesky factor of the $N \times N$ Toeplitz matrix $\Omega(b)$, whose i-th diagonal element is b^i if $i \leq 10$ and is zero otherwise. Hence, in this DGP, the cross-section correlation and cross-section heteroskedasticity of $e_{i,t}$ is controlled by b and $\sigma^2_{v,i}$, respectively.

The following five parameterizations are considered.

DGP1: $r = m = 1$, $b = 0$ and $\sigma^2_{v,i} = 1$, $\lambda_{1,i} \sim U[0,1]$.

DGP2: $r = m = 1$, $b = 0.5$ and $\sigma^2_{v,i} = 1$, $\lambda_{1,i} \sim U[0,1]$.

DGP3: $r = m = 1$, $b = 0$ and $\sigma^2_{v,i} \sim U[0.5, 1.5]$, $\lambda_{1,i} \sim U[0, 1]$.

DGP4: $r = m = 1$, $b = 0.5$ and $\sigma^2_{v,i} \sim U[0.5, 1.5]$, $\lambda_{1,i} \sim U[0, 1]$.

DGP5: $r = 1 < m = 2$, $b = 0$ and $\sigma^2_{v,i} = 1$, $\lambda_{1,i} \sim U[0, 1]$, $\lambda_{2,i} \sim U[0, 0.5]$.

DGP6: $r = 1 < m = 2$, $b = 0.5$ and $\sigma^2_{v,i} = 1$, $\lambda_{1,i} \sim U[0, 1]$, $\lambda_{2,i} \sim U[0, 0.5]$.

DGP7: $r = 1 < m = 2$, $b = 0$ and $\sigma^2_{v,i} \sim U[0.5, 1.5]$, $\lambda_{1,i} \sim U[0, 1]$, $\lambda_{2,i} \sim U[0, 0.5]$.

DGP8: $r = 1 < m = 2$, $b = 0.5$ and $\sigma^2_{v,i} \sim U[0.5, 1.5]$, $\lambda_{1,i} \sim U[0, 1]$, $\lambda_{2,i} \sim U[0, 0.5]$.

For the estimation of the variances we use three types of estimators for three different factor choices. For the CCE approach we propose in ths paper we use the formulas given in (13) - (14). For the PC approach we use the cross-section correlation and heteroskedasticity robust one with n set as in Bai and Ng (2006) to $n = \lfloor \min\{\sqrt{N}, \sqrt{T}\} \rfloor$. In order to construct the confidence interval for the case we use the true factors, as in Bai and Ng (2006) we ignore the terms that are induced by the estimation of F_t from the variance.

We report the empirical coverage rate and the empirical mean squared prediction error. We consider four different approaches regarding the unobserved factors: (i) cross-sectional averages (CA), (ii) principal components where we used the true number of factors, r, (iii) principal components where we used one factor estimate one more than the true number of factors, $r + 1$. 95% confidence intervals are constructed for both $y_{T+h|T}$ and y_{T+h}. The results are based on 5,000 replications of samples of size $(N, T) \in \{30, 50, 100, 200\}$.

Results for DGP1–DGP8 are reported in Tables 1–8, respectively. The results reported in Tables 1–8 suggest that the coverage rates are close to the nominal rate of 95%.

5 Conclusion

This paper shows that forecasts of a single time series based on CAs of $m \geq r$ panel data variables are consistent and asymptotically normal as N, $T \to \infty$ with $\sqrt{T}/N \to 0$. A problem arises in the empirically relevant case when $m > r$. In particular, the use of too many CAs causes an inconsistency in the estimator of the asymptotic variance of the conditional mean, which means that the coverage of the resulting confidence intervals is incorrect. The coverage is, however, shown to be upwards biased, which means that the confidence intervals will be conservative. The Monte Carlo results suggest that these theoretical results provide a useful guide to empirical work in datasets of the size typically encountered in macroeconomic forecasting.

References

Bai, J., and S. Ng (2006). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica* **74**, 1133–1150.

Bai, J., and S. Ng (2009). Boosting Diffusion Indices. *Journal of Applied Econometrics* **24**, 607–629.

Banerjee, A., M. Marcellino and I. Masten (2008). Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change. In Rapach, D., and M. Wohar (Eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald Publishing. Bingley, UK.

Breitung, J., and J. Tenhofen (2011). GLS Estimation of Dynamic Factor Models. *Journal of the American Statistical Association* 106, 1150–1166.

Breitung, J., and U. Pigorsch (2013). A Canonical Correlation Approach for Selecting the Number of Dynamic Factors. *Oxford Bulletin of Economics and Statistics* **75**, 23–36.

Chang, Y., and P. C. B. Phillips (1995). Time Series Regression with Mixtures of Integrated Regressors. *Econometric Theory* **12**, 1033–1094.

Cheng, X., and B. E. Hansen (2015). Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach. *Journal of Econometrics* **86**, 280–293.

Choi, I. (2012). Efficient Estimation of Factor Models. *Econometric Theory* **28**, 274–308.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and Strong Cross Section Dependence and Estimation of Large Panels. *Econometric Journal* **14**, C45–C90.

Corradi, V., and N. R. Swanson (2014). Testing for Structural Stability of Factor Augmented Forecasting Models. *Journal of Econometrics* **182**, 100–118.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics* **146**, 318–328.

Djogbenou, A., S. Gonçalves, and B. Perron (2015). Bootstrap Inference in Regressions with Estimated Factors and Serial Correlation. *Journal of Time Series Analysis* **36**, 481–502.

Djogbenou, A., S. Gonçalves, and B. Perron (2017). Bootstrap Prediction Intervals for Factor Models. *Journal of Business & Economics Statistics* **35**, 53–69.

Eickmeier, S., and C. Ziegler (2008). How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach. *Journal of Forecasting* **27**, 237–265.

Gonçalves, S., and B. Perron (2014). Bootstrapping Factor-Augmented Regression Models. *Journal of Econometrics* **82**, 156–173.

Karabiyik, H., S. Reese and J. Westerlund (2017). On the Role of the Rank Condition in CCE Estimation of Factor-Augmented Panel Regressions. *Journal of Econometrics* **197**, 60–64.

Moench, E., S. Ng and S. Potter (2013). *The Review of Economics and Statistics* **95**, 1811–1817.

Moon, H. R., and M. Weidner (2015). Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects. *Econometrica* **83**, 1543–1579.

Park, J., and P. C. B. Phillips (2000). Nonstationary Binary Choice Models. *Econometrica* **68**, 1249–1280.

Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica* **74**, 967–1012.

Pesaran, M. H., T. Schuermann, and L. V. Smith (2009). Forecasting Economic and Financial Variables with Global VARs. *International Journal of Forecasting* **25**, 642–675.

Smith, J., and K. F. Wallis (2009). A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics* **71**, 331–355.

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

Stock, J. H., and M. W. Watson (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics* **20**, 147–162.

Stock, J. H., and M. W. Watson (2004). Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting* **23**, 405–430.

Stock, J. H., and M. W. Watson (2005). Implications of Dynamic Factor Models for VAR Analysis. NBER Working Paper No. 11467.

Stock, J. H., and M. W. Watson (2009). Forecasting in Dynamic Factor Models Subject to Structural Instability. In Shephard, N., and J. Castle (Eds.), *The Methodology and Practice of Econometrics: Festschrift in Honor of David F. Hendry*, 1–57. Oxford University Press.

Timmermann, A. (2006). Forecast Combinations. In Elliot, G., C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, 135–196. Amsterdam. Elsevier.

Westerlund, J., and J.-P. Urbain (2015). Cross-Sectional Averages Versus Principal Components. *Journal of Econometrics* **85**, 372–377.

Appendix: Proofs

Proof of Theorem 1.

Consider the case when $r < m$. We have $\widehat{F}_t = \overline{x}_t = \overline{\lambda}' F_t + \overline{e}_t$, which we can write in matrix format as

$$\widehat{F} = \overline{x} = F\overline{\lambda} + \overline{e}, \tag{A1}$$

where $\overline{x} = [\overline{x}_1, ..., \overline{x}_{T-h}]'$ is $(T-h) \times m$ and $F = [F_1, ..., F_{T-h}]'$ is $(T-h) \times r$. Let $\overline{e} = [\overline{e}_r, \overline{e}_{-r}]$. Hence, since $\overline{\lambda} = [\overline{\lambda}_r, \overline{\lambda}_{-r}]$, we have

$$\widehat{F} = [F\overline{\lambda}_r, F\overline{\lambda}_{-r}] + [\overline{e}_r, \overline{e}_{-r}]. \tag{A2}$$

Post-multiplying (A2) by Λ yields

$$\widehat{F}\Lambda = F\overline{\lambda}\Lambda + \overline{e}\Lambda = [F, 0_{(T-h)\times(m-r)}] + [\overline{e}_r\overline{\lambda}_r^{-1}, \overline{e}_{-r} - \overline{e}_r\overline{\lambda}_r^{-1}\overline{\lambda}_{-r}].$$

We now look for a conformable normalization matrix D_N such that

$$\begin{aligned} &T^{-1}D_N'\Lambda'\widehat{F}'\widehat{F}\Lambda D_N \\ &= T^{-1}D_N' \begin{bmatrix} F'F + \overline{\lambda}_r^{-1\prime}\overline{e}_r'\overline{e}_r\overline{\lambda}_r^{-1} & -\overline{\lambda}_r^{-1\prime}\overline{e}_r'(\overline{e}_{-r} - \overline{e}_r\overline{\lambda}_r^{-1}\overline{\lambda}_{-r}) \\ -(\overline{e}_{-r}' - \overline{\lambda}_{-r}'\overline{\lambda}_r^{-1\prime}\overline{e}_r')\overline{e}_r\overline{\lambda}_r^{-1} & (\overline{e}_{-r}' - \overline{\lambda}_{-r}'\overline{\lambda}_r^{-1\prime}\overline{e}_r')(\overline{e}_{-r} - \overline{e}_r\overline{\lambda}_r^{-1}\overline{\lambda}_{-r}) \end{bmatrix} D_N \end{aligned}$$

converges to a positive matrix. Here we make use of Lemma 1 of Pesaran (2006), which states that $\|\overline{e}_t\| = O_p(N^{-1/2})$. Hence, while the upper left $r \times r$ block of $T^{-1}D_N'\Lambda'\widehat{F}'\widehat{F}\Lambda D_N$ converges to a positive definite matrix, the lower right $(m-r)\times(m-r)$ block is $O_p(N^{-1})$. This means that the required normalization matrix is given by $D_N = \text{diag}(I_r, \sqrt{N}I_{m-r})$. Hence, letting $F^0 = [F, 0_{(T-h)\times(m-r)}]$ and $\overline{e}^0 = \overline{e}\Lambda D_N = [\overline{e}_r^0, \overline{e}_{-r}^0] = [\overline{e}_r\overline{\lambda}_r^{-1}, \sqrt{N}(\overline{e}_{-r} - \overline{e}_r\overline{\lambda}_r^{-1}\overline{\lambda}_{-r})]$, the resulting normalized version of $\widehat{F}\Lambda$ is given by

$$\widehat{F}^0 = \widehat{F}\Lambda D_N = F\overline{\lambda}\Lambda D_N + \overline{e}\Lambda D_N = F^0 + \overline{e}^0, \tag{A3}$$

or, in vector form,

$$\widehat{F}_t^0 = D_N\Lambda'\widehat{F}_t = D_N\Lambda'\overline{\lambda}'F_t + D_N\Lambda'\overline{e}_t = F_t^0 + \overline{e}_t^0. \tag{A4}$$

Note that since $\text{rk}\,\overline{\lambda} = r$ ($\overline{\lambda}$ has full row rank), we have $\overline{\lambda}^+ = \overline{\lambda}'(\overline{\lambda}\overline{\lambda}')^{-1}$, such that $\overline{\lambda}\overline{\lambda}^+ = \overline{\lambda}^{+\prime}\overline{\lambda}' = I_r$ (see Abadir and Magnus, 2005, Exercise 10.31). Also, $\overline{\lambda}\Lambda = [I_r, 0_{r\times(m-r)}]$, and so $\overline{\lambda}\Lambda D_N = [I_r, 0_{r\times(m-r)}]$, whose rank is r. Therefore, in analogy with $\overline{\lambda}^+$, we have $[I_r, 0_{r\times(m-r)}]^+ =$

$[I_r, 0_{r\times(m-r)}]'$. Making use of this, and letting $d_t^0 = \widehat{F}_t^0 - F_t^0 = \bar{e}_t^0$, $\widehat{z}_t^0 = [\widehat{F}_t^{0\prime}, W_t']'$ and $\delta^0 = [\alpha', 0_{1\times(m-r)}, \beta']'$, (1) can be rewritten as

$$\begin{aligned}
y_{t+h} &= \alpha' F_t + \beta' W_t + \varepsilon_{t+h} \\
&= \alpha'(D_N \Lambda' \overline{\lambda}')^+ D_N \Lambda' \overline{\lambda}' F_t + \beta' W_t + \varepsilon_{t+h} \\
&= \alpha' [I_r, 0_{r\times(m-r)}] F_t^0 + \beta' W_t + \varepsilon_{t+h} \\
&= [\alpha', 0_{1\times(m-r)}] \widehat{F}_t^0 + \beta' W_t + \varepsilon_{t+h} - [\alpha', 0_{1\times(m-r)}](\widehat{F}_t^0 - F_t^0) \\
&= \delta^{0\prime} \widehat{z}_t^0 + \varepsilon_{t+h} - [\alpha', 0_{1\times(m-r)}] d_t^0. \qquad \text{(A5)}
\end{aligned}$$

Note how $d^0[\alpha', 0_{1\times(m-r)}]' = [\bar{e}_r^0, \bar{e}_{-r}^0][\alpha', 0_{1\times(m-r)}]' = \bar{e}_r^0 \alpha$. The above equation can therefore be written in the following matrix notation:

$$y = \widehat{z}^0 \delta^0 + \varepsilon - d^0 [\alpha', 0_{1\times(m-r)}]' = \widehat{z}^0 \delta^0 + \varepsilon - \bar{e}_r^0 \alpha, \qquad \text{(A6)}$$

where $y = [y_{h+1}, ..., y_T]'$ and $\varepsilon = [\varepsilon_{h+1}, ..., \varepsilon_T]'$ are $(T-h) \times 1$, $\widehat{z}^0 = [\widehat{F}^0, W] = [\widehat{z}_1^0, ..., \widehat{z}_{T-h}^0]'$ is $(T-h) \times (m+n)$ and $d^0 = [d_1^0, ..., d_{T-h}^0]'$ is $(T-h) \times m$. In order to appreciate how $\widehat{z}^0$ relates to $\widehat{z}$ it is useful to note that

$$Q_N = \begin{bmatrix} \Lambda D_N & 0_{m\times n} \\ 0_{n\times m} & I_n \end{bmatrix} = \begin{bmatrix} \overline{\lambda}_r^{-1} & -\overline{\lambda}_r^{-1} \overline{\lambda}_{-r} & 0_{r\times n} \\ 0_{(m-r)\times r} & \sqrt{N} I_{m-r} & 0_{(m-r)\times n} \\ 0_{n\times r} & 0_{n\times(m-r)} & I_n \end{bmatrix},$$

which is $(m+n) \times (n+m)$. Note also that since ΛD_N is invertible, Q_N is invertible too. In fact, since

$$\Lambda^{-1} = \begin{bmatrix} \overline{\lambda}_r & \overline{\lambda}_{-r} \\ 0_{(m-r)\times r} & I_{m-r} \end{bmatrix}$$

(see Abadir and Magnus, 2005, Exercise 5.13), we can show that

$$Q_N^{-1} = \begin{bmatrix} (\Lambda D_N)^{-1} & 0_{m\times n} \\ 0_{n\times m} & I_n \end{bmatrix} = \begin{bmatrix} \overline{\lambda}_r & \overline{\lambda}_{-r} & 0_{r\times n} \\ 0_{(m-r)\times r} & N^{-1/2} I_{m-r} & 0_{(m-r)\times n} \\ 0_{n\times r} & 0_{n\times m} & I_n \end{bmatrix}.$$

This implies

$$\widehat{z} Q_N = [\widehat{F} \Lambda D_N, W] = [\widehat{F}^0, W] = \widehat{z}^0, \qquad \text{(A7)}$$

which can be substituted into $\widehat{\delta}$, giving

$$\begin{aligned}
\widehat{\delta} &= (\widehat{z}'\widehat{z})^+ \widehat{z}' y = Q_N (Q_N' \widehat{z}' \widehat{z} Q_N)^+ Q_N' \widehat{z}' y = Q_N (\widehat{z}^{0\prime} \widehat{z}^0)^+ \widehat{z}^{0\prime} y \\
&= Q_N \delta^0 + Q_N (\widehat{z}^{0\prime} \widehat{z}^0)^+ \widehat{z}^{0\prime} (\varepsilon - \bar{e}_r^0 \alpha), \qquad \text{(A8)}
\end{aligned}$$

where

$$Q_N\delta^0 = \begin{bmatrix} \overline{\lambda}_r^{-1}\alpha \\ 0_{(m-r)\times 1} \\ \beta \end{bmatrix} = \begin{bmatrix} \Lambda_r\alpha \\ \beta \end{bmatrix}.$$

Therefore,

$$Q_N^{-1}\sqrt{T}(\widehat{\delta} - Q_N\delta^0) = (T^{-1}\widehat{z}^{0\prime}\widehat{z}^0)^{+}T^{-1/2}\widehat{z}^{0\prime}(\varepsilon - \overline{e}_r^0\alpha). \tag{A9}$$

We now evaluate in turn $T^{-1/2}\widehat{z}^{0\prime}\overline{e}_r^0$, $T^{-1/2}\widehat{z}^{0\prime}\varepsilon$ and $T^{-1}\widehat{z}^{0\prime}\widehat{z}^0$. Consider $T^{-1/2}\widehat{z}^{0\prime}\overline{e}_r^0$. Let us define $z_t^0 = [F_t^{0\prime}, W_t^{\prime}]^{\prime}$ and $z^0 = [F^0, W] = [z_1^0, ..., z_{T-h}^0]^{\prime}$, which are $(m+n)\times 1$ and $(T-h)\times(m+n)$, respectively. Note how $\widehat{z}^0 - z^0 = [d^0, 0_{T\times n}]$, where $d^0 = \overline{e}^0 = [\overline{e}_r^0, \overline{e}_{-r}^0]$, implying

$$T^{-1}(\widehat{z}^0 - z^0)^{\prime}\overline{e}_r^0 = \begin{bmatrix} T^{-1}d^{0\prime}\overline{e}_r^0 \\ 0_{n\times r} \end{bmatrix} = \begin{bmatrix} T^{-1}\overline{e}^{0\prime}\overline{e}_r^0 \\ 0_{n\times r} \end{bmatrix} = \begin{bmatrix} T^{-1}\overline{e}_r^{0\prime}\overline{e}_r^0 \\ T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_r^0 \\ 0_{n\times r} \end{bmatrix}.$$

By (A.10) in Lemma 2 of Pesaran (2006), we have $\|T^{-1}\overline{e}_r^{\prime}\overline{e}_r\| = O_p(N^{-1})$, and by further use of Assumption C, $\|\overline{\lambda}_r\|$ and $\|\overline{\lambda}_r^{-1}\|$ are $O_p(1)$. It follows that

$$\|T^{-1}\overline{e}_r^{0\prime}\overline{e}_r^0\| = N^{-1}\|\overline{\lambda}_r^{-1\prime}(NT^{-1}\overline{e}_r^{\prime}\overline{e}_r)\overline{\lambda}_r^{-1}\| \leq N^{-1}\|\overline{\lambda}_r^{-1}\|^2\,\|NT^{-1}\overline{e}_r^{\prime}\overline{e}_r\| = O_p(N^{-1}).$$

The order of $\|T^{-1}\overline{e}_r^{0\prime}\overline{e}_{-r}^0\|$ is similarly given by

$$\begin{aligned}\|T^{-1}\overline{e}_r^{0\prime}\overline{e}_{-r}^0\| &= \|\sqrt{N}T^{-1}\overline{\lambda}_r^{-1\prime}\overline{e}_r^{\prime}\overline{e}_{-r} - \sqrt{N}T^{-1}\overline{\lambda}_r^{-1\prime}\overline{e}_r^{\prime}\overline{e}_r\overline{\lambda}_r^{-1}\overline{\lambda}_{-r}\| \\ &\leq N^{-1/2}\|\overline{\lambda}_r^{-1}\|\,\|NT^{-1}\overline{e}_r^{\prime}\overline{e}_{-r}\| + N^{-1/2}\|\overline{\lambda}_r^{-1}\|^2\,\|NT^{-1}\overline{e}_r^{\prime}\overline{e}_r\|\,\|\overline{\lambda}_{-r}\| \\ &= O_p(N^{-1/2}),\end{aligned} \tag{A10}$$

which is too large to be treated as negligible. It follows that

$$T^{-1}(\widehat{z}^0 - z^0)^{\prime}\overline{e}_r^0 = \begin{bmatrix} 0_{r\times r} \\ T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_r^0 \\ 0_{n\times r} \end{bmatrix} + O_p(N^{-1}). \tag{A11}$$

Moreover, in view of Assumptions B and D,

$$\|T^{-1/2}z^{0\prime}\overline{e}_r^0\| = O_p(N^{-1/2}), \tag{A12}$$

implying

$$\begin{aligned}T^{-1/2}\widehat{z}^{0\prime}\overline{e}_r^0 &= T^{-1/2}z^{0\prime}\overline{e}_r^0 + T^{-1/2}(\widehat{z}^0 - z^0)^{\prime}\overline{e}_r^0 \\ &= \sqrt{T}\begin{bmatrix} 0_{r\times r} \\ T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_r^0 \\ 0_{n\times r} \end{bmatrix} + O_p(N^{-1/2}) + O_p(\sqrt{T}N^{-1}).\end{aligned} \tag{A13}$$

Note that the first term is the leading terms and is of order $O_p(\sqrt{T}N^{-1/2})$.

Consider $T^{-1/2}\widehat{z}^{0\prime}\varepsilon$. By Assumptions B and D,

$$T^{-1/2}\varepsilon' d^0 = T^{-1/2}[\varepsilon'\overline{e}_r^0, \varepsilon'\overline{e}_{-r}^0] = [0_{1\times r}, T^{-1/2}\varepsilon'\overline{e}_{-r}^0] + O_p(N^{-1/2}), \tag{A14}$$

from which it follows that

$$T^{-1/2}(\widehat{z}^0 - z^0)'\varepsilon = \begin{bmatrix} T^{-1/2}d^{0\prime}\varepsilon \\ 0_{n\times 1} \end{bmatrix} = \begin{bmatrix} 0_{r\times 1} \\ T^{-1/2}\overline{e}_{-r}^{0\prime}\varepsilon \\ 0_{n\times 1} \end{bmatrix} + O_p(N^{-1/2}).$$

Letting $g^0 = [F, \overline{e}_{-r}^0, W]$, we consequently have

$$\begin{aligned} T^{-1/2}\widehat{z}^{0\prime}\varepsilon &= T^{-1/2}z^{0\prime}\varepsilon + T^{-1/2}(\widehat{z}^0 - z^0)'\varepsilon \\ &= \begin{bmatrix} T^{-1/2}F'\varepsilon \\ 0_{(m-r)\times 1} \\ T^{-1/2}W'\varepsilon \end{bmatrix} + \begin{bmatrix} 0_{r\times 1} \\ T^{-1/2}\overline{e}_{-r}^{0\prime}\varepsilon \\ 0_{n\times 1} \end{bmatrix} + O_p(N^{-1/2}) \\ &= T^{-1/2}g^{0\prime}\varepsilon + O_p(N^{-1/2}). \end{aligned} \tag{A15}$$

The above results show that the numerator of $Q_N^{-1}\sqrt{T}(\widehat{\delta} - \delta^0)$ can be written as

$$\begin{aligned} & T^{-1/2}\widehat{z}^{0\prime}(\varepsilon - \overline{e}_r^0\alpha) \\ &= T^{-1/2}\widehat{z}^{0\prime}\varepsilon - T^{-1/2}\widehat{z}^{0\prime}\overline{e}_r^0\alpha \\ &= T^{-1/2}g^{0\prime}\varepsilon - \sqrt{T}\begin{bmatrix} 0_{r\times 1} \\ T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_r^0\alpha \\ 0_{n\times 1} \end{bmatrix} + O_p(N^{-1/2}) + O_p(\sqrt{T}N^{-1}). \end{aligned} \tag{A16}$$

Let us now consider the denominator of $Q_N^{-1}\sqrt{T}(\widehat{\delta} - \delta^0)$, which is given by

$$T^{-1}\widehat{z}^{0\prime}\widehat{z}^0 = T^{-1}\begin{bmatrix} \widehat{F}^{0\prime}\widehat{F}^0 & \widehat{F}^{0\prime}W \\ W'\widehat{F}^0 & W'W \end{bmatrix}.$$

From the definition of $\widehat{F}^0$,

$$T^{-1}\widehat{F}^{0\prime}\widehat{F}^0 = T^{-1}F^{0\prime}F^0 + T^{-1}F^{0\prime}\overline{e}^0 + T^{-1}\overline{e}^{0\prime}F^0 + T^{-1}\overline{e}^{0\prime}\overline{e}^0, \tag{A17}$$

where

$$T^{-1}F^{0\prime}F^0 = \begin{bmatrix} T^{-1}F'F & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)} \end{bmatrix}.$$

Also,

$$T^{-1}F^{0\prime}\overline{e}^0 = \begin{bmatrix} T^{-1}F'\overline{e}_r\overline{\lambda}_r^{-1} & \sqrt{N}T^{-1}(F'\overline{e}_{-r} - F'\overline{e}_r\overline{\lambda}_r^{-1}\overline{\lambda}_{-r}) \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)} \end{bmatrix}.$$

By (A.11) in Lemma 2 of Pesaran (2006), $\|T^{-1}F'\overline{e}_r\|$ and $\|T^{-1}F'\overline{e}_{-r}\|$ are both $O_p((NT)^{-1/2})$. This implies

$$\|T^{-1}F^{0\prime}\overline{e}^0\| = O_p(T^{-1/2}). \tag{A18}$$

It remains to consider

$$T^{-1}\overline{e}^{0\prime}\overline{e}^0 = T^{-1}\begin{bmatrix} \overline{e}_r^{0\prime}\overline{e}_r^0 & \overline{e}_r^{0\prime}\overline{e}_{-r}^0 \\ \overline{e}_{-r}^{0\prime}\overline{e}_r^0 & \overline{e}_{-r}^{0\prime}\overline{e}_{-r}^0 \end{bmatrix}.$$

We have already shown that $\|T^{-1}\overline{e}_r^{0\prime}\overline{e}_r^0\| = O_p(N^{-1})$ and $\|T^{-1}\overline{e}_r^{0\prime}\overline{e}_{-r}^0\| = O_p(N^{-1/2})$. Consider $T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_{-r}^0$. We begin by noting how $\overline{e}_{-r}^0 = \sqrt{N}\overline{e}\Lambda_{-r}$, where $\Lambda_{-r} = [-\overline{\lambda}_{-r}'\overline{\lambda}_r^{-1\prime}, I_{m-r}]$ is as before, giving

$$T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_{-r}^0 = \Lambda_{-r}'(NT^{-1}\overline{e}'\overline{e})\Lambda_{-r}'. \tag{A19}$$

A straightforward calculation reveals that $E(\|N^{-1}T^{-1/2}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{t=1}^{T-h} e_{i,t}e_{j,t}'\|^2) = O(1)$, implying $\|N^{-1}T^{-1/2}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{t=1}^{T-h} e_{i,t}e_{j,t}'\| = O_p(1)$, and by further use of Assumption B, we have that $\|(NT)^{-1/2}\sum_{i=1}^N\sum_{t=1}^{T-h}(e_{i,t}e_{i,t}' - \overline{\Sigma}_e)\|$ must be of the same order. Hence,

$$\begin{aligned} NT^{-1}\overline{e}'\overline{e} &= \frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^{T-h} e_{i,t}e_{j,t}' \\ &= \overline{\Sigma}_e + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^{T-h}(e_{i,t}e_{i,t}' - \overline{\Sigma}_e) + \frac{1}{NT}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{t=1}^{T-h} e_{i,t}e_{j,t}' \\ &= \overline{\Sigma}_e + O_p(T^{-1/2}), \end{aligned} \tag{A20}$$

which in turn implies

$$T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_{-r}^0 = \Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r} + O_p(T^{-1/2}). \tag{A21}$$

Note that $\overline{\Sigma}_e$ is positive definite by Assumption B. This means that the rank of $\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r}$ is determined by the rank of Λ_{-r}. Since this matrix has rank $m-r$, we have $\text{rk}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r}) = m-r$. Regardless, by adding the results,

$$T^{-1}\overline{e}^{0\prime}\overline{e}^0 = \begin{bmatrix} 0_{r\times r} & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & \Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r} \end{bmatrix} + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{A22}$$

Hence, letting

$$S_{F^0} = \begin{bmatrix} T^{-1}F'F & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_{-r}^0 \end{bmatrix},$$

we have

$$\|T^{-1}\widehat{F}^{0\prime}\widehat{F}^{0} - S_{F^0}\| = O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{A23}$$

But by Assumption B, we also have

$$T^{-1}d^{0\prime}W = T^{-1}\begin{bmatrix} \overline{e}_r^{0\prime}W \\ \overline{e}_{-r}^{0\prime}W \end{bmatrix} = O_p(T^{-1/2}),$$

and so, by adding the results,

$$\begin{aligned} T^{-1}\widehat{z}^{0\prime}\widehat{z}^{0} &= T^{-1}\begin{bmatrix} \widehat{F}^{0\prime}\widehat{F}^{0} & \widehat{F}^{0\prime}W \\ W'\widehat{F}^{0} & W'W \end{bmatrix} \\ &= \begin{bmatrix} S_{F^0} & T^{-1}(F^0)'W \\ T^{-1}W'F^0 & T^{-1}W'W \end{bmatrix} + \begin{bmatrix} T^{-1}\widehat{F}^{0\prime}\widehat{F}^{0} - S_{F^0} & T^{-1}d^{0\prime}W \\ T^{-1}W'd^0 & 0_{n\times n} \end{bmatrix} \\ &= S_{g^0} + \begin{bmatrix} T^{-1}\widehat{F}^{0\prime}\widehat{F}^{0} - S_{F^0} & T^{-1}d^{0\prime}W \\ T^{-1}W'd^0 & 0_{n\times n} \end{bmatrix} \\ &= S_{g^0} + O_p(N^{-1/2}) + O_p(T^{-1/2}). \end{aligned} \tag{A24}$$

where

$$S_{g^0} = \begin{bmatrix} S_{F^0} & T^{-1}(F^0)'W \\ T^{-1}W'F^0 & T^{-1}W'W \end{bmatrix} = \begin{bmatrix} T^{-1}F'F & 0_{r\times(m-r)} & T^{-1}F'W \\ 0_{(m-r)\times r} & T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_{-r}^{0} & 0_{(m-r)\times n} \\ T^{-1}W'F & 0_{n\times(m-r)} & T^{-1}W'W \end{bmatrix},$$

an $(m+n)\times(m+n)$ matrix.

Let us now consider $\|(T^{-1}\widehat{z}^{0\prime}\widehat{z}^{0})^{+} - S_{g^0}^{+}\|$. We begin by noting how $\text{rk}(T^{-1}F'F) \to_{a.s.} r$, $\text{rk}(T^{-1}W'W) \to_{a.s.} n$ and $\text{rk}(T^{-1}\overline{e}_{-r}^{0\prime}\overline{e}_{-r}^{0}) \to_{a.s.} m-r$ as $T\to\infty$, which in turn implies $\text{rk}\, S_{g^0} \to_{a.s.} m+n$. But we also have $\text{rk}(T^{-1}\widehat{z}^{0\prime}\widehat{z}^{0}) \to_{a.s.} m+n$ as $N, T\to\infty$, and so we obtain

$$\text{rk}(T^{-1}\widehat{z}^{0\prime}\widehat{z}^{0}) \to_{a.s.} \text{rk}\, S_{g^0}. \tag{A25}$$

Therefore, according to Andrews (1987),

$$\|(T^{-1}\widehat{z}^{0\prime}\widehat{z}^{0})^{+} - S_{g^0}^{-1}\| = O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{A26}$$

Let

$$D_1 = T^{-1}F'F - T^{-1}F'W(T^{-1}W'W)^{-1}T^{-1}W'F = T^{-1}F'M_WF,$$

$$D_2 = (T^{-1}W'W)^{-1} + (T^{-1}W'W)^{-1}T^{-1}W'FD_1^{-1}T^{-1}F'W(T^{-1}W'W)^{-1},$$

with $M_W = I_{T-h} - W(W'W)^{-1}W'$. Then, according to Abadir and Magnus (2005, Exercise 5.19), $S_{g^0}^{-1}$ has the following structure:

$$S_{g^0}^{-1} = \begin{bmatrix} D_1^{-1} & 0_{r\times(m-r)} & -D_1^{-1}T^{-1}F'W(T^{-1}W'W)^{-1} \\ 0_{(m-r)\times r} & (T^{-1}\bar{e}_{-r}^{0\prime}\bar{e}_{-r}^{0})^{-1} & 0_{(m-r)\times n} \\ -(T^{-1}W'W)^{-1}T^{-1}W'FD_1^{-1} & 0_{n\times(m-r)} & D_2 \end{bmatrix}.$$

By inserting the above results into (A9), we obtain

$$\begin{aligned} & Q_N^{-1}\sqrt{T}(\widehat{\delta} - \delta^0) \\ & = (T^{-1}\widehat{z}^{0\prime}\widehat{z}^0)^{+}T^{-1/2}\widehat{z}^{0\prime}(\varepsilon - \bar{e}_r^0\alpha) \\ & = S_{g^0}^{-1}\left(T^{-1/2}g^{0\prime}\varepsilon - \sqrt{T}\begin{bmatrix} 0_{r\times 1} \\ T^{-1}\bar{e}_{-r}^{0\prime}\bar{e}_r^0\alpha \\ 0_{n\times 1}\end{bmatrix}\right) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\ & + O_p(\sqrt{T}N^{-1}) \\ & = S_{g^0}^{-1}T^{-1/2}g^{0\prime}\varepsilon - \sqrt{T}N^{-1/2}\begin{bmatrix} 0_{r\times 1} \\ (T^{-1}\bar{e}_{-r}^{0\prime}\bar{e}_{-r}^{0})^{-1}\sqrt{N}T^{-1}\bar{e}_{-r}^{0\prime}\bar{e}_r^0\alpha \\ 0_{n\times 1}\end{bmatrix} \\ & + O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{T}N^{-1}). \end{aligned} \tag{A27}$$

Consider the second term on the right. Here, $\bar{e}_r^0 = \bar{e}\Lambda_r$ with $\Lambda_r = [\bar{\lambda}_r^{-1\prime}, 0_{r\times(m-r)}]$. By using this, (A20), $\bar{e}_{-r}^0 = \sqrt{N}\bar{e}\Lambda_{-r}$, and Assumption B,

$$\sqrt{N}T^{-1}\bar{e}_r^{0\prime}\bar{e}_{-r}^0 = \Lambda_r' NT^{-1}\bar{e}'\bar{e}\Lambda_{-r} = \Lambda_r'\overline{\Sigma}_e\Lambda_{-r} + O_p(T^{-1/2}). \tag{A28}$$

But we also have $T^{-1}\bar{e}_{-r}^{0\prime}\bar{e}_{-r}^0 = \Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r} + O_p(T^{-1/2})$, and so

$$\begin{aligned} Q_N^{-1}\sqrt{T}(\widehat{\delta} - Q_N\delta^0) = {} & S_{g^0}^{-1}T^{-1/2}g^{0\prime}\varepsilon - \sqrt{T}N^{-1/2}B + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\ & + O_p(\sqrt{T}N^{-1}). \end{aligned} \tag{A29}$$

where

$$B = \begin{bmatrix} 0_{r\times 1} \\ (\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha \\ 0_{n\times 1}\end{bmatrix}.$$

As for the asymptotic distribution of $S_{g^0}^{-1}T^{-1/2}g^{0\prime}\varepsilon$, note that under Assumptions C and D,

$$\begin{aligned}
\Sigma_{g^0\varepsilon} &= \lim_{N,T\to\infty} T^{-1}E(g^{0\prime}\varepsilon\varepsilon' g^0) = \lim_{N,T\to\infty}\frac{1}{T}\sum_{t=1}^{T-h}E(\varepsilon_{t+h}^2 g_t^0 g_t^{0\prime}) \\
&= \lim_{N,T\to\infty}\frac{1}{T}\sum_{t=1}^{T-h}\begin{bmatrix} E(\varepsilon_{t+h}^2 F_t F_t') & 0_{r\times(m-r)} & E(\varepsilon_{t+h}^2 F_t W_t') \\ 0_{(m-r)\times r} & E(\varepsilon_{t+h}^2 \bar{e}_{-r,t}^0 \bar{e}_{-r,t}^{0\prime}) & 0_{(m-r)\times n} \\ E(\varepsilon_{t+h}^2 W_t F_t') & 0_{n\times(m-r)} & E(\varepsilon_{t+h}^2 W_t W_t') \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_{F\varepsilon} & 0_{r\times(m-r)} & \Sigma_{FW\varepsilon} \\ 0_{(m-r)\times r} & \Sigma_{\bar{e}_{-r}^0\varepsilon} & 0_{(m-r)\times n} \\ \Sigma_{FW\varepsilon}' & 0_{n\times(m-r)} & \Sigma_{W\varepsilon} \end{bmatrix}.
\end{aligned}$$

We have $\lim_{N\to\infty} NE(\bar{e}_t\bar{e}_t') = \Sigma_{e,t}$ (Assumption B). Hence, letting $E(\varepsilon_t^2) = \sigma_{\varepsilon,t}^2$, we can show that $\lim_{N\to\infty} E(\varepsilon_{t+h}^2 \bar{e}_{-r,t}^0 \bar{e}_{-r,t}^{0\prime}) = \sigma_{\varepsilon,t+h}^2 \Lambda_{-r}' \Sigma_{e,t}\Lambda_{-r}$, which in turn implies that $\Sigma_{\bar{e}_{-r}^0\varepsilon}$ can be written as $\Sigma_{\bar{e}_{-r}^0\varepsilon} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T-h}\sigma_{\varepsilon,t+h}^2\Lambda_{-r}'\Sigma_{e,t}\Lambda_{-r} = \Lambda_{-r}'\Sigma_{\bar{e}\varepsilon}\Lambda_{-r}$ with an obvious definition of $\Sigma_{\bar{e}\varepsilon}$. Moreover,

$$S_{g^0}^{-1}\Sigma_{g^0\varepsilon}S_{g^0}^{-1} \to_p \Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}, \tag{A30}$$

as N, $T \to \infty$, with

$$\Sigma_{g^0} = \text{plim}_{N,T\to\infty} S_{g^0} = \begin{bmatrix} \Sigma_F & 0_{r\times(m-r)} & \Sigma_{FW} \\ 0_{(m-r)\times r} & \Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r} & 0_{(m-r)\times n} \\ \Sigma_{FW}' & 0_{n\times(m-r)} & \Sigma_W \end{bmatrix}.$$

By using this, and Assumption D, we can show that

$$S_{g^0}^{-1}T^{-1/2}g^{0\prime}\varepsilon \to_d N(0_{(m+n)\times 1}, \Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}) \tag{A31}$$

as N, $T \to \infty$. Hence, provided that $\sqrt{T}/N \to 0$,

$$\begin{aligned}
&Q_N^{-1}\sqrt{T}(\widehat{\delta} - Q_N\delta^0) + \sqrt{T}N^{-1/2}B \\
&= S_{g^0}^{-1}T^{-1/2}g^{0\prime}\varepsilon + O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{T}N^{-1}) \\
&\to_d N(0_{(m+n)\times 1}, \Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}),
\end{aligned} \tag{A32}$$

as required for the proof under $r < m$.

The proof under $r = m$ is analogous, and is obtained by setting $\bar{e} = \bar{e}_r$, $\bar{\lambda} = \bar{\lambda}_r$, $\Lambda = \bar{\lambda}_r^{-1}$ and $D_N = I_r$. ■

Proof of Theorem 2.

We provide the proof for the case when $r < m$. As in Proof of Theorem 1, the proof for $r = m$ follows by simple manipulations of that of $r < m$. We begin by noting how

$$y_{T+h|T} = \delta^{0\prime}\widehat{z}_T^0 - [\alpha', 0_{1\times(m-r)}]d_T^0 = \delta^{0\prime}\widehat{z}_T^0 - \alpha'\bar{e}_{r,T}^0,$$

where $\bar{e}_{r,t}^0$ is the t-th row of $\bar{e}_r^0$. We similarly have

$$\widehat{y}_{T+h|T} = \widehat{\alpha}'\widehat{F}_T + \widehat{\beta}'W_T = \widehat{\alpha}'(D_N\Lambda')^+D_N\Lambda'\widehat{F}_T + \widehat{\beta}'W_T = [\widehat{\alpha}'(D_N\Lambda')^+, \widehat{\beta}']\widehat{z}_T^0 = \widehat{\delta}'Q_N^{-1\prime}\widehat{z}_T^0,$$

where $\widehat{\delta} = [\widehat{\alpha}', \widehat{\beta}']'$, where $\widehat{\alpha}$ is $m \times 1$ and $\widehat{\beta}$ is $n \times 1$. Here

$$\widehat{z}_t^0 = \begin{bmatrix} \widehat{F}_t^0 \\ W_t \end{bmatrix} = \begin{bmatrix} F_t + \bar{e}_{r,t}^0 \\ \bar{e}_{-r,t}^0 \\ W_t \end{bmatrix} = \begin{bmatrix} F_t \\ \bar{e}_{-r,t}^0 \\ W_t \end{bmatrix} + O_p(N^{-1/2}) = g_t^0 + O_p(N^{-1/2}) \tag{A33}$$

uniformly in t, where $g_t^0 = [F_t', \bar{e}_{-r,t}^{0\prime}, W_t']'$, implying

$$\begin{aligned}
&\widehat{y}_{T+h|T} - y_{T+h|T} \\
&= (Q_N^{-1}\widehat{\delta} - \delta^0)'\widehat{z}_T^0 + \alpha'\bar{e}_{r,T}^0 \\
&= \sqrt{T}(\widehat{\delta} - Q_N\delta^0)'Q_N^{-1\prime}T^{-1/2}\widehat{z}_T^0 + N^{-1/2}\alpha'\sqrt{N}\bar{e}_{r,T}^0 \\
&= \sqrt{T}(\widehat{\delta} - Q_N\delta^0)'Q_N^{-1\prime}T^{-1/2}g_T^0 + N^{-1/2}\alpha'\sqrt{N}\bar{e}_{r,T}^0 + O_p((NT)^{-1/2}) \\
&= [\sqrt{T}(\widehat{\delta} - Q_N\delta^0)'Q_N^{-1\prime} + \sqrt{T}N^{-1/2}B']T^{-1/2}g_T^0 - N^{-1/2}B'g_T^0 + N^{-1/2}\alpha'\sqrt{N}\bar{e}_{r,T}^0 \\
&+ O_p((NT)^{-1/2}).
\end{aligned} \tag{A34}$$

By the definitions of $\bar{e}_{r,t}^0$, $\bar{e}_{-r,t}^0$ and B, letting $\Phi^0 = [I_m - \overline{\Sigma}_e\Lambda_{-r}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}']'\Lambda_r\alpha$,

$$\begin{aligned}
&-N^{-1/2}B'g_T^0 + N^{-1/2}\alpha'\sqrt{N}\bar{e}_{r,T}^0 \\
&= -N^{-1/2}\alpha'\Lambda_r'\overline{\Sigma}_e\Lambda_{-r}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\bar{e}_{-r,T}^0 + N^{-1/2}\alpha'\sqrt{N}\bar{e}_{r,T}^0 \\
&= N^{-1/2}\alpha'[\sqrt{N}\bar{e}_{r,T}^0 - \Lambda_r'\overline{\Sigma}_e\Lambda_{-r}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\bar{e}_{-r,T}^0] \\
&= N^{-1/2}\alpha'\Lambda_r'[I_m - \overline{\Sigma}_e\Lambda_{-r}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}']\sqrt{N}\bar{e}_T \\
&= N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T,
\end{aligned}$$

which can be substituted into the expression for $\widehat{y}_{T+h|T} - y_{T+h|T}$. By doing this and using the result in (A29) we obtain

$$\begin{aligned}
\widehat{y}_{T+h|T} - y_{T+h|T} &= T^{-1/2}(T^{-1/2}\varepsilon'g^0S_{g^0}^{-1}g_T^0) + N^{-1/2}(\Phi^{0\prime}\sqrt{N}\bar{e}_T) + O_p((NT)^{-1/2}) \\
&+ O_p(T^{-1}) + O_p(N^{-1}).
\end{aligned} \tag{A35}$$

Because $T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1}$ and $\sqrt{N}\bar{e}_T$ are both mean zero asymptotically normal, the linear combination $T^{-1/2}(T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0) + N^{-1/2}(\Phi^{0\prime}\sqrt{N}\bar{e}_T)$ is mean zero and asymptotically normal too. The asymptotic variance given z_T is given by

$$\begin{aligned}
&\text{Avar}(\widehat{y}_{T+h|T} - y_{T+h|T}|z_T) \\
&= \text{Avar}[(T^{-1/2}T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0 + N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T)|z_T] \\
&= \text{Avar}(T^{-1/2}T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0|z_T) + \text{Avar}(N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T|z_T) \\
&+ 2\text{Acov}(T^{-1/2}T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0, N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T|z_T).
\end{aligned} \tag{A36}$$

The first term on the right is

$$\begin{aligned}
\text{Avar}(T^{-1/2}T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0|z_T) &= \lim_{N,T\to\infty} T^{-1}E(g_T^{0\prime} S_{g^0}^{-1} T^{-1} g^{0\prime}\varepsilon\varepsilon' g^0 S_{g^0}^{-1} g_T^0|z_T) \\
&= \lim_{N,T\to\infty} T^{-1}E[g_T^{0\prime}\Sigma_{g^0}^{-1} E(T^{-1}g^{0\prime}\varepsilon\varepsilon' g^0|g_T^0)\Sigma_{g^0}^{-1} g_T^0|z_T] \\
&= \lim_{N,T\to\infty} T^{-1}E(g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1} g_T^0|z_T).
\end{aligned} \tag{A37}$$

Note that

$$E(g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1} g_T^0|z_T) = \text{tr}[\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}E(g_T^0 g_T^{0\prime}|z_T)], \tag{A38}$$

where, by Assumption D, which ensures the independence of $(F_T', W_T')'$ and $\bar{e}_{-r,T}^0$,

$$\begin{aligned}
E(g_T^0 g_T^{0\prime}|z_T) &= z_T^0 z_T^{0\prime} + \text{diag}(0_{r\times r}, E(\bar{e}_{-r,T}^0\bar{e}_{-r,T}^{0\prime}), 0_{n\times n}) \\
&= z_T^0 z_T^{0\prime} + \text{diag}(0_{r\times r}, \Lambda_{-r}'\Sigma_{e,T}\Lambda_{-r}, 0_{n\times n}).
\end{aligned}$$

Here we use (A28) to obtain the last equality.

Because of the structures of $z_T^0 z_T^{0\prime}$ and $\text{diag}(0_{r\times r}, \Lambda_{-r}'\Sigma_{e,T}\Lambda_{-r}, 0_{n\times n})$ on the one hand, and the structure of $\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}$ on the other hand,

$$\begin{aligned}
&E(g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1} g_T^0|z_T) \\
&= z_T^{0\prime}\Sigma_{z^0}^{-1}\Sigma_{z^0\varepsilon}\Sigma_{z^0}^{-1}z_T^0 + \text{tr}[(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Sigma_{\bar{e}_{-r}^0\varepsilon}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\Sigma_{e,T}\Lambda_{-r}] \\
&= z_T'\Sigma_z^{-1}\Sigma_{z\varepsilon}\Sigma_z^{-1}z_T + \text{tr}[(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\Sigma_{\bar{e}\varepsilon}\Lambda_{-r}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\Sigma_{e,T}\Lambda_{-r}] \\
&= z_T'\Sigma_z^{-1}\Sigma_{z\varepsilon}\Sigma_z^{-1}z_T + \text{tr}(P_{\Lambda_{-r}}\Sigma_{\bar{e}\varepsilon}P_{\Lambda_{-r}}\Sigma_{e,T}) = \phi,
\end{aligned} \tag{A39}$$

where $P_{\Lambda_{-r}} = \Lambda_{-r}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'$.

For $\text{Avar}(N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T|z_T)$, we make use of the fact that $\lim_{N\to\infty} NE(\bar{e}_t\bar{e}_t') = \Sigma_{e,t}$ (Assumption B). By using this and $\Sigma_e = \lim_{T\to\infty}\Sigma_{e,T}$, we can show that

$$\begin{aligned}\text{Avar}(N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T|z_T) &= \lim_{N,T\to\infty} N^{-1}\Phi^{0\prime}NE(\bar{e}_T\bar{e}_T')\Phi^0 = \lim_{N,T\to\infty} N^{-1}\Phi^{0\prime}\Sigma_{e,T}\Phi^0. \\ &= \lim_{N\to\infty} N^{-1}\Phi^{0\prime}\Sigma_e\Phi^0. \end{aligned} \tag{A40}$$

Moreover, since by Assumption D, ε is uncorrelated with both g and $\bar{e}$, we can show that

$$\begin{aligned}&\text{Acov}(T^{-1/2}T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0, N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T|z_T) \\ &= \lim_{N,T\to\infty}(NT)^{-1/2}E(T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0\sqrt{N}\bar{e}_T'\Phi^0|z_T) = 0. \end{aligned} \tag{A41}$$

Direct insertion into the expression for $\text{Avar}(\widehat{y}_{T+h|T} - y_{T+h|T}|z_T)$ now yields

$$\begin{aligned}&\text{Avar}(\widehat{y}_{T+h|T} - y_{T+h|T}|z_T) \\ &= \text{Avar}(T^{-1/2}T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0|z_T) + \text{Avar}(N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T|z_T) \\ &+ 2\text{Acov}(T^{-1/2}T^{-1/2}\varepsilon' g^0 S_{g^0}^{-1} g_T^0, N^{-1/2}\Phi^{0\prime}\sqrt{N}\bar{e}_T|z_T) \\ &= \lim_{N,T\to\infty}[T^{-1}E(g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0|z_T) + N^{-1}\Phi^{0\prime}\Sigma_e\Phi^0] \\ &= \lim_{N,T\to\infty}(T^{-1}\phi + N^{-1}\Phi^{0\prime}\Sigma_e\Phi^0), \end{aligned} \tag{A42}$$

where the first term is $O(T^{-1})$, while the second is $O(N^{-1})$. Hence, by putting everything together,

$$\frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{T^{-1}\phi + N^{-1}\Phi^{0\prime}\Sigma_e\Phi^0}} \to_d N(0,1), \tag{A43}$$

as N, $T\to\infty$ with $\sqrt{T}/N \to 0$. This completes the proof of the theorem. ■

Proof of Theorem 3.

By definition,

$$\begin{aligned}\widehat{\Sigma}_z^+\widehat{\Sigma}_{z\varepsilon}\widehat{\Sigma}_z^+ &= (T^{-1}\widehat{z}'\widehat{z})^+\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t\widehat{z}_t'(T^{-1}\widehat{z}'\widehat{z})^+ \\ &= Q_N(T^{-1}Q_N'\widehat{z}'\widehat{z}Q_N)^+\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2 Q_N'\widehat{z}_t\widehat{z}_t'Q_N(T^{-1}Q_N'\widehat{z}'\widehat{z}Q_N)^+Q_N' \\ &= Q_N(T^{-1}\widehat{z}^{0\prime}\widehat{z}^0)^+\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t^0\widehat{z}_t^{0\prime}(T^{-1}\widehat{z}^{0\prime}\widehat{z}^0)^+Q_N'. \end{aligned} \tag{A44}$$

We begin by noting that $\widehat{\varepsilon}_{t+h} = y_{t+h} - \widehat{\delta}'\widehat{z}_t = \varepsilon_{t+h} - (\widehat{\delta} - Q_N\delta^0)'Q_N^{-1\prime}\widehat{z}_t^0 - \alpha'\overline{e}_{r,t}^0$ and $\widehat{\varepsilon}_{t+h}^2 - \varepsilon_{t+h}^2 = \widehat{\varepsilon}_{t+h}(\widehat{\varepsilon}_{t+h} - \varepsilon_{t+h}) + \varepsilon_{t+h}(\widehat{\varepsilon}_{t+h} - \varepsilon_{t+h})$, giving

$$
\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t^0\widehat{z}_t^{0\prime} - \frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^2\widehat{z}_t^0\widehat{z}_t^{0\prime}\right\| = \left\|\frac{1}{T}\sum_{t=1}^{T-h}(\widehat{\varepsilon}_{t+h}^2 - \varepsilon_{t+h}^2)\widehat{z}_t^0\widehat{z}_t^{0\prime}\right\| \\
&\leq \left\|\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}\widehat{z}_t^{0\prime}Q_N^{-1}(\widehat{\delta} - Q_N\delta^0)\widehat{z}_t^0\widehat{z}_t^{0\prime}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T-h}(\widehat{\delta} - Q_N\delta^0)'Q_N^{-1\prime}\widehat{z}_t^0\overline{e}_{r,t}^{0\prime}\alpha\widehat{z}_t^0\widehat{z}_t^{0\prime}\right\| \\
&+ \left\|\frac{1}{T}\sum_{t=1}^{T-h}\alpha'\overline{e}_{r,t}^0\overline{e}_{r,t}^{0\prime}\alpha\widehat{z}_t^0\widehat{z}_t^{0\prime}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}\widehat{z}_t^{0\prime}Q_N^{-1}(\widehat{\delta} - Q_N\delta^0)\widehat{z}_t^0\widehat{z}_t^{0\prime}\right\| \\
&+ 2\left\|\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}\overline{e}_{r,t}^{0\prime}\alpha\widehat{z}_t^0\widehat{z}_t^{0\prime}\right\| \\
&\leq \|Q_N^{-1}(\widehat{\delta} - Q_N\delta^0)\|\frac{1}{T}\sum_{t=1}^{T-h}\|\widehat{\varepsilon}_{t+h}\|\|\widehat{z}_t^0\|^3 \\
&+ \|Q_N^{-1}(\widehat{\delta} - Q_N\delta^0)\|\|\alpha\|\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\overline{e}_{r,t}^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\widehat{z}_t^0\|^6\right)^{1/2} \\
&+ \|\alpha\|^2\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\overline{e}_{r,t}^0\|^4\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\widehat{z}_t^0\|^4\right)^{1/2} + \|Q_N^{-1}(\widehat{\delta} - Q_N\delta^0)\|\frac{1}{T}\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|\|\widehat{z}_t^0\|^3 \\
&+ 2\|\alpha\|\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\overline{e}_{r,t}^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|^2\|\widehat{z}_t^0\|^6\right)^{1/2} \\
&= O_p(T^{-1/2}) + O_p(N^{-1/2}).
\end{aligned}
$$

Moreover, $\widehat{z}_t^0\widehat{z}_t^{0\prime} - g_t^0 g_t^{0\prime} = (\widehat{z}_t^0 - g_t^0)\widehat{z}_t^{0\prime} - g_t^0(g_t^0 - \widehat{z}_t^0)'$, where $\widehat{z}_t^0 - g_t^0 = [\overline{e}_{r,t}^{0\prime}, 0'_{(n+m-r)\times 1}]'$ and $\|\overline{e}_{r,t}^0\| = O_p(N^{-1/2})$ uniformly in t. This implies

$$
\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^2\widehat{z}_t^0\widehat{z}_t^{0\prime} - \frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^2 z_t^0 z_t^{0\prime}\right\| \\
&= \left\|\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^2(\widehat{z}_t^0\widehat{z}_t^{0\prime} - z_t^0 z_t^{0\prime})\right\| \\
&= \left\|\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^2(\widehat{z}_t^0 - g_t^0)\widehat{z}_t^{0\prime}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^2 g_t^0(g_t^0 - \widehat{z}_t^0)'\right\| \\
&\leq \left(\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^4\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\widehat{z}_t^0 - g_t^0\|^4\right)^{1/4}\left[\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\widehat{z}_t^0\|^4\right)^{1/4} + \left(\frac{1}{T}\sum_{t=1}^{T-h}\|g_t^0\|^4\right)^{1/4}\right] \\
&= O_p(N^{-1/2}),
\end{aligned}
$$

and so

$$\begin{aligned}
&\left\| \frac{1}{T}\sum_{t=1}^{T-h} \widehat{\varepsilon}^2_{t+h}\widehat{z}^0_t\widehat{z}^{0\prime}_t - \frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h} g^0_t g^{0\prime}_t \right\| \\
&\leq \left\| \frac{1}{T}\sum_{t=1}^{T-h} \widehat{\varepsilon}^2_{t+h}\widehat{z}^0_t\widehat{z}^{0\prime}_t - \frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h}\widehat{z}^0_t\widehat{z}^{0\prime}_t \right\| + \left\| \frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h}\widehat{z}^0_t\widehat{z}^{0\prime}_t - \frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h} g^0_t g^{0\prime}_t \right\| \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}). \qquad \text{(A45)}
\end{aligned}$$

Hence, since $\|(T^{-1}\widehat{z}^{0\prime}\widehat{z}^0)^+ - S^{-1}_{g^0}\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$ (Proof of Theorem 1), we can show that

$$\begin{aligned}
&Q_N(T^{-1}\widehat{z}^{0\prime}\widehat{z}^0)^+ \frac{1}{T}\sum_{t=1}^{T-h} \widehat{\varepsilon}^2_{t+h}\widehat{z}^0_t\widehat{z}^{0\prime}_t (T^{-1}\widehat{z}^{0\prime}\widehat{z}^0)^+ Q_N' \\
&= Q_N S^{-1}_{g^0}\frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h} g^0_t g^{0\prime}_t S^{-1}_{g^0} Q_N' + O_p(N^{-1/2}) + O_p(T^{-1/2}), \qquad \text{(A46)}
\end{aligned}$$

which, together with $\|\widehat{z}^0_T - g^0_T\| = O_p(N^{-1/2})$ (Proof of Theorem 1), gives the following result for $\widehat{z}_T'\widehat{\Sigma}^+_z\widehat{\Sigma}_{z\varepsilon}\widehat{\Sigma}^+_z\widehat{z}_T$:

$$\begin{aligned}
\widehat{\phi} &= \widehat{z}_T'\widehat{\Sigma}^+_z\widehat{\Sigma}_{z\varepsilon}\widehat{\Sigma}^+_z\widehat{z}_T \\
&= \widehat{z}_T' Q_N S^{-1}_{g^0}\frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h} g^0_t g^{0\prime}_t S^{-1}_{g^0} Q_N'\widehat{z}_T + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= (\widehat{z}^0_T)' S^{-1}_{g^0}\frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h} g^0_t g^{0\prime}_t S^{-1}_{g^0}\widehat{z}^0_T + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= g^{0\prime}_T S^{-1}_{g^0}\frac{1}{T}\sum_{t=1}^{T-h} \varepsilon^2_{t+h} g^0_t g^{0\prime}_t S^{-1}_{g^0} g^0_T + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= g^{0\prime}_T \Sigma^{-1}_{g^0}\Sigma_{g^0\varepsilon}\Sigma^{-1}_{g^0} g^0_T + O_p(N^{-1/2}) + O_p(T^{-1/2}). \qquad \text{(A47)}
\end{aligned}$$

In analogy with the decomposition of ϕ in Proof of Theorem 2, we can show that

$$\begin{aligned}
g^{0\prime}_T \Sigma^{-1}_{g^0}\Sigma_{g^0\varepsilon}\Sigma^{-1}_{g^0} g^0_T &= z_T'\Sigma^{-1}_z\Sigma_{z\varepsilon}\Sigma^{-1}_z z_T \\
&\quad + \text{tr}[(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Sigma_{\overline{e}^0_{-r}\varepsilon}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'N\overline{e}_T\overline{e}_T'\Lambda_{-r}] \\
&= z_T'\Sigma^{-1}_z\Sigma_{z\varepsilon}\Sigma^{-1}_z z_T + \text{tr}(P_{\Lambda_{-r}}\Sigma_{\overline{e}\varepsilon}P_{\Lambda_{-r}}N\overline{e}_T\overline{e}_T'). \qquad \text{(A48)}
\end{aligned}$$

Let $Z_m \sim N(0_{m\times 1}, I_m)$. Note how $\sqrt{N}\Sigma^{-1/2}_{e,T}\overline{e}_T \to_d Z_m$ as $N \to \infty$, where $\Sigma_{e,t} = \Sigma^{1/2}_{e,t}\Sigma^{1/2\prime}_{e,t}$ with $\Sigma^{1/2}_{e,t}$ being the lower triangular Choleski factor. It follows that

$$\begin{aligned}
\text{tr}(P_{\Lambda_{-r}}\Sigma_{\overline{e}\varepsilon}P_{\Lambda_{-r}}N\overline{e}_T\overline{e}_T') &= (\sqrt{N}\Sigma^{-1/2}_{e,T}\overline{e}_T)'\Sigma^{1/2}_{e,T}P_{\Lambda_{-r}}\Sigma_{\overline{e}\varepsilon}P_{\Lambda_{-r}}\Sigma^{1/2}_{e,T}(\sqrt{N}\Sigma^{-1/2}_{e,T}\overline{e}_T) \\
&\to_d Z_m'\Sigma^{1/2}_e P_{\Lambda_{-r}}\Sigma_{\overline{e}\varepsilon}P_{\Lambda_{-r}}\Sigma^{1/2}_e Z_m. \qquad \text{(A49)}
\end{aligned}$$

Next, consider $\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}$. The formula for $\widehat{\Sigma}_e$ depend on whether or not $e_{i,t}$ is correlated across the cross-section. Here we assume that the correlation is absent. The proof for the correlated case is analogous. We begin by noting how $\|Q_N\| = O_p(\sqrt{N})$, from which it follows that

$$\widehat{\delta} - Q_N\delta^0 + N^{-1/2}Q_N B = Q_N T^{-1/2} S_{g^0}^{-1} T^{-1/2} g^{0\prime}\varepsilon + O_p(T^{-1/2}) + O_p(\sqrt{N}T^{-1}) + O_p(N^{-1/2}). \tag{A50}$$

In terms of the notation introduced in Proof of Theorem 1,

$$\begin{aligned}
& -(T^{-1}W'W)^{-1}T^{-1}W'FD_1^{-1}T^{-1/2}F'\varepsilon + D_2 T^{-1/2}W'\varepsilon \\
&= -(T^{-1}W'W)^{-1}T^{-1}W'F(T^{-1}F'M_W F)^{-1}T^{-1/2}F'\varepsilon \\
&+ (T^{-1}W'W)^{-1}T^{-1/2}W'\varepsilon + (T^{-1}W'W)^{-1}T^{-1}W'FD_1^{-1}T^{-1}F'W(T^{-1}W'W)^{-1}T^{-1/2}W'\varepsilon \\
&= (T^{-1}W'W)^{-1}T^{-1/2}W'[I_{T-h} + T^{-1}F(T^{-1}F'M_W F)^{-1}T^{-1}F'W(T^{-1}W'W)^{-1}W' \\
&- T^{-1}F(T^{-1}F'M_W F)^{-1}F']\varepsilon \\
&= (T^{-1}W'W)^{-1}T^{-1/2}W'[I_{T-h} + F(F'M_W F)^{-1}F'M_W]\varepsilon.
\end{aligned}$$

This implies

$$\begin{aligned}
S_{g^0}^{-1}T^{-1/2}g^{0\prime}\varepsilon &= \begin{bmatrix} D_1^{-1}T^{-1/2}F'\varepsilon - D_1^{-1}T^{-1}F'W(T^{-1}W'W)^{-1}T^{-1/2}W'\varepsilon \\ (T^{-1}\overline{e}^{0\prime}_{-r}\overline{e}^{0}_{-r})^{-1}T^{-1/2}\overline{e}^{0\prime}_{-r}\varepsilon \\ -(T^{-1}W'W)^{-1}T^{-1}W'FD_1^{-1}T^{-1/2}F'\varepsilon + D_2 T^{-1/2}W'\varepsilon \end{bmatrix} \\
&= \begin{bmatrix} (T^{-1}F'M_W F)^{-1}T^{-1/2}F'M_W\varepsilon \\ (T^{-1}\overline{e}^{0\prime}_{-r}\overline{e}^{0}_{-r})^{-1}T^{-1/2}\overline{e}^{0\prime}_{-r}\varepsilon \\ (T^{-1}W'W)^{-1}T^{-1/2}W'[I_{T-h} - F(F'M_W F)^{-1}F'M_W]\varepsilon \end{bmatrix},
\end{aligned} \tag{A51}$$

and so

$$\begin{aligned}
& Q_N T^{-1/2}S_{g^0}^{-1}T^{-1/2}g^{0\prime}\varepsilon \\
&= \begin{bmatrix} \overline{\lambda}_r^{-1}T^{-1/2}[(T^{-1}F'M_W F)^{-1}T^{-1/2}F'M_W\varepsilon - \overline{\lambda}_{-r}(T^{-1}\overline{e}^{0\prime}_{-r}\overline{e}^{0}_{-r})^{-1}T^{-1/2}\overline{e}^{0\prime}_{-r}\varepsilon] \\ \sqrt{N}T^{-1/2}(T^{-1}\overline{e}^{0\prime}_{-r}\overline{e}^{0}_{-r})^{-1}T^{-1/2}\overline{e}^{0\prime}_{-r}\varepsilon \\ T^{-1/2}(T^{-1}W'W)^{-1}T^{-1/2}W'[I_{T-h} - F(F'M_W F)^{-1}F'M_W]\varepsilon \end{bmatrix} \\
&= \begin{bmatrix} 0_{r\times 1} \\ \sqrt{N}T^{-1/2}(T^{-1}\overline{e}^{0\prime}_{-r}\overline{e}^{0}_{-r})^{-1}T^{-1/2}\overline{e}^{0\prime}_{-r}\varepsilon \\ 0_{n\times 1} \end{bmatrix} + O_p(T^{-1/2}).
\end{aligned} \tag{A52}$$

Let

$$\Omega = (\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}\Sigma_{\overline{e}\varepsilon}\Lambda_{-r}(\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}.$$

By using the results provided in Proof of Theorem 1, we can show that

$$(T^{-1}\bar{e}^{0\prime}_{-r}\bar{e}^{0}_{-r})^{-1}T^{-1/2}\bar{e}^{0\prime}_{-r}\varepsilon \to_d \Omega^{1/2}Z_{m-r} \tag{A53}$$

as $N, T \to \infty$. Hence, provided that $N/T \to \tau < \infty$ and $\sqrt{T}/N \to 0$, we can show the following:

$$\widehat{\delta} - Q_N\delta^0 + N^{-1/2}Q_NB \to_d \begin{bmatrix} 0_{r\times 1} \\ \tau\Omega^{1/2}Z_{m-r} \\ 0_{n\times 1} \end{bmatrix}, \tag{A54}$$

where

$$\begin{aligned} N^{-1/2}Q_NB &= \begin{bmatrix} -N^{-1/2}\overline{\lambda}_r^{-1}\overline{\lambda}_{-r}(\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}\overline{\Sigma}_e\Lambda_r\alpha \\ (\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}\overline{\Sigma}_e\Lambda_r\alpha \\ 0_{n\times 1} \end{bmatrix} \\ &= \begin{bmatrix} 0_{r\times 1} \\ (\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}\overline{\Sigma}_e\Lambda_r\alpha \\ 0_{n\times 1} \end{bmatrix} + O_p(N^{-1/2}), \end{aligned}$$

and so

$$\begin{aligned} Q_N\delta^0 - N^{-1/2}Q_NB &= \begin{bmatrix} \overline{\lambda}_r^{-1}\alpha \\ 0_{(m-r)\times 1} \\ \beta \end{bmatrix} - \begin{bmatrix} 0_{r\times 1} \\ (\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}\overline{\Sigma}_e\Lambda_r\alpha \\ 0_{n\times 1} \end{bmatrix} + O_p(N^{-1/2}) \\ &= \begin{bmatrix} \overline{\lambda}_r^{-1}\alpha \\ -(\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}\overline{\Sigma}_e\Lambda_r\alpha \\ \beta \end{bmatrix} + O_p(N^{-1/2}), \end{aligned} \tag{A55}$$

which means that while $\widehat{\beta}$ is consistent for β, $\widehat{\alpha}$ is inconsistent. In fact,

$$\widehat{\alpha} \to_d \begin{bmatrix} \overline{\lambda}_r^{-1}\alpha \\ -(\Lambda'_{-r}\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda'_{-r}\overline{\Sigma}_e\Lambda_r\alpha + \sqrt{\tau}\Omega^{1/2}Z_{m-r} \end{bmatrix} = \Phi$$

as $N, T \to \infty$.

Let us now consider $\widehat{\Sigma}_e$. Note how

$$\begin{aligned} x_i &= F\lambda_i + e_i = F\overline{\lambda}\Lambda D_N(\overline{\lambda}\Lambda D_N)^{+}\lambda_i + e_i = F^0 \begin{bmatrix} I_r \\ 0_{(m-r)\times r} \end{bmatrix} \overline{\lambda}_r^{-1}\overline{\lambda}_r\lambda_i + e_i \\ &= F^0\Lambda_r\overline{\lambda}_r\lambda_i + e_i = \widehat{F}^0\Lambda_r\overline{\lambda}_r\lambda_i - d^0\Lambda_r\overline{\lambda}_r\lambda_i + e_i, \end{aligned}$$

implying

$$\begin{aligned}
\widehat{\lambda}_i &= (\widehat{F}'\widehat{F})^+\widehat{F}x_i \\
&= \Lambda D_N(D_N\Lambda'\widehat{F}'\widehat{F}\Lambda D_N)^+D_N\Lambda'\widehat{F}'x_i \\
&= \Lambda D_N(T^{-1}\widehat{F}^{0\prime}\widehat{F}^0)^+T^{-1}\widehat{F}^{0\prime}x_i \\
&= \Lambda D_N\Lambda_r\overline{\lambda}_r\lambda_i - \Lambda D_N(T^{-1}\widehat{F}^{0\prime}\widehat{F}^0)^+T^{-1}\widehat{F}^{0\prime}d^0\Lambda_r\overline{\lambda}_r\lambda_i + \Lambda D_N(T^{-1}\widehat{F}^{0\prime}\widehat{F}^0)^+T^{-1}\widehat{F}^{0\prime}e_i.
\end{aligned}$$

Here,

$$\begin{aligned}
T^{-1}\widehat{F}^{0\prime}d^0 &= T^{-1}F^{0\prime}d^0 + T^{-1}d^{0\prime}d^0 = T^{-1}F^{0\prime}\overline{e}^0 + T^{-1}\overline{e}^{0\prime}\overline{e}^0 \\
&= T^{-1}\overline{e}^{0\prime}\overline{e}^0 + O_p(T^{-1/2}) \\
&= \begin{bmatrix} 0_{r\times r} & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & T^{-1}\overline{e}^{0\prime}_{-r}\overline{e}^0_{-r} \end{bmatrix} + O_p(N^{-1/2}) + O_p(T^{-1/2}), \\
T^{-1}\widehat{F}^{0\prime}e_i &= T^{-1}F^{0\prime}e_i + T^{-1}\overline{e}^{0\prime}e_i \\
&= T^{-1}\overline{e}^{0\prime}e_i + O_p(T^{-1/2}) \\
&= (NT)^{-1}e_i^{0\prime}e_i + \frac{1}{NT}\sum_{j\neq i}^{N} e_j^{0\prime}e_i + O_p(T^{-1/2}) \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}).
\end{aligned}$$

By using this and the fact that the leading matrix in $T^{-1}\widehat{F}^{0\prime}d^0$ is zero when post-multiplied by Λ_r, we obtain

$$\begin{aligned}
&(\Lambda D_N)^+(\widehat{\lambda}_i - \Lambda D_N\Lambda_r\overline{\lambda}_r\lambda_i) \\
&= -(T^{-1}\widehat{F}^{0\prime}\widehat{F}^0)^+T^{-1}\widehat{F}^{0\prime}d^0\Lambda_r\overline{\lambda}_r\lambda_i + (T^{-1}\widehat{F}^{0\prime}\widehat{F}^0)^+T^{-1}\widehat{F}^{0\prime}e_i \\
&= -(T^{-1}\widehat{F}^{0\prime}\widehat{F}^0)^+\begin{bmatrix} 0_{r\times r} & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & T^{-1}\overline{e}^{0\prime}_{-r}\overline{e}^0_{-r} \end{bmatrix}\Lambda_r\overline{\lambda}_r\lambda_i + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}). \qquad \text{(A56)}
\end{aligned}$$

Let us now consider $\widehat{e}_i$;

$$\begin{aligned}
\widehat{e}_i &= x_i - \widehat{F}\widehat{\lambda}_i \\
&= F\overline{\lambda}\Lambda D_N(\overline{\lambda}\Lambda D_N)^+\lambda_i - \widehat{F}\Lambda D_N(\Lambda D_N)^+\widehat{\lambda}_i + e_i \\
&= F^0\Lambda_r\overline{\lambda}_r\lambda_i - \widehat{F}^0(\Lambda D_N)^+\widehat{\lambda}_i + e_i \\
&= -(\widehat{F}^0 - F^0)\Lambda_r\overline{\lambda}_r\lambda_i - \widehat{F}^0[(\Lambda D_N)^+\widehat{\lambda}_i - \Lambda_r\overline{\lambda}_r\lambda_i] + e_i \\
&= -(\widehat{F}^0 - F^0)\Lambda_r\overline{\lambda}_r\lambda_i - \widehat{F}^0(\Lambda D_N)^+(\widehat{\lambda}_i - \Lambda D_N\Lambda_r\overline{\lambda}_r\lambda_i) + e_i, \qquad \text{(A57)}
\end{aligned}$$

or, in vector notation,

$$\widehat{e}_{i,t} = -\lambda_i'\overline{\lambda}_r'\Lambda_r'(\widehat{F}_t^0 - F_t^0) - (\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i)'[(\Lambda D_N)^+]'\widehat{F}_t^0 + e_{i,t}, \tag{A58}$$

which we can use to obtain

$$\begin{aligned}
&\widehat{e}_{i,t}\widehat{e}_{i,t}' - e_{i,t}e_{i,t}' \\
&= \lambda_i'\overline{\lambda}_r'\Lambda_r'(\widehat{F}_t^0 - F_t^0)(\widehat{F}_t^0 - F_t^0)'\Lambda_r\overline{\lambda}_r\lambda_i + (\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i)'[(\Lambda D_N)^+]'\widehat{F}_t^0(\widehat{F}_t^0 - F_t^0)'\Lambda_r\overline{\lambda}_r\lambda_i \\
&- e_{i,t}(\widehat{F}_t^0 - F_t^0)'\Lambda_r\overline{\lambda}_r\lambda_i + \lambda_i'\overline{\lambda}_r'\Lambda_r'(\widehat{F}_t^0 - F_t^0)\widehat{F}_t^{0\prime}(\Lambda D_N)^+(\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i) \\
&+ (\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i)'[(\Lambda D_N)^+]'\widehat{F}_t^0\widehat{F}_t^{0\prime}(\Lambda D_N)^+(\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i) \\
&- e_{i,t}\widehat{F}_t^{0\prime}(\Lambda D_N)^+(\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i) - \lambda_i'\overline{\lambda}_r'\Lambda_r'(\widehat{F}_t^0 - F_t^0)e_{i,t}' \\
&- (\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i)'[(\Lambda D_N)^+]'\widehat{F}_t^0 e_{i,t}'.
\end{aligned}$$

By using this, and the known orders of $(\Lambda D_N)^+(\widehat{\lambda}_i - \Lambda D_N \Lambda_r \overline{\lambda}_r \lambda_i)$ and $\Lambda_r'(\widehat{F}_t^0 - F^0) = (\overline{\lambda}_r^{-1})'\overline{e}_{r,t}^0$, we can show that

$$\left\| \frac{1}{N}\sum_{i=1}^{N}(\widehat{e}_{i,t}\widehat{e}_{i,t}' - e_{i,t}e_{i,t}') \right\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) \tag{A59}$$

uniformly in t. This implies

$$\begin{aligned}
\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha} &= \widehat{\alpha}'\frac{1}{N}\sum_{i=1}^{N}\widehat{e}_{i,T}\widehat{e}_{i,T}'\widehat{\alpha} = \widehat{\alpha}'\frac{1}{N}\sum_{i=1}^{N}e_{i,T}e_{i,T}'\widehat{\alpha} + o_p(1) = \widehat{\alpha}'\Sigma_e\widehat{\alpha} + o_p(1) \\
&\to_d \Phi'\Sigma_e\Phi
\end{aligned} \tag{A60}$$

as N, $T \to \infty$ with $\sqrt{T}/N \to 0$ and $N/T \to \tau$, as required. ■

Proof of Theorem 4.

From Proof of Theorem 2, we have that

$$\begin{aligned}
\text{Avar}[\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T}|z_T)] &= \lim_{N,T\to\infty}[E(g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0|z_T) + TN^{-1}\Phi^{0\prime}\Sigma_e\Phi^0] \\
&= \lim_{N,T\to\infty} E(g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0|z_T) + O_p(TN^{-1})
\end{aligned} \tag{A61}$$

as N, $T \to \infty$ with $T/N \to 0$. But, according to Proof of Theorem 3, we also have

$$\widehat{\varphi} = g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0 + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{A62}$$

These results hold regardless of whether $r = m$ or $r < m$. This implies

$$\frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{T^{-1}\widehat{\varphi} + N^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}} = \frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{\widehat{\varphi} + TN^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}}$$
$$= \frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0 + TN^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}} + o_p(1). \tag{A63}$$

Consider $TN^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}$. If $r = m$, we know from Corollary 2 that $\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha} \to_p \alpha'\overline{\lambda}^{-1\prime}\Sigma_e\overline{\lambda}^{-1}\alpha$, and hence $TN^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha} = O_p(TN^{-1})$, which is $o_p(1)$ if $T/N \to 0$. If, on the other hand, $r < m$, then we have from Theorem 3 that

$$TN^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha} \to_d \lim_{T\to\infty}(\sqrt{T}N^{-1/2}\Phi)'\Sigma_e\sqrt{T}N^{-1/2}\Phi, \tag{A64}$$

where

$$\sqrt{T}N^{-1/2}\Phi = \begin{bmatrix} \sqrt{T}N^{-1/2}\overline{\lambda}_r^{-1}\alpha \\ -\sqrt{T}N^{-1/2}(\Lambda_{-r}'\overline{\Sigma}_e\Lambda_{-r})^{-1}\Lambda_{-r}'\overline{\Sigma}_e\Lambda_r\alpha + \Omega^{1/2}Z_{m-r} \end{bmatrix} + o_p(1)$$
$$= \begin{bmatrix} 0_{r\times 1} \\ \Omega^{1/2}Z_{m-r} \end{bmatrix} + O_p(\sqrt{T}N^{-1/2}). \tag{A65}$$

Hence,

$$(\sqrt{T}N^{-1/2}\Phi)'\Sigma_e\sqrt{T}N^{-1/2}\Phi = Z_{m-r}'\Omega^{1/2\prime}\Sigma_e\Omega^{1/2}Z_{m-r} + O_p(\sqrt{T}N^{-1/2}), \tag{A66}$$

where the first term on the right is almost surely positive under $r < m$. This implies

$$\frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0 + TN^{-1}\Phi'\Sigma_e\Phi}}$$
$$= \frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0 + Z_{m-r}'\Omega^{1/2\prime}\Sigma_e\Omega^{1/2}Z_{m-r}}} + O_p(\sqrt{T}N^{-1/2}), \tag{A67}$$

with

$$\frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0 + Z_{m-r}'\Omega^{1/2\prime}\Sigma_e\Omega^{1/2}Z_{m-r}}} \le \frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0}} \to_d N(0,1) \tag{A68}$$

almost surely, as $N, T \to \infty$ with $T/N \to 0$. Let us denote by z_α the $(1-\alpha)$-th quantile of the standard normal cumulative distribution function. The above results imply that the asymptotic size of the absolute value of the feasible test statistic when compared to $z_{\alpha/2}$ is given by

$$\lim_{N,T\to\infty} P\left(\left|\frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{T^{-1}\widehat{\phi} + N^{-1}\widehat{\alpha}'\widehat{\Sigma}_e\widehat{\alpha}}}\right| > z_{\alpha/2}\right)$$

$$= \lim_{N,T\to\infty} P\left(\left|\frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0 + Z_{m-r}'\Omega^{1/2\prime}\Sigma_e\Omega^{1/2}Z_{m-r}}}\right| > z_{\alpha/2}\right)$$

$$\leq \lim_{N,T\to\infty} P\left(\left|\frac{\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T})}{\sqrt{g_T^{0\prime}\Sigma_{g^0}^{-1}\Sigma_{g^0\varepsilon}\Sigma_{g^0}^{-1}g_T^0}}\right| > z_{\alpha/2}\right) = \alpha, \tag{A69}$$

which supposes that $T/N \to 0$. ∎

Table 1: Monte Carlo results for DGP1 with $m = r = 1$, and cross-section uncorrelated and homoskedastic errors.

		Forecasting $y_{T+h\|T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.92	0.90	0.98	0.92	0.16	0.16	0.19	0.08	0.97	0.96	0.99	0.96	1.04	1.03	0.99	0.08
50	30	0.93	0.91	0.98	0.92	0.15	0.14	0.17	0.08	0.97	0.96	0.99	0.96	1.01	0.98	0.95	0.08
100	30	0.94	0.92	0.99	0.92	0.11	0.10	0.13	0.08	0.96	0.96	0.99	0.96	0.96	0.95	0.92	0.08
200	30	0.93	0.92	0.99	0.92	0.10	0.09	0.13	0.08	0.97	0.97	0.99	0.97	0.92	0.92	0.89	0.08
30	50	0.92	0.90	0.99	0.93	0.15	0.13	0.15	0.05	0.96	0.96	0.99	0.96	1.07	1.07	1.04	0.05
50	50	0.94	0.91	1.00	0.94	0.12	0.11	0.12	0.04	0.96	0.96	0.99	0.96	1.03	1.01	0.99	0.04
100	50	0.94	0.94	1.00	0.93	0.07	0.07	0.09	0.04	0.96	0.96	0.99	0.96	0.99	0.98	0.96	0.04
200	50	0.95	0.94	1.00	0.94	0.06	0.06	0.08	0.04	0.96	0.96	0.99	0.96	0.96	0.96	0.94	0.04
30	100	0.92	0.91	1.00	0.94	0.13	0.11	0.12	0.02	0.95	0.95	1.00	0.95	1.13	1.11	1.10	0.02
50	100	0.94	0.92	1.00	0.94	0.11	0.08	0.09	0.02	0.96	0.96	1.00	0.95	1.06	1.03	1.02	0.02
100	100	0.94	0.94	1.00	0.94	0.06	0.05	0.06	0.02	0.96	0.95	1.00	0.95	1.02	1.01	1.00	0.02
200	100	0.94	0.94	1.00	0.94	0.04	0.04	0.05	0.02	0.95	0.95	1.00	0.95	1.02	1.02	1.01	0.02
30	200	0.92	0.92	1.00	0.94	0.12	0.10	0.10	0.01	0.97	0.96	1.00	0.96	1.07	1.04	1.04	0.01
50	200	0.93	0.92	1.00	0.94	0.10	0.08	0.09	0.01	0.96	0.96	1.00	0.95	1.06	1.04	1.04	0.01
100	200	0.94	0.94	1.00	0.94	0.04	0.04	0.04	0.01	0.95	0.95	1.00	0.95	1.04	1.02	1.02	0.01
200	200	0.94	0.94	1.00	0.94	0.03	0.03	0.03	0.01	0.95	0.95	1.00	0.95	1.02	1.01	1.01	0.01

Notes: CA, PC1, PC2 and F refer to the results based on factors estimated using cross-section averages, principal components with true number of factors, r, principal components with $k = r + 1$ and true factors, respectively. CR and MSE denote the coverage rate and the empirical mean squared prediction error, respectively. k is the number of factors used by the principal components approach.

Table 2: Monte Carlo results for DGP2 with $m = r = 1$, and cross-section correlated and homoskedastic errors.

		Forecasting $y_{T+h\|T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.88	0.87	0.98	0.92	0.20	0.18	0.20	0.08	0.97	0.96	0.99	0.96	1.07	1.05	1.01	0.08
50	30	0.89	0.91	0.99	0.92	0.18	0.14	0.17	0.08	0.97	0.97	0.99	0.96	1.04	1.00	0.97	0.08
100	30	0.92	0.92	1.00	0.92	0.12	0.11	0.14	0.08	0.96	0.96	0.99	0.96	0.97	0.96	0.92	0.08
200	30	0.92	0.93	0.99	0.92	0.11	0.10	0.13	0.08	0.97	0.97	0.99	0.97	0.93	0.92	0.88	0.08
30	50	0.87	0.87	0.99	0.93	0.18	0.15	0.16	0.05	0.96	0.96	0.99	0.96	1.11	1.08	1.05	0.05
50	50	0.90	0.87	1.00	0.94	0.16	0.12	0.13	0.04	0.96	0.96	0.99	0.96	1.06	1.02	1.00	0.04
100	50	0.93	0.94	1.00	0.93	0.08	0.07	0.09	0.04	0.96	0.96	1.00	0.96	1.00	0.99	0.97	0.04
200	50	0.94	0.92	1.00	0.94	0.07	0.06	0.08	0.04	0.96	0.96	1.00	0.96	0.97	0.96	0.94	0.04
30	100	0.87	0.83	1.00	0.94	0.17	0.13	0.14	0.02	0.95	0.95	0.99	0.95	1.17	1.13	1.11	0.02
50	100	0.89	0.77	1.00	0.94	0.14	0.10	0.10	0.02	0.96	0.95	0.99	0.95	1.09	1.05	1.03	0.02
100	100	0.92	0.92	1.00	0.94	0.07	0.05	0.06	0.02	0.96	0.95	1.00	0.95	1.03	1.01	1.00	0.02
200	100	0.93	0.91	1.00	0.94	0.05	0.04	0.05	0.02	0.95	0.95	1.00	0.95	1.03	1.02	1.01	0.02
30	200	0.88	0.94	1.00	0.94	0.16	0.12	0.12	0.01	0.97	0.97	1.00	0.96	1.11	1.07	1.07	0.01
50	200	0.89	0.81	1.00	0.94	0.14	0.09	0.10	0.01	0.96	0.95	1.00	0.95	1.10	1.05	1.05	0.01
100	200	0.92	0.92	1.00	0.94	0.06	0.04	0.05	0.01	0.95	0.95	1.00	0.95	1.05	1.03	1.02	0.01
200	200	0.93	0.90	1.00	0.94	0.04	0.03	0.04	0.01	0.95	0.95	1.00	0.95	1.03	1.02	1.01	0.01

Notes: See the notes to Table 1.

Table 3: Monte Carlo results for DGP3 with $m = r = 1$, and cross-section uncorrelated and heteroskedastic errors.

		Forecasting $y_{T+h\|T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.91	0.88	0.98	0.92	0.16	0.16	0.19	0.08	0.97	0.96	0.98	0.96	1.03	1.03	0.99	0.08
50	30	0.92	0.91	0.98	0.92	0.15	0.14	0.17	0.08	0.97	0.97	0.99	0.96	1.01	0.98	0.95	0.08
100	30	0.93	0.93	0.99	0.92	0.11	0.10	0.13	0.08	0.96	0.96	0.99	0.96	0.96	0.95	0.92	0.08
200	30	0.93	0.92	0.99	0.92	0.10	0.09	0.13	0.08	0.97	0.96	0.99	0.97	0.92	0.92	0.89	0.08
30	50	0.90	0.89	0.99	0.93	0.15	0.13	0.15	0.05	0.96	0.96	0.99	0.96	1.07	1.06	1.04	0.05
50	50	0.92	0.89	0.99	0.94	0.12	0.11	0.12	0.04	0.96	0.96	0.99	0.96	1.03	1.01	0.98	0.04
100	50	0.94	0.94	1.00	0.93	0.07	0.07	0.09	0.04	0.96	0.96	0.99	0.96	0.99	0.98	0.96	0.04
200	50	0.95	0.93	1.00	0.94	0.06	0.06	0.08	0.04	0.96	0.96	0.99	0.96	0.96	0.96	0.94	0.04
30	100	0.91	0.88	1.00	0.94	0.13	0.11	0.12	0.02	0.95	0.95	0.99	0.95	1.13	1.11	1.10	0.02
50	100	0.92	0.83	1.00	0.94	0.10	0.08	0.09	0.02	0.96	0.96	0.99	0.95	1.06	1.03	1.02	0.02
100	100	0.94	0.93	1.00	0.94	0.06	0.05	0.06	0.02	0.96	0.95	1.00	0.95	1.02	1.01	1.00	0.02
200	100	0.94	0.93	1.00	0.94	0.04	0.04	0.05	0.02	0.95	0.95	1.00	0.95	1.02	1.02	1.01	0.02
30	200	0.92	0.96	1.00	0.94	0.12	0.10	0.10	0.01	0.96	0.97	1.00	0.96	1.07	1.04	1.04	0.01
50	200	0.92	0.85	1.00	0.94	0.10	0.08	0.09	0.01	0.96	0.95	1.00	0.95	1.06	1.04	1.04	0.01
100	200	0.93	0.93	1.00	0.94	0.04	0.04	0.04	0.01	0.95	0.95	1.00	0.95	1.03	1.02	1.02	0.01
200	200	0.94	0.92	1.00	0.94	0.03	0.03	0.03	0.01	0.95	0.95	1.00	0.95	1.02	1.02	1.01	0.01

Notes: See the notes to Table 1.

Table 4: Monte Carlo results for DGP4 with $m = r = 1$, and cross-section correlated and heteroskedastic errors.

		Forecasting $y_{T+h\|T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.88	0.87	0.98	0.92	0.19	0.18	0.20	0.08	0.96	0.96	0.99	0.96	1.07	1.04	1.00	0.08
50	30	0.89	0.90	0.99	0.92	0.18	0.14	0.17	0.08	0.97	0.97	0.99	0.96	1.04	0.99	0.96	0.08
100	30	0.92	0.92	0.99	0.92	0.12	0.11	0.14	0.08	0.96	0.96	0.99	0.96	0.97	0.96	0.92	0.08
200	30	0.92	0.93	0.99	0.92	0.11	0.10	0.13	0.08	0.97	0.96	0.99	0.97	0.93	0.92	0.88	0.08
30	50	0.87	0.87	0.99	0.93	0.18	0.15	0.16	0.05	0.96	0.96	0.99	0.96	1.11	1.08	1.05	0.05
50	50	0.90	0.88	1.00	0.94	0.15	0.12	0.13	0.04	0.96	0.96	0.99	0.96	1.06	1.02	1.00	0.04
100	50	0.93	0.94	1.00	0.93	0.09	0.07	0.09	0.04	0.96	0.96	1.00	0.96	1.00	0.99	0.97	0.04
200	50	0.93	0.92	1.00	0.94	0.07	0.06	0.08	0.04	0.96	0.96	0.99	0.96	0.97	0.96	0.94	0.04
30	100	0.86	0.83	1.00	0.94	0.17	0.13	0.14	0.02	0.95	0.95	0.99	0.95	1.17	1.13	1.11	0.02
50	100	0.89	0.77	1.00	0.94	0.14	0.09	0.10	0.02	0.96	0.96	0.99	0.95	1.09	1.05	1.03	0.02
100	100	0.92	0.92	1.00	0.94	0.07	0.05	0.06	0.02	0.96	0.95	1.00	0.95	1.03	1.02	1.00	0.02
200	100	0.93	0.92	1.00	0.94	0.05	0.04	0.05	0.02	0.95	0.95	1.00	0.95	1.03	1.02	1.01	0.02
30	200	0.88	0.94	1.00	0.94	0.15	0.12	0.12	0.01	0.96	0.97	1.00	0.96	1.11	1.07	1.06	0.01
50	200	0.89	0.81	1.00	0.94	0.14	0.09	0.10	0.01	0.96	0.95	1.00	0.95	1.10	1.05	1.05	0.01
100	200	0.92	0.93	1.00	0.94	0.06	0.04	0.05	0.01	0.95	0.95	1.00	0.95	1.05	1.03	1.02	0.01
200	200	0.93	0.90	1.00	0.94	0.04	0.03	0.04	0.01	0.95	0.95	1.00	0.95	1.03	1.02	1.01	0.01

Notes: See the notes to Table 1.

Table 5: Monte Carlo results for DGP5 with $m = 2 > r = 1$, and cross-section uncorrelated and homoskedastic errors.

		Forecasting $y_{T+h\mid T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.98	0.90	0.98	0.92	0.16	0.16	0.19	0.08	0.98	0.97	0.99	0.97	0.93	0.98	0.94	0.08
50	30	0.98	0.91	0.98	0.92	0.14	0.13	0.16	0.08	0.98	0.97	0.99	0.96	0.91	0.98	0.94	0.08
100	30	0.98	0.92	0.99	0.92	0.13	0.11	0.14	0.08	0.97	0.96	0.99	0.96	0.94	1.00	0.96	0.08
200	30	0.98	0.93	0.99	0.93	0.12	0.09	0.12	0.08	0.98	0.97	0.99	0.97	0.87	0.91	0.87	0.08
30	50	0.98	0.89	0.99	0.93	0.14	0.16	0.17	0.04	0.97	0.96	0.99	0.95	1.04	1.09	1.07	0.04
50	50	0.98	0.93	1.00	0.93	0.10	0.10	0.12	0.05	0.97	0.96	0.99	0.96	0.98	1.02	1.00	0.05
100	50	0.98	0.93	1.00	0.93	0.08	0.07	0.09	0.04	0.96	0.96	0.99	0.96	0.93	0.97	0.95	0.04
200	50	0.98	0.93	1.00	0.94	0.07	0.06	0.08	0.04	0.96	0.96	0.99	0.96	0.96	0.99	0.96	0.04
30	100	0.99	0.92	1.00	0.94	0.09	0.09	0.10	0.02	0.97	0.96	1.00	0.96	0.98	1.01	1.00	0.02
50	100	0.99	0.92	1.00	0.94	0.07	0.09	0.10	0.02	0.96	0.95	1.00	0.95	1.04	1.07	1.07	0.02
100	100	0.99	0.94	1.00	0.94	0.05	0.05	0.06	0.02	0.96	0.95	1.00	0.95	0.98	1.00	0.99	0.02
200	100	0.99	0.95	1.00	0.94	0.04	0.04	0.05	0.02	0.96	0.95	1.00	0.95	0.99	1.01	1.00	0.02
30	200	0.99	0.89	1.00	0.94	0.09	0.12	0.12	0.01	0.96	0.95	1.00	0.95	1.08	1.13	1.12	0.01
50	200	0.99	0.93	1.00	0.94	0.05	0.07	0.07	0.01	0.96	0.96	1.00	0.96	1.00	1.04	1.03	0.01
100	200	0.99	0.93	1.00	0.95	0.03	0.04	0.05	0.01	0.95	0.95	1.00	0.95	1.01	1.04	1.03	0.01
200	200	0.99	0.94	1.00	0.95	0.03	0.03	0.03	0.01	0.95	0.95	1.00	0.95	1.01	1.02	1.01	0.01

Notes: See the notes to Table 1.

Table 6: Monte Carlo results for DGP6 with $m = 2 > r = 1$, and cross-section correlated and homoskedastic errors.

		Forecasting $y_{T+h\mid T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.95	0.87	0.98	0.92	0.18	0.17	0.20	0.08	0.97	0.97	0.99	0.97	0.94	0.99	0.95	0.08
50	30	0.94	0.91	0.99	0.92	0.16	0.14	0.17	0.08	0.97	0.97	0.99	0.96	0.92	0.98	0.95	0.08
100	30	0.94	0.90	0.99	0.92	0.14	0.11	0.14	0.08	0.97	0.96	0.99	0.96	0.95	1.00	0.96	0.08
200	30	0.94	0.92	0.99	0.93	0.12	0.09	0.12	0.08	0.97	0.97	0.99	0.97	0.88	0.91	0.88	0.08
30	50	0.95	0.87	0.99	0.93	0.17	0.17	0.19	0.04	0.97	0.96	0.99	0.95	1.07	1.11	1.09	0.04
50	50	0.96	0.85	0.99	0.93	0.11	0.11	0.13	0.05	0.97	0.96	0.99	0.96	0.99	1.03	1.00	0.05
100	50	0.96	0.89	1.00	0.93	0.09	0.08	0.10	0.04	0.96	0.96	0.99	0.96	0.94	0.97	0.96	0.04
200	50	0.94	0.92	1.00	0.94	0.08	0.06	0.08	0.04	0.96	0.96	1.00	0.96	0.97	0.99	0.96	0.04
30	100	0.96	0.86	1.00	0.94	0.11	0.11	0.12	0.02	0.97	0.96	1.00	0.96	1.00	1.03	1.02	0.02
50	100	0.96	0.92	1.00	0.94	0.09	0.10	0.11	0.02	0.96	0.95	1.00	0.95	1.05	1.09	1.07	0.02
100	100	0.97	0.91	1.00	0.94	0.06	0.06	0.07	0.02	0.96	0.96	1.00	0.95	0.99	1.01	1.00	0.02
200	100	0.97	0.95	1.00	0.94	0.05	0.04	0.05	0.02	0.96	0.95	1.00	0.95	1.00	1.02	1.01	0.02
30	200	0.97	0.91	1.00	0.94	0.12	0.14	0.14	0.01	0.96	0.96	1.00	0.95	1.10	1.14	1.13	0.01
50	200	0.97	0.96	1.00	0.94	0.06	0.08	0.08	0.01	0.96	0.96	1.00	0.96	1.02	1.05	1.04	0.01
100	200	0.98	0.82	1.00	0.95	0.04	0.05	0.05	0.01	0.95	0.95	1.00	0.95	1.02	1.05	1.04	0.01
200	200	0.98	0.96	1.00	0.95	0.03	0.03	0.03	0.01	0.95	0.95	1.00	0.95	1.01	1.02	1.02	0.01

Notes: See the notes to Table 1.

Table 7: Monte Carlo results for DGP7 with $m = 2 > r = 1$, and cross-section uncorrelated and heteroskedastic errors.

		Forecasting $y_{T+h\|T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.95	0.88	0.98	0.92	0.16	0.16	0.19	0.08	0.97	0.97	0.99	0.97	0.92	0.98	0.94	0.08
50	30	0.95	0.91	0.98	0.92	0.14	0.14	0.16	0.08	0.97	0.97	0.99	0.96	0.91	0.98	0.94	0.08
100	30	0.94	0.91	0.98	0.92	0.13	0.11	0.14	0.08	0.97	0.96	0.98	0.96	0.94	1.00	0.96	0.08
200	30	0.94	0.92	0.99	0.93	0.12	0.09	0.12	0.08	0.97	0.97	0.99	0.97	0.87	0.91	0.88	0.08
30	50	0.97	0.88	0.99	0.93	0.14	0.16	0.17	0.04	0.97	0.96	0.99	0.95	1.04	1.10	1.07	0.04
50	50	0.96	0.87	0.99	0.93	0.10	0.10	0.12	0.05	0.96	0.96	0.98	0.96	0.98	1.02	1.00	0.05
100	50	0.96	0.91	1.00	0.93	0.08	0.07	0.09	0.04	0.96	0.96	0.99	0.96	0.93	0.97	0.95	0.04
200	50	0.94	0.93	1.00	0.94	0.07	0.06	0.08	0.04	0.96	0.96	0.99	0.96	0.96	0.98	0.96	0.04
30	100	0.97	0.90	1.00	0.94	0.09	0.09	0.10	0.02	0.97	0.96	1.00	0.96	0.98	1.01	1.00	0.02
50	100	0.98	0.92	1.00	0.94	0.07	0.09	0.10	0.02	0.96	0.95	1.00	0.95	1.04	1.07	1.07	0.02
100	100	0.97	0.92	1.00	0.94	0.05	0.05	0.06	0.02	0.96	0.95	1.00	0.95	0.98	1.00	0.99	0.02
200	100	0.97	0.95	1.00	0.94	0.04	0.04	0.05	0.02	0.96	0.95	1.00	0.95	0.99	1.01	1.01	0.02
30	200	0.98	0.93	1.00	0.94	0.09	0.12	0.12	0.01	0.96	0.95	1.00	0.95	1.07	1.13	1.12	0.01
50	200	0.98	0.96	1.00	0.94	0.05	0.07	0.07	0.01	0.96	0.96	1.00	0.96	1.00	1.04	1.03	0.01
100	200	0.98	0.84	1.00	0.95	0.03	0.04	0.05	0.01	0.95	0.95	1.00	0.95	1.01	1.04	1.03	0.01
200	200	0.98	0.96	1.00	0.95	0.03	0.03	0.03	0.01	0.95	0.95	1.00	0.95	1.01	1.02	1.01	0.01

Notes: See the notes to Table 1.

Table 8: Monte Carlo results for DGP8 with $m = 2 > r = 1$, and cross-section correlated and heteroskedastic errors.

		Forecasting $y_{T+h\|T}$								Forecasting y_{T+h}							
		CR				MSE				CR				MSE			
N	T	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F	CA	PC1	PC2	F
30	30	0.94	0.87	0.98	0.92	0.18	0.17	0.20	0.08	0.97	0.97	0.99	0.97	0.94	0.99	0.95	0.08
50	30	0.94	0.91	0.99	0.92	0.16	0.14	0.17	0.08	0.97	0.97	0.99	0.96	0.92	0.98	0.95	0.08
100	30	0.94	0.90	0.99	0.92	0.14	0.11	0.14	0.08	0.97	0.96	0.99	0.96	0.95	1.00	0.96	0.08
200	30	0.93	0.92	0.99	0.93	0.12	0.09	0.12	0.08	0.97	0.97	0.99	0.97	0.88	0.91	0.88	0.08
30	50	0.95	0.87	0.99	0.93	0.17	0.17	0.18	0.04	0.97	0.96	0.99	0.95	1.07	1.11	1.08	0.04
50	50	0.96	0.85	0.99	0.93	0.11	0.11	0.13	0.05	0.96	0.96	0.99	0.96	0.99	1.03	1.00	0.05
100	50	0.96	0.89	1.00	0.93	0.09	0.08	0.10	0.04	0.96	0.96	0.99	0.96	0.94	0.97	0.96	0.04
200	50	0.94	0.92	1.00	0.94	0.08	0.06	0.08	0.04	0.96	0.96	1.00	0.96	0.97	0.99	0.96	0.04
30	100	0.96	0.86	1.00	0.94	0.11	0.11	0.12	0.02	0.97	0.96	1.00	0.96	1.00	1.03	1.02	0.02
50	100	0.96	0.91	1.00	0.94	0.09	0.10	0.11	0.02	0.96	0.95	1.00	0.95	1.05	1.09	1.08	0.02
100	100	0.97	0.91	1.00	0.94	0.06	0.06	0.07	0.02	0.96	0.96	1.00	0.95	0.99	1.01	1.00	0.02
200	100	0.97	0.95	1.00	0.94	0.05	0.04	0.05	0.02	0.96	0.95	1.00	0.95	1.00	1.02	1.01	0.02
30	200	0.96	0.91	1.00	0.94	0.12	0.14	0.14	0.01	0.96	0.95	1.00	0.95	1.10	1.14	1.13	0.01
50	200	0.97	0.95	1.00	0.94	0.06	0.08	0.08	0.01	0.96	0.96	1.00	0.96	1.02	1.05	1.04	0.01
100	200	0.98	0.83	1.00	0.95	0.04	0.05	0.05	0.01	0.96	0.95	1.00	0.95	1.02	1.04	1.04	0.01
200	200	0.98	0.96	1.00	0.95	0.03	0.03	0.03	0.01	0.95	0.95	1.00	0.95	1.01	1.02	1.02	0.01

Notes: See the notes to Table 1.